Development of new and innovative tools for the characterisation, modelling and forecasting of the hydrogeological behaviour of karst aquifers in response to climatic and anthropogenic changes

Guillaume Cinkus

▶ **To cite this version:**

Guillaume Cinkus. Development of new and innovative tools for the characterisation, modelling and forecasting of the hydrogeological behaviour of karst aquifers in response to climatic and anthropogenic changes. Hydrology. Montpellier University, 2023. English. NNT : . tel-04441501

HAL Id: tel-04441501
https://hal.science/tel-04441501

Submitted on 6 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Sciences de l'Eau

École doctorale GAIA

Unité de recherche HydroSciences Montpellier

Développement d'outils nouveaux et innovants pour la caractérisation, la modélisation et la prévision du fonctionnement des aquifères karstiques en réponse aux changements climatiques et anthropiques

Présentée par Guillaume CINKUS
Le 05 octobre 2023

Sous la direction de Hervé JOURDE et Naomi MAZZILLI

Devant le jury composé de

Roger MOUSSA, Directeur de recherche, INRAE centre Occitanie-Montpellier	Président du jury
Wendy GRAHAM, Carl S. Swisher Eminent Scholar in Water Resources & Director, University of Florida	Rapporteure
Andreas HARTMANN, Full Professor, Technical University of Dresden	Rapporteur
Francesco FIORILLO, Full Professor of Engineering Geology, University of Sannio	Examinateur
Hervé JOURDE, Professeur, Université de Montpellier	Directeur de thèse
Naomi MAZZILLI, Maîtresse de conférences, Université d'Avignon	Co-Directrice de thèse
Joanna DOUMMAR, Associate Professor, Chair, American University of Beirut	Invitée

UNIVERSITÉ DE MONTPELLIER

Contents

Remerciements

Je suis extrêmement reconnaissant envers Hervé Jourde et Naomi Mazzilli pour m'avoir fait confiance et donné l'opportunité de travailler sur cette thèse. C'est grâce à leur capacité d'écoute, leurs conseils et leurs compétences que mon expérience a été si intéressante, stimulante et complète – me permettant de me développer autant personnellement que professionnellement.

Je tiens également à remercier les unités mixtes de recherche HydroSciences Montpellier et EMMAH pour leur excellent accueil. Merci à toutes les personnes au sein de ces structures qui m'ont accompagné et conseillé tout au long de ma thèse, sur les aspects administratifs, matériels et scientifiques.

Un grand merci aux membres du jury d'avoir pris le temps de lire et évaluer mon travail. Merci à Wendy Graham et Andreas Hartmann d'avoir accepté d'être les rapporteurs de cette thèse. Merci à Francesco Fiorillo et Roger Moussa d'avoir accepté le rôle d'examinateur. Merci à Joanna Doummar pour sa présence en tant qu'invitée.

Je remercie vivement :

i. Nico Goldscheider, Christian Leduc, Albert Olioso et Yves Tramblay pour leur participation aux comités de suivi individuel et leurs conseils avisés.

ii. Nataša Ravbar pour son accueil au Karst Research Institute et sa disponibilité lors de nos collaborations.

iii. Zhao Chen pour son aide à l'utilisation d'une routine de neige.

iv. Vianney Sivelle pour son aide sur de nombreux aspects techniques et scientifiques, ainsi que pour les différentes collaborations.

v. Leïla Serene pour son aide et ses conseils lors de la finalisation du manuscrit et la préparation à la soutenance.

vi. les membres du projet KARMA pour les nombreuses interactions scientifiques tout au long de ma thèse, le partage des données, ainsi que pour m'avoir donné l'opportunité de travailler dans un cadre international.

vii. les membres du SNO KARST pour les stimulants échanges scientifiques depuis de nombreuses années et le partage des données.

viii. les personnes qui m'ont accompagné lors des activités de monitorat.

ix. tous ceux qui, de près ou de loin, m'ont aidé, conseillé et soutenu.

Enfin, un immense merci à ma famille et à mes amis.

"Have you ever seen flowing water?... Have you ever seen still water?... If your mind is peaceful it will be just like still, flowing water. Have you ever seen still, flowing water? There! You've only ever seen flowing water and still water, haven't you? But you've never seen still, flowing water. Right there, right where your thinking cannot take you, even though it's peaceful you can develop wisdom. Your mind will be like flowing water, and yet it's still. It's almost as if it were still, and yet it's flowing. So I call it "still, flowing water." Wisdom can arise here."

– Ajahn Chah, *Wat Tham Saeng Phet, rains retreat of 1981.*

Résumé étendu

Problématique

Les formations carbonatées karstiques affleurent sur environ 15.2 % de la surface continentale mondiale hors glaciers (Goldscheider et al., 2020) et sont largement représentées dans le sous-sol (Chen et al., 2017). Elles sont essentielles pour l'alimentation en eau potable, ainsi que pour l'agriculture et le maintien du fonctionnement des écosystèmes, en particulier dans la région méditerranéenne où 15 % de l'eau domestique, et jusqu'à 90 % dans certaines régions, provient des aquifères karstiques (Bakalowicz, 2015; Stevanović, 2019). Ces systèmes présentent généralement un haut degré d'hétérogénéité et sont gouvernés par des processus complexes régulant le flux et le stockage de l'eau, ce qui les rend difficiles à étudier et particulièrement vulnérables aux changements environnementaux (Parise et al., 2018).

Le changement climatique, les pressions anthropiques, ainsi que les évolutions des régimes de précipitations et d'évapotranspiration pourraient avoir un impact significatif sur la recharge des aquifères karstiques. Les propriétés hétérogènes de ces systèmes et la non-linéarité de leurs processus internes les rendent particulièrement difficiles à étudier – une exploration plus approfondie de leur réponse aux changements globaux est donc nécessaire. Les analyses hydrologiques et les modèles numériques sont des outils pertinents pour ce type d'investigation (Hartmann et al., 2014) et sont utilisés dans les environnements karstiques à différentes fins : (i) simuler le débit des sources (Fleury et al., 2007; Wunsch et al., 2022), (ii) prédire et prévoir les inondations (Kong-A-Siou et al., 2011; Wu et al., 2008), (iii) gérer l'exploitation des aquifères (Fleury et al., 2009; Xanke et al., 2016), (iv) caractériser le fonctionnement spécifique et interne du système (Baudement et al., 2017; Dubois et al., 2020), (v) enquêter sur la genèse et l'évolution du système (Kaufmann and Romanov, 2020; Liedl et al., 2003), (vi) évaluer la qualité et la vulnérabilité de la ressource en eau (Husic et al., 2019; Sullivan et al., 2019), et (vii) étudier l'impact du changement climatique sur le fonctionnement du système (Hartmann et al., 2012; Sivelle et al., 2021). En particulier, les modèles à paramètres globaux permettent d'étudier les systèmes karstiques sans nécessiter de données météorologiques et systémiques à haute résolution spatiale. Ils sont l'aspect central de cette thèse, qui vise à développer et améliorer les analyses hydrologiques et les modèles réservoir pour caractériser le fonctionnement des systèmes karstiques en réponse aux changements climatiques et anthropiques. Un accent est mis sur les questions de recherche suivantes :

i. Comment caractériser le fonctionnement d'un système karstique dans un contexte de manque de données ?

ii. Sur quels aspects peut-on encore améliorer les modèles simples, unidimensionnels ?

iii. Peut-on se fier aux critères de performance pour la calibration et l'évaluation des modèles hydrologiques ?

iv. Quels sont les avantages et les inconvénients des différentes approches de modélisation empiriques et conceptuelles en hydrologie karstique ?

Organisation du mémoire et principaux résultats

Le Chapitre 1 fournit une brève description de la formation, du fonctionnement et des particularités des systèmes karstiques. La modélisation des écoulements pluie-débit dans les environnements karstiques est introduite, avec un rappel des défis rencontrés et une description des différentes approches de modélisation couramment utilisées. En particulier, les modèles à paramètres globaux offrent un bon compromis entre accessibilité et pertinence du modèle. Leur utilisation dans les environnements karstiques est complexe en raison de l'hétérogénéité caractéristique de ces systèmes, ajoutant un autre défi en plus des incertitudes et biais typiques associés à la modélisation hydrologique. Ils permettent néanmoins d'obtenir des informations précieuses pour la caractérisation et la prédiction du fonctionnement de ces systèmes (Hartmann et al., 2014).

La première partie est dédiée à la caractérisation du fonctionnement hydrologique des systèmes karstiques.

Le Chapitre 2 présente une nouvelle classification des systèmes karstiques basée sur l'analyse des débits à la source. Cette étude tire profit de larges bases de données de débit de sources (Jourde et al., 2018; Olarinoye et al., 2020) pour prendre en compte la grande diversité des fonctionnements hydrologiques existants. Les séries temporelles de débit sont étudiées avec l'analyse des courbes de récession, ainsi que des analyses statistiques et de signal pour identifier des indicateurs de fonctionnement hydrologique. La sélection des indicateurs les plus pertinents et la proposition de la classification sont basées sur des analyses multivariées. La classification repose sur l'analyse des courbes de récession et permet de discriminer six classes à l'aide de trois indicateurs reflétant la capacité de stockage dynamique, la dynamique de drainage de la fonction capacitive, et la variabilité du fonctionnement hydrologique (Figure 1). La méthode est développée pour être adaptée aux contextes de manque de données et est évaluée sur une base de données de 78 systèmes karstiques situés dans le monde entier. Tous les systèmes se répartissent de manière homogène parmi les six classes proposées, ce qui met en évidence la pertinence de l'approche et la représentativité des différentes classes de fonctionnement hydrologique. Les résultats de la méthodologie proposée sont finalement discutés pour explorer ses limites et définir des conditions pour son application.

Le Chapitre 3 présente une boîte à outils développée dans le cadre de cette thèse. KarstID aide les utilisateurs à analyser les séries chronologiques de débit et à caractériser le fonctionnement hydrologique des systèmes karstiques en utilisant la classification décrite dans le Chapitre 2. Les analyses de séries chronologiques de débit de sources sont souvent utilisées pour obtenir une première compréhension du fonctionnement hydrologique d'un système karstique. Le logiciel permet (i) d'effectuer des analyses statistiques, de courbes de récession (Figure 2), de classements des débits, et de signal (analyses corrélatoire et spectrale simples) ; (ii) de calculer des indicateurs représentatifs de différents aspects du fonctionnement d'un système ; (iii) de classer le fonctionnement hydrologique du système étudié ; et (iv) de comparer les résultats à une base de données de 78 systèmes karstiques. La réalisation de ces analyses est facilitée par une interface graphique qui ne nécessite aucune compétence en programmation – KarstID peut être ainsi utilisé à la fois pour la recherche et à des fins pédagogiques. Le logiciel est également gratuit, open source, et activement développé sur une plateforme communautaire

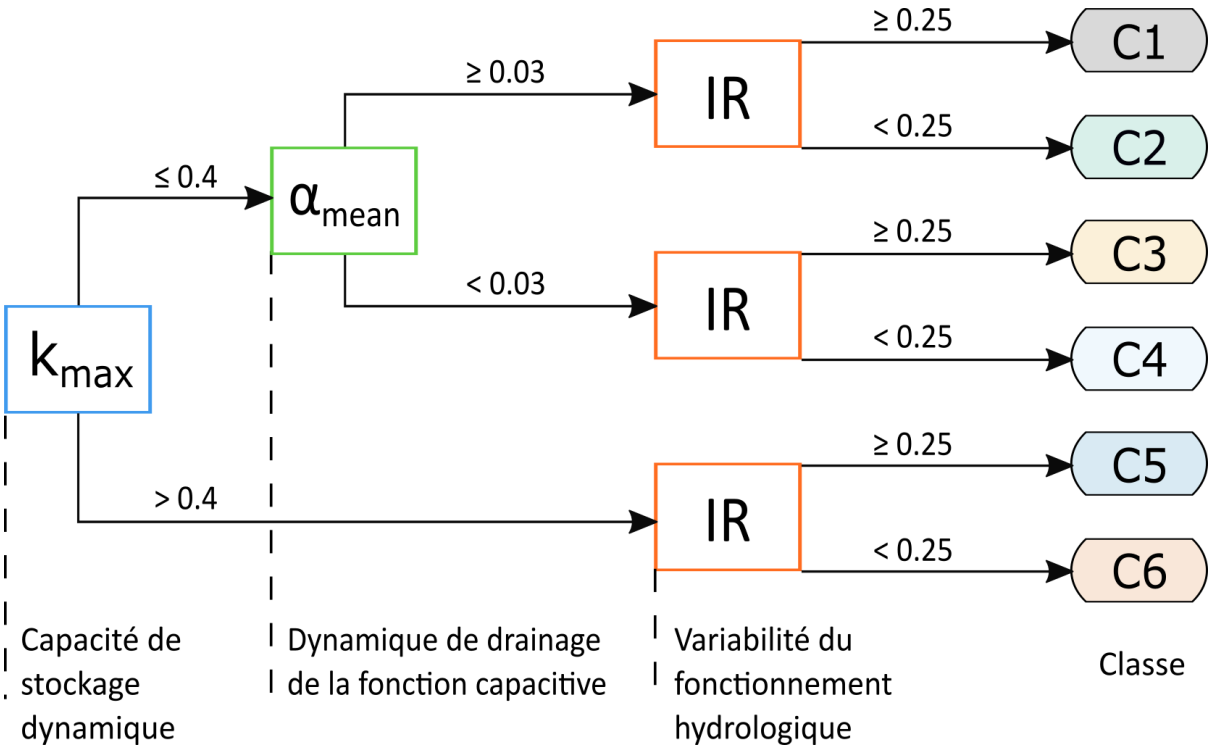


Figure 1: Diagramme pour la classification des systèmes karstiques à partir des trois indicateurs issus de l'analyse des courbes de récession.

de développeurs. L'application et son manuel utilisateur sont tous deux disponibles sur le site web du SNO KARST (Jourde et al., 2018).

La deuxième partie est consacrée à l'amélioration et l'évaluation des performances des modèles à paramètres globaux pour la simulation des débits et des niveaux piézométriques des systèmes karstiques.

Le Chapitre 4 présente les nouvelles fonctionnalités développées pour la version 3.0 du logiciel KarstMod, une plateforme dédiée à la modélisation pluie-débit des aquifères karstiques (Mazzilli et al., 2019; Mazzilli et al., 2023). KarstMod offre un environnement de modélisation modulaire et accessible pour des fins éducatives, de recherche et opérationnelles. Il comprend également des outils numériques pour l'analyse des séries temporelles, l'évaluation des modèles et les analyses de sensibilité. La modularité de la plateforme facilite les opérations courantes relatives aux modèles à paramètres globaux, telles que (i) la mise en place et l'estimation des paramètres d'une structure de modèle pertinente, et (ii) l'évaluation de la sensibilité des paramètres et des caractéristiques des simulations. Ce travail a porté sur deux aspects principaux : (i) l'utilisation de données d'entrée pertinentes à travers des modules optionnels (routines de neige et d'évapotranspiration potentielle), (ii) l'expansion des possibilités de calibration multi-objectifs et multi-variables, offrant davantage de flexibilité en termes de fonctions objectifs ainsi que de variables de calibration, et (iii) des outils supplémentaires pour l'évaluation des performances du modèle, incluant de nouveaux critères de performance et des graphiques associés. La nouvelle version est appliquée sur deux études de cas en France (les systèmes karstiques de la Touvre et du Lez).

Le Chapitre 5 examine le mécanisme de compensation des erreurs sur le critère de

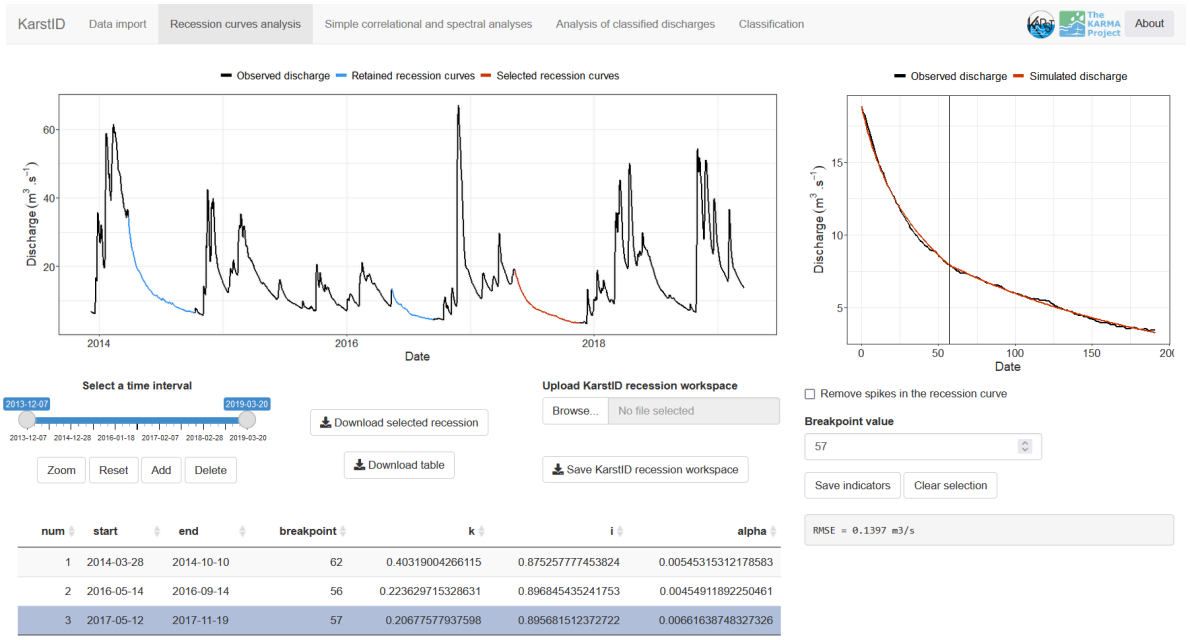
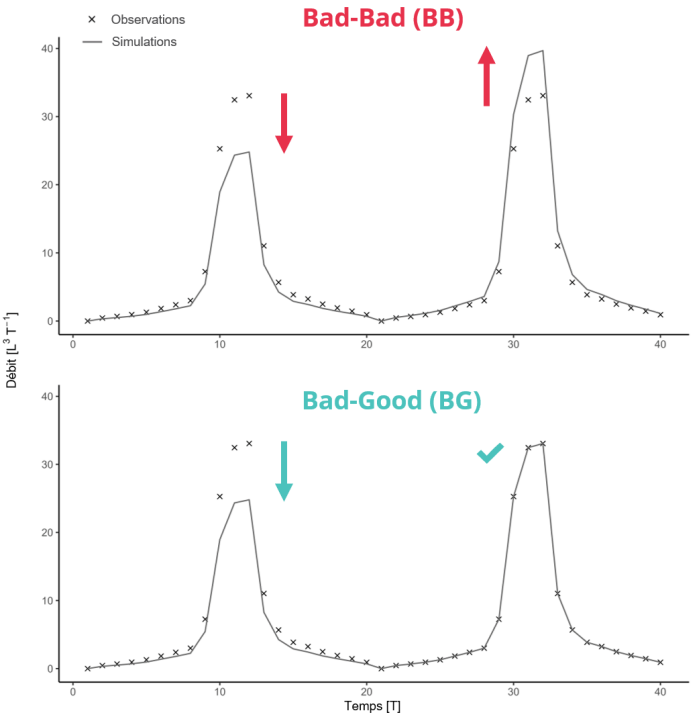


Figure 2: Onglet pour la sélection et l'analyse des courbes de récession dans KarstID. Le graphique de gauche présente les séries temporelles étudiées et les courbes de récession sélectionnées. Le graphique de droite affiche la récession sélectionnée avec le modèle de récession de Mangin (1975). Le tableau montre les détails de chaque courbe de récession et les valeurs des indicateurs correspondants.

performance *Kling-Gupta Efficiency* (KGE) (Gupta et al., 2009) et ses variantes (Kling et al., 2012; Lee and Choi, 2022; Liu, 2020; Pool et al., 2018; Schwemmle et al., 2021; Tang et al., 2021). Les critères de performance jouent un rôle clé dans la calibration et l'évaluation des modèles hydrologiques et ont été largement développés et étudiés, mais certains des critères les plus utilisés présentent encore des écueils mal identifiés. Au total, neuf critères de performance – incluant le KGE et ses variantes, ainsi que le critère de Nash-Sutcliffe (NSE) (Nash and Sutcliffe, 1970) et l'*Index of Agreement* (d_1) (Willmott et al., 1985) – et leur décomposition sont comparés sur des séries chronologiques synthétiques et un cas d'étude réel. L'étude démontre que l'évaluation d'un modèle présentant des surestimations et sous-estimations simultanées peut amener à des scores globaux de critères plus élevés sans pour autant qu'ils soient associés à une amélioration de la pertinence du modèle (Figure 3). Ces erreurs de compensation favorise les paramètres de biais et de variabilité, qui représentent généralement deux tiers du poids dans l'équation du KGE et de ses variantes. Il est recommandé d'utiliser (i) des critères de performance qui sont peu ou pas sujets aux erreurs de compensation (d_1, KGE modifié, KGE non paramétrique, *Diagnostic Efficiency*), et/ou (ii) des facteurs d'échelle dans l'équation pour réduire l'influence des paramètres relatifs (Gupta et al., 2009).

Le Chapitre 6 présente une comparaison de deux approches de modélisation largement utilisées en hydrologie karstique : les réseaux de neurones artificiels (ANN, pour *Artificial Neural Network*) et les modèles réservoir. De nombreuses approches de modélisation peuvent être utilisées pour l'étude des ressources en eau karstiques, ce qui peut rendre difficile le choix de la méthode la plus appropriée. Les modèles empiriques et conceptuels présentent un intérêt particulier dans les environnements karstiques en raison de leur accessibilité vis-à-vis de la complexité et de l'hétérogénéité de ces systèmes. Cinq systèmes



	BB	BG
KGE	**0.95**	0.83
KGE′	**0.94**	0.87
KGE″	**0.96**	0.86
KGE_{NP}	**0.92**	0.86
DE	**0.89**	0.85
LME	**0.96**	0.82
LCE	**0.94**	0.79
NSE	0.92	**0.95**
d_1	0.89	**0.94**

Figure 3: Chroniques synthétiques mettant en évidence l'influence de la compensation des biais et variabilité sur l'optimisation et l'évaluation par le KGE et ses variantes.

karstiques aux caractéristiques différentes en termes de conditions climatiques, de propriétés hydrogéologiques et de disponibilité des données sont étudiés pour la comparaison des approches de modélisation. Les résultats de chaque modèle sont comparés sur la base de plusieurs critères de performance appliqués à différentes périodes hydrologiques. Les résultats montrent que les modèles ANN et réservoir peuvent simuler avec précision le débit des sources karstiques (Figure 4, Figure 5), mais également qu'ils ont différents avantages et inconvénients : (i) Les modèles ANN sont très flexibles concernant le format et la quantité des données d'entrée du modèle, (ii) les modèles réservoir peuvent fournir de bons résultats même avec seulement quelques années d'observations de débit pour la période de calibration, et (iii) les modèles ANN semblent robustes pour reproduire des conditions de débit élevé tandis que les modèles réservoir sont supérieurs pour reproduire des conditions de débit faible. Cependant, les deux approches de modélisation ont des difficultés à reproduire les événements extrêmes (étiages, inondations), ce qui est une limitation connue en modélisation hydrologique. Pour des objectifs de recherche, les modèles ANN peuvent être utiles pour dériver des informations à partir des données d'entrée, par exemple l'identification des zones de recharge. Les modèles réservoir peuvent quant à eux permettre de caractériser le fonctionnement hydrologique d'un système, en étudiant la structure et les paramètres du modèle.

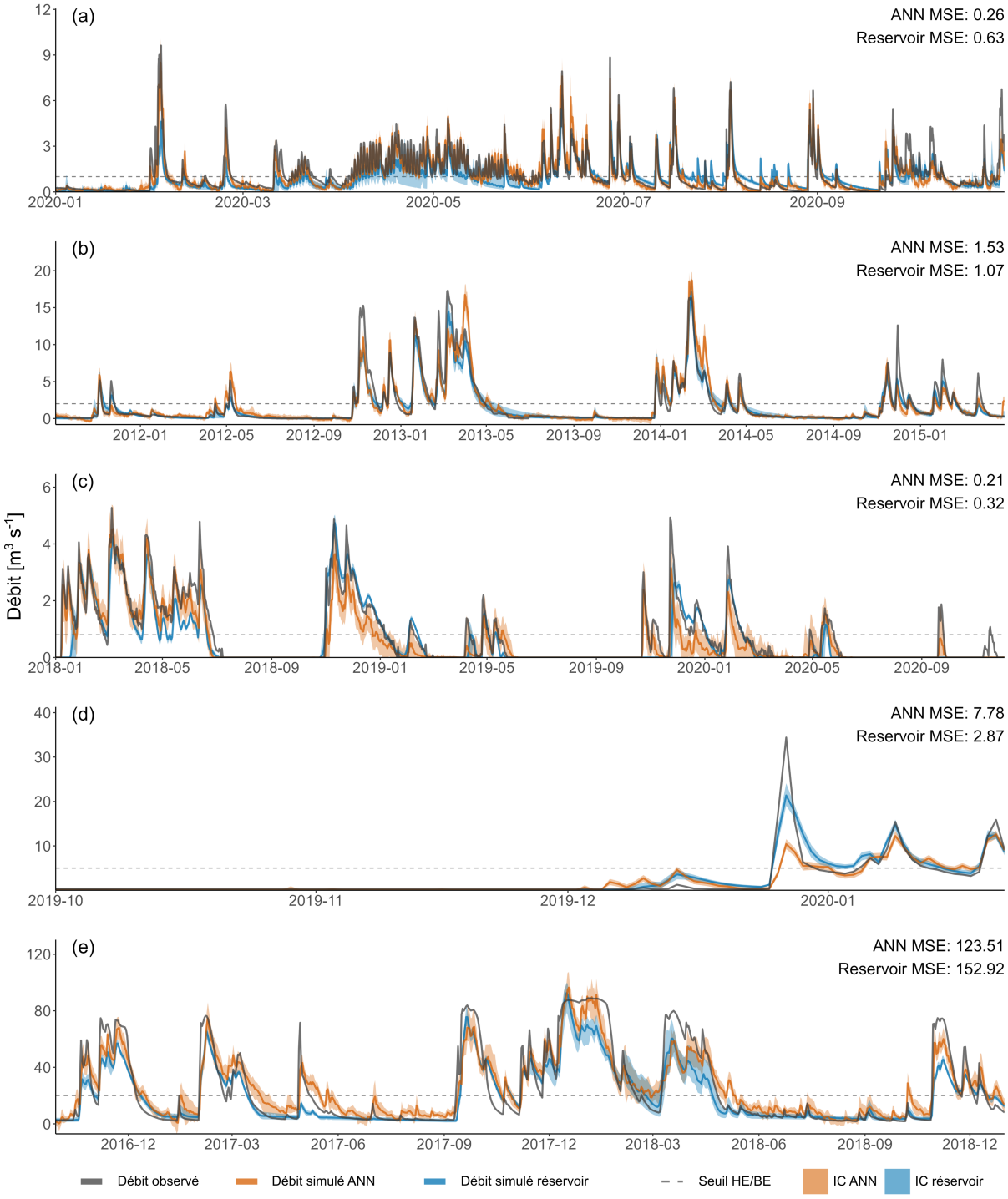


Figure 4: Séries temporelles des débits observés et simulés lors de la période de validation avec (i) des intervalles de confiance (IC) à 90 %, et (ii) le seuil pour les hautes eaux (HE) et faibles eaux (BE) utilisé pour le calcul des critères de performance. (a) Source d'Aubach, (b) source de Gato Cave, (c) source du Lez, (d) source de Qachqouch, et (e) sources d'Unica.

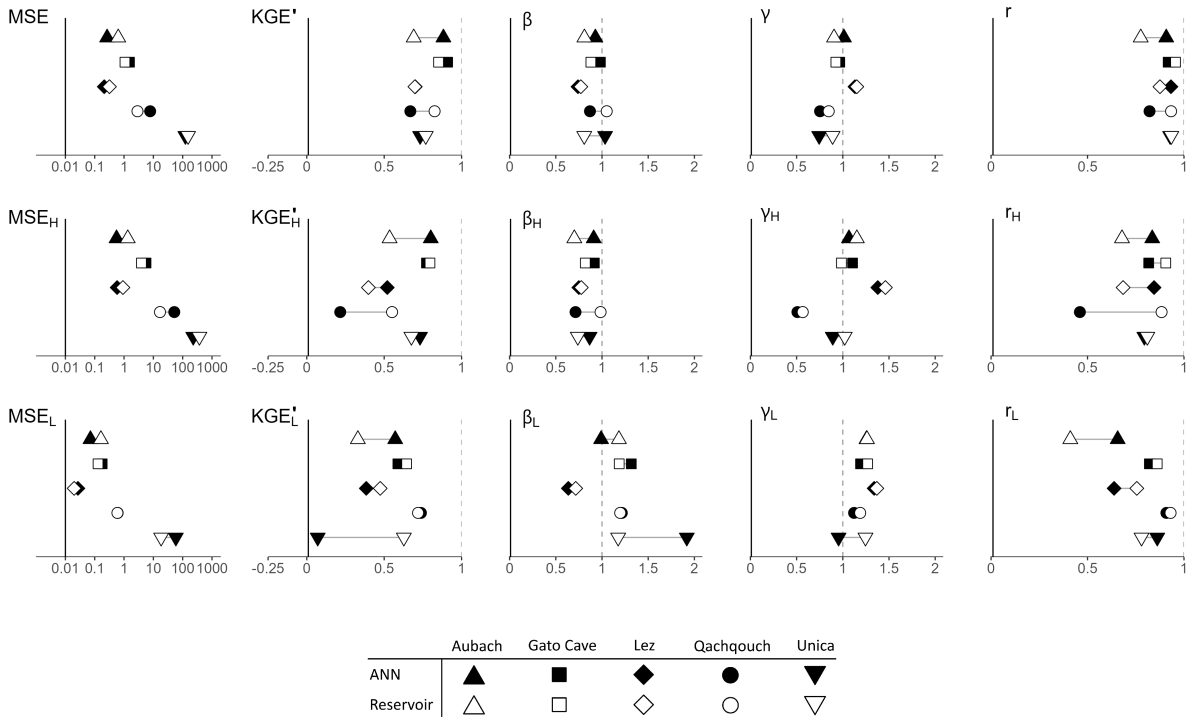


Figure 5: Performance des modèles ANN et réservoir sur la période de validation, selon différents critères de performance : erreur moyenne quadratique (MSE), le *Kling-Gupta Efficiency* modifié (KGE', Kling et al., 2012) et sa décompisition sous les différents aspects de volume (β), variabilité (γ) et corrélation (r). Chaque critère de performance est appliqué sur la chronique entière, mais aussi sur les périodes de basses (L) et hautes eaux (H) uniquement.

Conclusions

Cette thèse vise à développer et utiliser les méthodes numériques et les modèles hydrologiques pour caractériser le fonctionnement hydrologique des systèmes karstiques en réponse aux changements climatiques et anthropiques, en mettant l'accent sur la modélisation réservoir conceptuelle.

Moyens. Ce travail repose sur différentes bases de données et approches :

i. Les bases de données SNO KARST (Jourde et al., 2018) et World Karst Spring (WoKaS, Olarinoye et al., 2020) ont été utilisées pour l'analyse des séries chronologiques de débit des sources (Chapitre 2, Chapitre 3 et Chapitre 4).

ii. Des méthodes numériques telles que les analyses des courbes de récession, ainsi que les analyses statistiques et de signal ont été utilisées pour la classification du fonctionnement hydrologique des systèmes karstiques (Chapitre 2 et Chapitre 3).

iii. Le cadre du projet européen KARMA a permis de travailler sur plusieurs systèmes karstiques à travers la région méditerranéenne et de bénéficier des connaissances expertes locales (Chapitre 5 et Chapitre 6).

iv. Les approches à paramètres globaux, en particulier les modèles réservoir conceptuels, ont été utilisés pour approfondir la compréhension des dynamiques internes des systèmes karstiques (Chapitre 6).

Questions de recherche. Les méthodes numériques et les modèles hydrologiques ont été utilisés pour répondre aux questions de recherche initiales :

(i) *Comment caractériser le fonctionnement d'un système karstique dans un contexte de manque de données ?*

Un contexte de manque de données implique l'étude d'un système avec des données limitées, que ce soit en termes de couverture spatiale, de portée temporelle, de résolution, ou de diversité. Le Chapitre 2 montre les informations sur le fonctionnement d'un système qui peuvent être déduites de l'analyse des séries chronologiques de débit d'une source. L'analyse des courbes de récession permet une évaluation pertinente de différents aspects du fonctionnement hydrologique d'un système karstique : capacité de stockage dynamique, dynamique de drainage de la fonction capacitive et variabilité du fonctionnement hydrologique. La considération de la variabilité du fonctionnement hydrologique dans la classification proposée est un aspect nouveau qui permet une meilleure différenciation des systèmes karstiques. L'utilisation d'analyses multivariées et l'application de la classification sur de larges bases de données a permis d'identifier un nombre de classes de fonctionnement idéal – six étant un bon compromis entre le gain de connaissance, la pertinence, et l'accessibilité de la méthode. Le Chapitre 2 montre que cette caractérisation est pertinente compte tenu d'un nombre minimal d'observations dépendant des dynamiques du système et des caractéristiques du climat, restant généralement compatible avec les contextes de manque de données. Le Chapitre 6 montre que les modèles à paramètres globaux peuvent fournir des résultats intéressants avec peu de données en entrée (précipitations et températures unidimensionnelles) et avec de courtes séries chronologiques. L'utilisation des modèles réservoir peut être bénéfique dans un contexte de manque de

données ou pour réaliser une caractérisation préliminaire du fonctionnement d'un système karstique peu connu (e.g. répartition des volumes entre matrice et conduits, dynamique des flux internes).

(ii) *Sur quels aspects peut-on encore améliorer les modèles simples, unidimensionnels ?*

Les modèles unidimensionnels peuvent être améliorés sur plusieurs aspects : données d'entrée, calibration et évaluation. Le Chapitre 4 propose plusieurs ajouts à la plateforme de modélisation KarstMod sur ces aspects. Dans la nouvelle version 3.0 du logiciel, la définition des paramètres pour le prétraitement des données d'entrée peut être réalisée par calibration du modèle en utilisant des modules optionnels (neige ou évapotranspiration), permettant l'utilisation de données plus adéquates. La fonction objectif supporte maintenant plus de deux éléments, permettent ainsi de calibrer les modèles avec plusieurs variables d'observation (débit de la source, piézométrie, débit de source de trop-plein) et avec un ou plusieurs critères de performance, comme démontré pour le cas d'étude du Lez. Des outils supplémentaires ont été implémentés pour l'évaluation des modèles, qui peut être réalisée sur différents aspects du fonctionnement d'un système : bonne reproduction des volumes, de la variabilité, et de la corrélation. Le Chapitre 5 souligne l'importance de choisir des critères de performance adaptés.

(iii) *Peut-on se fier aux critères de performance pour la calibration et l'évaluation des modèles hydrologiques ?*

Le Chapitre 5 démontre l'impact des erreurs de compensation lors de l'utilisation de certain critères de performance pour la calibration et l'évaluation d'un modèle hydrologique. Les critères de performance ne sont pas entièrement fiables car chacun présente des limites spécifiques. Ils doivent être choisis avec précaution en fonction de l'objectif du modèle, et de préférence dans un cadre multi-critères.

(iv) *Quels sont les avantages et les inconvénients des différentes approches de modélisation empiriques et conceptuelles en hydrologie karstique ?*

Le Chapitre 6 propose une comparaison des approches de modélisation réservoir et à réseaux de neurones artificiels (ANN). Les modèles ANN, issus de l'apprentissage automatique, montrent une grande flexibilité par rapport aux données d'entrée et sont particulièrement performants pour reproduire les forts débits. Les modèles réservoir peuvent opérer avec des séries temporelles relativement courtes et semblent reproduire avec précision les faibles débits. Ces deux approches peuvent aider à caractériser différents aspects d'un système : (i) les modèles ANN en ce qui concerne la pertinence des données d'entrée et la délimitation du bassin versant, et (ii) les modèles réservoir en ce qui concerne le fonctionnement interne de l'aquifère.

Résultats clés. Les points clés de ce travail sont les suivants :

i. Malgré la forte hétérogénéité des systèmes karstiques, une caractérisation préliminaire de leur fonctionnement hydrologique est possible même dans un contexte de manque de données. L'analyse des séries chronologiques de débit ou l'utilisation de modèles à paramètres globaux unidimensionnels peuvent apporter de nombreuses informations sur les caractéristiques du système, son fonctionnement interne et sa réponse à diverses perturbations anthropiques ou climatiques, moyennant peu de données et de moyens.

ii. Les modèles hydrologiques doivent être appliqués avec discernement pour que l'interprétation soit pertinente. Chaque étape d'un processus de modélisation nécessite une étude approfondie des options disponibles, que ce soit pour le choix de l'approche de modélisation, la définition de la structure du modèle, l'utilisation des données d'entrée, la procédure de calibration, la fonction objectif, les critères de performance pour l'évaluation des résultats, ou l'analyse de sensibilité des paramètres. En particulier, cette thèse montre que tous les critères de performance ont des limites et doivent être sélectionnés en fonction de l'objectif du modèle hydrologique et de l'étude.

Perspectives. Les recherches futures peuvent porter sur :

i. L'impact du changement climatique sur les systèmes karstiques en Méditerranée à partir d'une même méthodologie appliquée à plusieurs sites ayant des caractéristiques différentes en termes de localisation, conditions climatiques, propriétés hydrogéologiques et disponibilité des données. Un travail en cours (commencé récemment dans le cadre de cette thèse) tire profit du cadre du projet KARMA pour étudier les tendances futures à large échelle de la dynamique et de la disponibilité des ressources en eau des aquifères karstiques, en utilisant des modèles réservoir. Pour tenir compte des incertitudes des projections climatiques, 8 modèles de climat couplés GCM/RCM sont considérés avec deux scénarios d'émission (RCP 4.5 et RCP 8.5). Une telle étude pourrait apporter des connaissances pertinentes sur la réponse des systèmes karstiques aux changements des précipitations et sous des conditions marquées de réchauffement et de sécheresse.

ii. L'éventuel lien entre la classe de fonctionnement hydrologique d'un système karstique et la structure du modèle réservoir considérée pour la simulation du débit à la source ou du niveau piézométrique. L'identification d'élements structuraux (compartiments, fonctions de transfert) spécifiques à certaines classes pourrait simplifier la tâche de modélisation.

iii. L'utilisation d'indicateurs de fonctionnement hydrologique pour la calibration des modèles (par exemple, les indicateurs issus de l'analyse des courbes de récession). Contraindre un modèle pour qu'il reproduise les aspects clés du fonctionnement d'un système (capacité de stockage dynamique, dynamique de drainage de la fonction capacitive et variabilité du fonctionnement hydrologique) pourrait amener à une structure et un paramétrage plus pertinents.

iv. La proposition d'un indicateur pour évaluer l'ampleur des erreurs de compensation d'un modèle, qui pourrait aider à renforcer l'évaluation de la pertinence des résultats. Une autre perspective serait de développer un critère de performance alternatif non sujet aux erreurs de compensation pour éviter des sous-estimations et surestimations simultanées trop prononcées dans les simulations.

v. Le couplage des modèles ANN et réservoir pour bénéficier des avantages de chaque approche. Une utilisation possible serait d'utiliser les modèles ANN pour contourner les limitations des modèles réservoir issues du traitement des données d'entrée. Cela permettrait de bénéficier des informations relatives à la structure et aux paramètres issus de la modélisation réservoir.

Références pour le résumé étendu

Bakalowicz, M. (2015). Karst and karst groundwater resources in the Mediterranean. *Environ Earth Sci*, *74*(1), 5–14. https://doi.org/10.1007/s12665-015-4239-4 (cit. on p. iii).

Baudement, C., Arfib, B., Mazzilli, N., Jouves, J., Lamarque, T., & Guglielmi, Y. (2017). Groundwater management of a highly dynamic karst by assessing baseflow and quickflow with a rainfall-discharge model (Dardennes springs, SE France). *Bull. Soc. géol. Fr.*, *188*(6), 40. https://doi.org/10.1051/bsgf/2017203 (cit. on p. iii).

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., & Goldscheider, N. (2017). The World Karst Aquifer Mapping project: Concept, mapping procedure and map of Europe. *Hydrogeol J*, *25*(3), 771–785. https://doi.org/10.1007/s10040-016-1519-3 (cit. on p. iii).

Dubois, E., Doummar, J., Pistre, S., & Larocque, M. (2020). Calibration of a lumped karst system model and application to the Qachqouch karst spring (Lebanon) under climate change conditions. *Hydrol. Earth Syst. Sci.*, *24*(9), 4275–4290. https://doi.org/10.5194/hess-24-4275-2020 (cit. on p. iii).

Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., & Dörfliger, N. (2009). Modelling the hydrologic functions of a karst aquifer under active water management – The Lez spring. *J. Hydrol.*, *365*(3-4), 235–243. https://doi.org/10.1016/j.jhydrol.2008.11.037 (cit. on p. iii).

Fleury, P., Plagnes, V., & Bakalowicz, M. (2007). Modelling of the functioning of karst aquifers with a reservoir model: Application to Fontaine de Vaucluse (South of France). *J. Hydrol.*, *345*(1), 38–49. https://doi.org/10.1016/j.jhydrol.2007.07.014 (cit. on p. iii).

Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., & Veni, G. (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeol J*, *28*(5), 1661–1677. https://doi.org/10.1007/s10040-020-02139-5 (cit. on p. iii).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003 (cit. on p. vi).

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.*, *52*(3), 218–242. https://doi.org/10.1002/2013RG000443 (cit. on pp. iii, iv).

Hartmann, A., Lange, J., Aguado, A. V., Mizyed, N., Smiatek, G., & Kunstmann, H. (2012). A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *J. Hydrol.*, *468–469*, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024 (cit. on p. iii).

Husic, A., Fox, J., Adams, E., Ford, W., Agouridis, C., Currens, J., & Backus, J. (2019). Nitrate Pathways, Processes, and Timing in an Agricultural Karst System: Development and Application of a Numerical Model. *Water Resour. Res.*, *55*(3), 2079–2103. https://doi.org/10.1029/2018WR023703 (cit. on p. iii).

Jourde, H., Massei, N., Mazzilli, N., Binet, S., Batiot-Guilhe, C., Labat, D., Steinmann, M., Bailly-Comte, V., Seidel, J., Arfib, B., Charlier, J., Guinot, V., Jardani, A.,

Fournier, M., Aliouache, M., Babic, M., Bertrand, C., Brunet, P., Boyer, J., ... Wang, X. (2018). SNO KARST: A French Network of Observatories for the Multidisciplinary Study of Critical Zone Processes in Karst Watersheds and Aquifers. *Vadose Zone J.*, *17*(1), 180094. https://doi.org/10.2136/vzj2018.04.0094 (cit. on pp. iv, v, x).

Kaufmann, G., & Romanov, D. (2020). Modelling long-term and short-term evolution of karst in vicinity of tunnels. *Journal of Hydrology*, *581*, 124282. https://doi.org/10.1016/j.jhydrol.2019.124282 (cit. on p. iii).

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011 (cit. on pp. vi, ix).

Kong-A-Siou, L., Johannet, A., Borrell, V., & Pistre, S. (2011). Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *J. Hydrol.*, *403*(3), 367–380. https://doi.org/10.1016/j.jhydrol.2011.04.015 (cit. on p. iii).

Lee, J. S., & Choi, H. I. (2022). A rebalanced performance criterion for hydrological model calibration. *J. Hydrol.*, *606*, 127372. https://doi.org/10.1016/j.jhydrol.2021.127372 (cit. on p. vi).

Liedl, R., Sauter, M., Hückinghaus, D., Clemens, T., & Teutsch, G. (2003). Simulation of the development of karst aquifers using a coupled continuum pipe flow model: DEVELOPMENT OF KARST AQUIFERS. *Water Resour. Res.*, *39*(3). https://doi.org/10.1029/2001WR001206 (cit. on p. iii).

Liu, D. (2020). A rational performance criterion for hydrological model. *J. Hydrol.*, *590*, 125488. https://doi.org/10.1016/j.jhydrol.2020.125488 (cit. on p. vi).

Mangin, A. (1975). *Contribution à l'étude hydrodynamique des aquifères karstiques* [Doctoral dissertation, Université de Dijon]. (Cit. on p. vi).

Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Soglio, L. D., & Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.*, *122*, 103927. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on p. v).

Mazzilli, N., Sivelle, V., Cinkus, G., Jourde, H., & Bertin, D. (2023). *KarstMod User Guide - version 3.0*. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on p. v).

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, *10*(3), 282–290 (cit. on p. vi).

Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., Allocca, V., Andreo, B., Apaéstegui, J., Apolit, C., Arfib, B., Auler, A., Bailly-Comte, V., Barberá, J. A., Batiot-Guilhe, C., Bechtel, T., Binet, S., Bittner, D., Blatnik, M., Bolger, T., ... Hartmann, A. (2020). Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater. *Sci. Data*, *7*(1), 59. https://doi.org/10.1038/s41597-019-0346-5 (cit. on pp. iv, x).

Parise, M., Gabrovsek, F., Kaufmann, G., & Ravbar, N. (2018). Recent advances in karst research: From theory to fieldwork and applications. *SP*, *466*(1), 1–24. https://doi.org/10.1144/SP466.26 (cit. on p. iii).

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.*, *63*(13-14), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002 (cit. on p. vi).

Schwemmle, R., Demand, D., & Weiler, M. (2021). Technical note: Diagnostic efficiency – specific evaluation of model performance. *Hydrol. Earth Syst. Sci.*, *25*(4), 2187–2198. https://doi.org/10.5194/hess-25-2187-2021 (cit. on p. vi).

Sivelle, V., Jourde, H., Bittner, D., Mazzilli, N., & Tramblay, Y. (2021). Assessment of the relative impacts of climate changes and anthropogenic forcing on spring discharge of a Mediterranean karst system. *J. Hydrol.*, *598*, 126396. https://doi.org/10.1016/j.jhydrol.2021.126396 (cit. on p. iii).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on p. iii).

Sullivan, T., Gao, Y., & Reimann, T. (2019). Nitrate transport in a karst aquifer: Numerical model development and source evaluation. *Journal of Hydrology*, *573*, 432–448. https://doi.org/10.1016/j.jhydrol.2019.03.078 (cit. on p. iii).

Tang, G., Clark, M. P., & Papalexiou, S. M. (2021). SC-Earth: A Station-Based Serially Complete Earth Dataset from 1950 to 2019. *J. Clim.*, *34*(16), 6493–6511. https://doi.org/10.1175/JCLI-D-21-0067.1 (cit. on p. vi).

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, *90*(C5), 8995. https://doi.org/10.1029/JC090iC05p08995 (cit. on p. vi).

Wu, Q., Xu, H., & Pang, W. (2008). GIS and ANN coupling model: An innovative approach to evaluate vulnerability of karst water inrush in coalmines of north China. *Environ. Geol.*, *54*(5), 937–943. https://doi.org/10.1007/s00254-007-0887-3 (cit. on p. iii).

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., & Goldscheider, N. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrol. Earth Syst. Sci.*, *26*(9), 2405–2430. https://doi.org/10.5194/hess-26-2405-2022 (cit. on p. iii).

Xanke, J., Jourde, H., Liesch, T., & Goldscheider, N. (2016). Numerical long-term assessment of managed aquifer recharge from a reservoir into a karst aquifer in Jordan. *Journal of Hydrology*, *540*, 603–614. https://doi.org/10.1016/j.jhydrol.2016.06.058 (cit. on p. iii).

Introduction

Scientific context

Karst carbonate formations outcrop on approximately 15.2 % of the global ice-free continental surface (Goldscheider et al., 2020) and are extensively represented in the subsurface (Chen et al., 2017). They are essential for providing drinking water, supporting agriculture and maintaining ecosystems, especially in the Mediterranean where 15 % of the domestic water (Bakalowicz, 2015) and up to 90 % in some regions originate from karst aquifers (Stevanović, 2019). These systems generally exhibit a high degree of heterogeneity and involve complex processes that govern water flow and storage, which make them challenging to study and particularly vulnerable to environmental changes (Parise et al., 2018).

Aim of this work

Climate change, increase in anthropogenic influence, as well as shifts in precipitation and evapotranspiration patterns could have a significant impact on the recharge of aquifers. The heterogeneous properties of karst systems and the non-linearity of their internal processes make them particularly challenging to study – further exploration of their response to these major changes is thus needed. Hydrological methods and numerical models constitute valuable tools for such investigations (Hartmann et al., 2014) and are widely used for different purposes in karst environments: (i) simulating spring discharge (Fleury et al., 2007; Wunsch et al., 2022), (ii) predicting and forecasting water flood/inrush (Kong-A-Siou et al., 2011; Wu et al., 2008), (iii) managing the exploitation of aquifers (Fleury et al., 2009; Xanke et al., 2016), (iv) characterising specific and internal functioning (Baudement et al., 2017; Dubois et al., 2020), (v) investigating the system's genesis and evolution (Kaufmann and Romanov, 2020; Liedl et al., 2003), (vi) assessing water quality and vulnerability (Husic et al., 2019; Sullivan et al., 2019), and (vii) studying the impact of climate change on water resources (Hartmann et al., 2012; Sivelle et al., 2021). In particular, lumped parameter models permit the study of complex and heterogeneous karst systems without requiring extensive meteorological and system-related data with high spatial resolution. They are the central aspect of this thesis, which aims to develop and improve hydrological methods and reservoir models for characterising the functioning of karst systems in response to climatic and anthropogenic changes. A focus is made on the following research questions:

i. How to characterise the functioning of a karst system in data-scarce contexts?

ii. On which aspects can simple, one-dimensional models be further improved?

iii. Can performance criteria be trusted for the calibration and evaluation of hydrological models?

iv. What are the advantages and drawbacks of different lumped parameter modelling approaches in karst hydrology?

Structure of this thesis

This document is structured as follows.

Chapter 1 provides a brief description of the formation, functioning, and particularities of karst systems. Rainfall-runoff modelling in karst environments is introduced, with an emphasis on lumped parameter models.

The first part is dedicated to the characterisation of karst systems hydrological functioning.

Chapter 2 proposes a new classification of karst systems hydrological functioning based on the analysis of the discharge at the spring. The classification relies on the analysis of recession curves and allows to characterise a karst system using three indicators reflecting the capacity of dynamic storage, the drainage dynamic of the capacitive function, and the variability of hydrological functioning. The method is developed to be also adapted for data-scarce contexts and is tested on a database of 78 karst systems located worldwide.

Chapter 3 presents a toolbox developed within the scope of this thesis. KarstID helps users in analysing karst spring discharge time series and characterising the hydrological functioning of karst systems using the classification outlined in Chapter 2. The software supports statistical, recession curve, classified discharge, and signal (correlational and spectral) analyses. The completion of these analyses is facilitated through a graphical interface, which require no programming skills.

The second part is devoted to the improvement and evaluation of the performance of lumped parameter models for the simulation of spring discharge and water level of karst systems.

Chapter 4 presents the new features introduced in the KarstMod modelling platform for the update to version 3.0. This work focused on two major aspects: (i) the use of relevant input data through optional modules (snow and potential evapotranspiration routines), and (ii) the proposition of appropriate objective functions and performance criteria for model calibration and evaluation. The new version is applied on two case studies in France (the Touvre and Lez karst systems).

Chapter 5 investigates the mechanism of counterbalancing errors on the Kling-Gupta Efficiency (KGE) and its variants, which are extensively used to evaluate hydrological models. In total, nine performance criteria and their decomposition are compared on synthetic time series and a real case study. The study demonstrates that, assessing a simulation, concurrent over- and underestimation of discharge can lead to an overall higher criterion score without being associated to an increase in model relevance.

Chapter 6 presents a comparison of two widely used lumped parameter modelling approaches in karst hydrology: artificial neural networks (ANN) and reservoir models. Numerous modelling approaches can be used for studying karst water resources, which can make it difficult for a stakeholder or researcher to choose the appropriate method. Five karst systems with different characteristics in terms of climatic conditions, hydrological properties and data availability are investigated. The study presents the advantages and drawbacks of each modelling approaches, as well as the common uncertainties faced in karst hydrological modelling.

Doctoral research – Timeline and contributions

This thesis started in October 2020, following a six-month Master's internship on the same topic, and took place within the framework of the European project KARMA (Karst Aquifer Resources availability and quality in the Mediterranean Area). The project aims to achieve significant progress in the hydrogeological understanding and sustainable management of karst groundwater resources in the Mediterranean area, in terms of both water availability and quality. This context provided the opportunity to access a large amount of data from the various partners and led to numerous interactions with hydrogeologists and modellers from the different universities involved in the project.

Throughout the duration of the thesis, several contributions have been made to the project, mainly in the form of deliverables:

i. Deliverable 4.1: Report of typology of the karst system (Cinkus et al., 2020)

ii. Deliverable 4.2: Application of lumped parameter modelling at the KARMA test sites (Cinkus et al., 2021c)

iii. Deliverable 4.4: Release of modelling tools: Lumped parameter modelling (Cinkus et al., 2021d)

iv. Policy Brief: Reservoir modelling (Cinkus et al., 2022b)

v. Deliverable 2.2: Recharge evaluation (Andreo et al., 2021)

vi. Deliverable 2.6: Spring discharge monitoring (Barberá et al., 2021)

vii. Deliverable 2.7: Uncertainties in water budget (Petitta et al., 2022)

viii. Deliverable 2.8: Water Availability (Petitta et al., 2023)

ix. Contribution to the KARMA Final report

During the three years of the thesis, 158 hours of teaching were realised and two practical course subjects were entirely developed from the outset for a Master's program. Six articles were produced: four as first author (Cinkus et al., 2021a; Cinkus et al., 2023d; Cinkus et al., 2023a; Cinkus et al., 2023b) and two as co-author (Sivelle et al., 2023; Wunsch et al., 2022). A total of nine oral presentations and one poster were presented during progress meetings, workshop or international conferences (Table 1).

Date	Event	City	Topic
2020-10	KARMA Progress meeting	Montpellier (FR)	Oral – Chapter 2
2021-03	KARMA Progress meeting	Rome (IT)	Oral – Chapter 6
2021-04	EGU21	Online	Oral[1] – Chapter 2
2021-06	SNO Karst Workshop	Nîmes (FR)	Oral – Chapter 3
2021-09	KARMA Progress meeting	Beirut (LB)	Oral – Chapter 6
2022-03	KARMA Progress meeting	Tunis (TN)	Oral – Chapter 6
2022-05	EGU22	Vienna (AT)	Oral[2] – Chapter 5
2022-06	SNO Karst Workshop	Saint-Martin-le-Nœud (FR)	Oral – Chapter 3
2022-06	Eurokarst 2022	Málaga (ES)	Poster[3] – Chapter 3
2023-04	EGU23	Vienna (AT)	Oral[4]

[1] Cinkus et al., 2021b
[2] Cinkus et al., 2022c
[3] Cinkus et al., 2022a
[4] Cinkus et al., 2023c

Table 1: Oral presentations and poster realised during the thesis.

References for Introduction

Andreo, B., Barberá, J. A., Chen, Z., Cinkus, G., Doummar, J., Fharmeier, N., Fernández-Ortega, J., Goldscheider, N., Jourde, H., Lorenzi, V., Ana Isabel, M. G., Martín Rodríguez, J. F., Othman, J., Petitta, M., & Sbarbati, C. (2021). *Deliverable 2.2: Recharge evaluation*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/D22_Recharge_evaluation (cit. on p. xix).

Bakalowicz, M. (2015). Karst and karst groundwater resources in the Mediterranean. *Environ Earth Sci*, *74*(1), 5–14. https://doi.org/10.1007/s12665-015-4239-4 (cit. on p. xvii).

Barberá, J. A., Batiot-Guilhe, C., Frank, S., Goeppert, N., Fharmeier, N., Goldscheider, N., Rakowski, S., Jourde, H., Andreo, B., Fernández-Ortega, J., Slama, F., Gargouri-Ellouze, E., Bouhlila, R., Brunet, P., Cinkus, G., Sivelle, V., Ulloa-Cedamanos, F., Petitta, M., Lorenzi, V., ... Doummar, J. (2021). *Deliverable 2.6: Spring discharge monitoring*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/D26_Spring_discharge_monitoring_DEF (cit. on p. xix).

Baudement, C., Arfib, B., Mazzilli, N., Jouves, J., Lamarque, T., & Guglielmi, Y. (2017). Groundwater management of a highly dynamic karst by assessing baseflow and quickflow with a rainfall-discharge model (Dardennes springs, SE France). *Bull. Soc. géol. Fr.*, *188*(6), 40. https://doi.org/10.1051/bsgf/2017203 (cit. on p. xvii).

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., & Goldscheider, N. (2017). The World Karst Aquifer Mapping project: Concept, mapping procedure and map of Europe. *Hydrogeol J*, *25*(3), 771–785. https://doi.org/10.1007/s10040-016-1519-3 (cit. on p. xvii).

Cinkus, G., Mazzilli, N., & Jourde, H. (2020). *Deliverable 4.1: Report of typology of the karst systems*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/D41_Report_of_typology_of_karst_systems (cit. on p. xix).

Cinkus, G., Mazzilli, N., & Jourde, H. (2021a). Identification of relevant indicators for the assessment of karst systems hydrological functioning: Proposal of a new classification. *J. Hydrol.*, *603*, 127006. https://doi.org/10.1016/j.jhydrol.2021.127006 (cit. on p. xix).

Cinkus, G., Mazzilli, N., & Jourde, H. (2021b). *Proposal of a typology of karst systems functioning based on relevant indicators of karst springs hydrodynamics* (Oral presentation). vEGU21, the 23rd EGU General Assembly. online. https://doi.org/10.5194/egusphere-egu21-4009 (cit. on p. xx).

Cinkus, G., Mazzilli, N., & Jourde, H. (2022a). *KarstID: An interactive R application for karst spring hydrograph analysis that also provides a classification of karst hydrological functioning* (Poster). Eurokarst 2022 – The European Conference on Karst Hydrogeology and Carbonate Reservoirs. Málaga, Spain. (Cit. on p. xx).

Cinkus, G., Mazzilli, N., & Jourde, H. (2022b). *Policy Brief: Reservoir modelling*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Policy_briefs/KARMA_Policy_Brief_Reservoir_modeling.pdf (cit. on p. xix).

Cinkus, G., Mazzilli, N., & Jourde, H. (2023a). KarstID: An R Shiny application for the analysis of karst spring discharge time series and the classification of karst system hydrological functioning. *Environ Earth Sci*, *82*(136), 6. https://doi.org/10.1007/s12665-023-10830-5 (cit. on p. xix).

Cinkus, G., Mazzilli, N., Jourde, H., Batiot-Guilhe, C., Wunsch, A., Liesch, T., Chen, Z., Petitta, M., Lorenzi, V., Sbarbati, C., Barberá, J. A., Andreo, B., Ravbar, N., Doummar, J., Bouhlila, R., Gargouri-Ellouze, E., Slama, F., & Goldscheider, N. (2021c). *Deliverable 4.2: Application of lumped parameter modelling at the KARMA test sites*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/D42_Application_of_lumped_parameter_modelling_at_the_KARMA_test_sites.pdf (cit. on p. xix).

Cinkus, G., Mazzilli, N., Jourde, H., & Chen, Z. (2021d). *Deliverable 4.4: Release of modelling tools: Lumped parameter modelling*. Retrieved June 14, 2023, from https://github.com/HSM-hydro/KARMA_D44 (cit. on p. xix).

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Goldscheider, N., Ravbar, N., Fernández-Ortega, J., Barberá, J. A., Andreo, B., & Chen, Z. (2022c). *Possible bias in the assessment of karst hydrological model performance. Example of alpha and beta parameters compensation when using the KGE as performance criterion.* (Oral presentation). EGU22, the 24th EGU General Assembly. Vienna, Austria. https://doi.org/10.5194/egusphere-egu22-5149 (cit. on p. xx).

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., & Goldscheider, N. (2023b). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrol. Earth Syst. Sci.*, *27*(13), 2397–2411. https://doi.org/10.5194/hess-27-2397-2023 (cit. on p. xix).

Cinkus, G., Sivelle, V., Jourde, H., Mazzilli, N., Tramblay, Y., Andreo, B., Barberá, J. A., Bouhlila, R., Doummar, J., Fernández-Ortega, J., Gargouri-Ellouze, E., Lorenzi, V., Petitta, M., Ravbar, N., Slama, F., & Goldscheider, N. (2023c). *Impact of climate change on groundwater level dynamics and karst spring discharge of several karst systems in the Mediterranean area* (Oral presentation). EGU22, the 25th EGU General Assembly. Vienna, Austria. https://doi.org/10.5194/egusphere-egu23-12285 (cit. on p. xx).

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J. A., Andreo, B., Goldscheider, N., & Jourde, H.

(2023d). Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions. *Hydrol. Earth Syst. Sci.*, *27*(10), 1961–1985. https://doi.org/10.5194/hess-27-1961-2023 (cit. on p. xix).

Dubois, E., Doummar, J., Pistre, S., & Larocque, M. (2020). Calibration of a lumped karst system model and application to the Qachqouch karst spring (Lebanon) under climate change conditions. *Hydrol. Earth Syst. Sci.*, *24*(9), 4275–4290. https://doi.org/10.5194/hess-24-4275-2020 (cit. on p. xvii).

Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., & Dörfliger, N. (2009). Modelling the hydrologic functions of a karst aquifer under active water management – The Lez spring. *J. Hydrol.*, *365*(3-4), 235–243. https://doi.org/10.1016/j.jhydrol.2008.11.037 (cit. on p. xvii).

Fleury, P., Plagnes, V., & Bakalowicz, M. (2007). Modelling of the functioning of karst aquifers with a reservoir model: Application to Fontaine de Vaucluse (South of France). *J. Hydrol.*, *345*(1), 38–49. https://doi.org/10.1016/j.jhydrol.2007.07.014 (cit. on p. xvii).

Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., & Veni, G. (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeol J*, *28*(5), 1661–1677. https://doi.org/10.1007/s10040-020-02139-5 (cit. on p. xvii).

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.*, *52*(3), 218–242. https://doi.org/10.1002/2013RG000443 (cit. on p. xvii).

Hartmann, A., Lange, J., Aguado, A. V., Mizyed, N., Smiatek, G., & Kunstmann, H. (2012). A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *J. Hydrol.*, *468–469*, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024 (cit. on p. xvii).

Husic, A., Fox, J., Adams, E., Ford, W., Agouridis, C., Currens, J., & Backus, J. (2019). Nitrate Pathways, Processes, and Timing in an Agricultural Karst System: Development and Application of a Numerical Model. *Water Resour. Res.*, *55*(3), 2079–2103. https://doi.org/10.1029/2018WR023703 (cit. on p. xvii).

Kaufmann, G., & Romanov, D. (2020). Modelling long-term and short-term evolution of karst in vicinity of tunnels. *Journal of Hydrology*, *581*, 124282. https://doi.org/10.1016/j.jhydrol.2019.124282 (cit. on p. xvii).

Kong-A-Siou, L., Johannet, A., Borrell, V., & Pistre, S. (2011). Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *J. Hydrol.*, *403*(3), 367–380. https://doi.org/10.1016/j.jhydrol.2011.04.015 (cit. on p. xvii).

Liedl, R., Sauter, M., Hückinghaus, D., Clemens, T., & Teutsch, G. (2003). Simulation of the development of karst aquifers using a coupled continuum pipe flow model: DEVELOPMENT OF KARST AQUIFERS. *Water Resour. Res.*, *39*(3). https://doi.org/10.1029/2001WR001206 (cit. on p. xvii).

Parise, M., Gabrovsek, F., Kaufmann, G., & Ravbar, N. (2018). Recent advances in karst research: From theory to fieldwork and applications. *SP*, *466*(1), 1–24. https://doi.org/10.1144/SP466.26 (cit. on p. xvii).

Petitta, M., Lorenzi, V., Andreo, B., Barberá, J. A., Hmid, A. B., Bouhlila, R., Cinkus, G., Gargouri-Ellouze, E., Doummar, J., Fernández-Ortega, J., Goldscheider, N.,

Jourde, H., Marín, A. I., Martín Rodríguez, J. F., Mazzilli, N., & Slama, F. (2023). *Deliverable 2.8: Water availability*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/Deliverable_28_Water_availability.pdf (cit. on p. xix).

Petitta, M., Lorenzi, V., Barbiero, D., Sbarbati, C., Al Ali, M., Andreo, B., Barberá, J. A., Bouhlila, R., Cinkus, G., Gargouri-Ellouze, E., Doummar, J., Fernández-Ortega, J., Goldscheider, N., Jourde, H., Martín Rodríguez, J. F., Mazzilli, N., Othman, J., & Slama, F. (2022). *Deliverable 2.7: Uncertainties in water budget*. Retrieved June 14, 2023, from http://karma-project.org/images/Documents/Deliverables/D27_Uncertainties_Water_Budget.pdf (cit. on p. xix).

Sivelle, V., Cinkus, G., Mazzilli, N., Labat, D., Arfib, B., Massei, N., Cousquer, Y., Bertin, D., & Jourde, H. (2023). Improvement of the KarstMod modeling platform for a better assessment of karst groundwater resources. *Hydrol. Earth Syst. Sci.*, 1–26. https://doi.org/10.5194/hess-2023-17 (cit. on p. xix).

Sivelle, V., Jourde, H., Bittner, D., Mazzilli, N., & Tramblay, Y. (2021). Assessment of the relative impacts of climate changes and anthropogenic forcing on spring discharge of a Mediterranean karst system. *J. Hydrol.*, *598*, 126396. https://doi.org/10.1016/j.jhydrol.2021.126396 (cit. on p. xvii).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on p. xvii).

Sullivan, T., Gao, Y., & Reimann, T. (2019). Nitrate transport in a karst aquifer: Numerical model development and source evaluation. *Journal of Hydrology*, *573*, 432–448. https://doi.org/10.1016/j.jhydrol.2019.03.078 (cit. on p. xvii).

Wu, Q., Xu, H., & Pang, W. (2008). GIS and ANN coupling model: An innovative approach to evaluate vulnerability of karst water inrush in coalmines of north China. *Environ. Geol.*, *54*(5), 937–943. https://doi.org/10.1007/s00254-007-0887-3 (cit. on p. xvii).

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., & Goldscheider, N. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrol. Earth Syst. Sci.*, *26*(9), 2405–2430. https://doi.org/10.5194/hess-26-2405-2022 (cit. on pp. xvii, xix).

Xanke, J., Jourde, H., Liesch, T., & Goldscheider, N. (2016). Numerical long-term assessment of managed aquifer recharge from a reservoir into a karst aquifer in Jordan. *Journal of Hydrology*, *540*, 603–614. https://doi.org/10.1016/j.jhydrol.2016.06.058 (cit. on p. xvii).

Chapter 1

Scientific context

Contents

This chapter provides the general scientific context of the thesis. Specific and extensive states of the art are further detailed in each Chapter. Section 1.1 gives an introduction to the characteristics, functioning and uniqueness of karst systems. Section 1.2 presents the main challenges in applying hydrological models to karst environments. The main modelling approaches are briefly introduced, highlighting their concepts, advantages and drawbacks, and main applications. Section 1.3 contrasts the scientific context against the goals and structure of the thesis.

1.1 Introduction to karst systems

1.1.1 Formation and characteristics

The term *karst* defines a landscape in a carbonate or evaporite environment, characterised by unique surface or underground formations resulting from the infiltration and flow of acidic waters (Ford and Williams, 2007). Karst is predominantly found in carbonate rocks (limestone, dolomite, and to a lesser extent, chalk and marble), but it can also be present in evaporite rocks (gypsum, halite, anhydrite) (Stevanović, 2015). From a hydrogeological perspective, carbonate rocks are by far the most significant when considering karst (Goldscheider et al., 2020). This type of environment is known for its high heterogeneity resulting from various processes that occur both during and after the rock's diagenesis and over geological time.

The primary heterogeneity of karst is linked to the environment and deposition conditions. The precipitation of carbonates is dependent on many biotic and abiotic factors that influence the characteristics of the deposit over time and space (Lowenstam and Weiner, 1989). Thus, the characteristics of the produced rock vary spatially on both lateral and vertical scales, according to the conditions and processes that constrained

the deposit, the diagenesis, and the simultaneous or subsequent alteration of the carbonates (Tendil, 2018).

Heterogeneity over geological time develops after the diagenesis of the rock. Unlike silicate rocks, karst is highly chemically reactive, and the phenomena of alteration of the parent rock occur over very short geological timescales (White and Blum, 1995). The infiltration and circulation of meteoric waters in the rock are unevenly distributed due to the already present heterogeneity; the preferential expansion of certain voids by dissolution thus favours flow within these same voids. This positive feedback mechanism is the origin of a hierarchy of flows within the karst, where different degrees of porosity (matrix, fractures, conduits) induce different flow velocities. These differences result in a complex, diversified morphological landscape, unique to karst (Figure 1.1).

The heterogeneity of a karst system thus depends on the rock formation conditions and alteration phenomena over time (Jouves et al., 2017). As a result, the evolution of primary properties of carbonates (e.g. mineralogical composition, porosity, permeability) in space and time contributes to the general and multi-scale complexity of karst systems (Tendil, 2018). This leads to a unique hydrodynamic functioning specific to each system; the presence of major karst drains favours rapid flows while the less altered matrix results in slower flows and water storage areas (Király, 1998). All the waters that are infiltrated and transported in the system generally converge towards one or several outlets referred to as *springs*, which can be either surficial or underwater.

1.1.2 Global distribution and importance

Karst carbonate formations outcrop on approximately 15.2 % of the global ice-free continental surface (Goldscheider et al., 2020), and are extensively represented in the subsurface (Chen et al., 2017). About two-thirds of carbonate outcrops are present across three continents: Europe (21.8 %), North America (19.6 %) and Asia (18.6 %), the latter having the largest surface area in absolute terms (Goldscheider et al., 2020). Goldscheider et al. (2020) also demonstrated that around 16.5 % of the global population lived in karst areas in 2015, with the most populated continents in relative proportion being Europe (25.3 %, 172.1 million) and North America (23.5 %, 134.2 million), and Asia in the highest absolute population (15.1 %, 661.7 million). Karst is highly represented in the Mediterranean basin – situated in the southeast region of Europe and north of Africa, centred on the Mediterranean Sea (Figure 1.2) – especially along the coastlines. Well-known examples include the Dinaric karst along the Adriatic coast, and the coastlines of Libya and Egypt (Goldscheider et al., 2020).

Carbonate aquifers play a crucial role in the Mediterranean region as they provide on average about 15 % of domestic water (Bakalowicz, 2015), and can supply up to 90 % in certain areas (Stevanović, 2019). They also contribue to surface waters and are important for irrigation and heating purposes (Stevanović, 2015), geochemical and carbon cycling (Chen et al., 2023; Liu et al., 2010), their unique and valuable ecosystems and biodiversity (Bonacci et al., 2009; Pipan and Culver, 2013), geotourism (Ravbar and Šebela, 2015; Ruban, 2018), palaeoclimatology (Vaks et al., 2018), and archaeological and cultural heritage (Gu et al., 2023; Williams, 2008).

However, due to their relatively rapid transfer times, karst aquifers are particularly vulnerable to pollution (Bakalowicz, 2015), especially when the rock is not undercover (Goldscheider, 2005). In coastal areas, the direct connection of karst systems with the sea makes them sensitive to seawater intrusion (Arfib et al., 2007; Pinault et al.,

Figure 1.1: Karst landscape (Unica catchment, Slovenia).

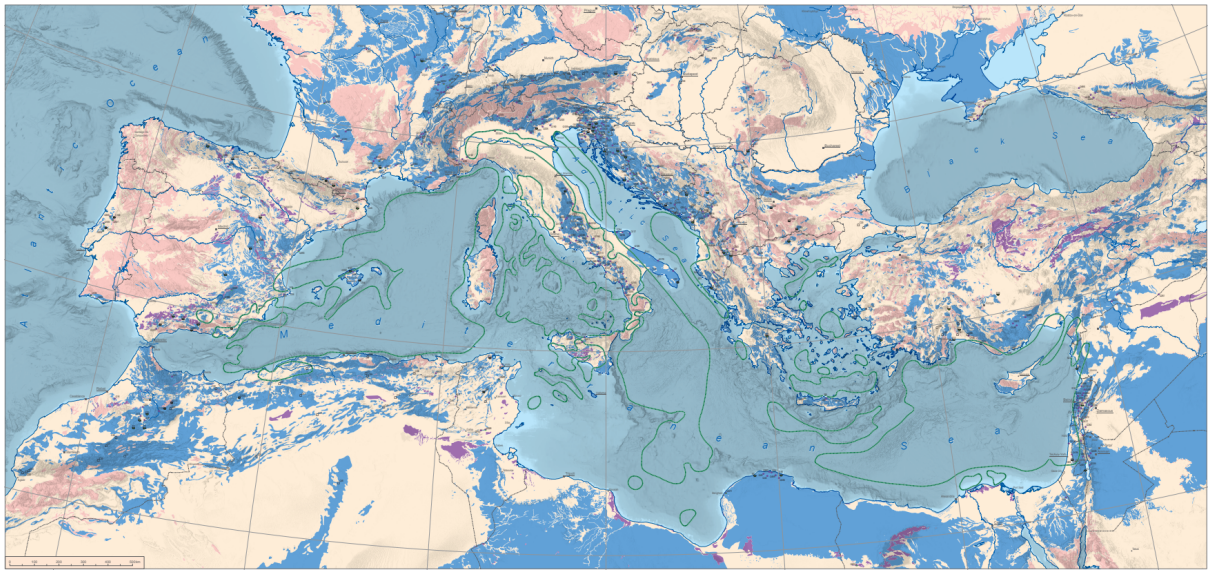


Figure 1.2: Map from MEDKAM (Xanke et al., 2022). Blue areas correspond to karst aquifers in sedimentary and metamorphic carbonate rocks. Purple areas represent karst aquifers in evaporite rocks. The detailed legend can be found on the online MEDKAM map.

2004) – a problem exacerbated by overexploitation in certain regions (Goldscheider et al., 2020). The unique structure of karst systems also makes them vulnerable to a variety of geohazards (De Waele et al., 2011; Gutiérrez et al., 2014). All these vulnerabilities are likely to intensify in the future amid climate disruptions and increasing human pressures (Parise et al., 2018; Stevanović, 2019). Furthermore, the Mediterranean basin is considered a *hot-spot* of climate change, where decreases in precipitation and increases in temperature (more severe than in the rest of the world) are expected (Lionello and Scarascia, 2018). Numerous studies have thus lately been dedicated to studying future karst water resources, under various scenarios of climate change and anthropogenic pressure (Dirnböck et al., 2016; Hartmann et al., 2012b; Nerantzaki and Nikolaidis, 2020; Sivelle et al., 2021).

1.2 Rainfall-runoff modelling in karst environments

1.2.1 Challenges in (karst) hydrological modelling

Hydrological models can provide very valuable insights into the functioning of hydrosystems, as well as predict and forecast flow conditions at short and long terms. However, they have a fair amount of limitations and improving models has been identified as a crucial area of research (Blöschl et al., 2019). Issues limiting the performance and relevance of hydrological models include:

i. Data-scarcity: some catchments are not sufficiently instrumented or have not been monitored for a long enough period to allow relevant flow modelling; this problem is particularly pronounced in karst environments (Hartmann et al., 2014).

ii. Reliability of discharge measurements: discharge time series are generally derived from water height, using water level–discharge calibration curves. Numerous uncertainties are associated with this determination method (Pelletier, 1988), including

extrapolation errors for extreme values (Di Baldassarre and Montanari, 2009; Moges et al., 2021).

iii. Input data: input data does not always accurately reflect the heterogeneity of processes occuring on a catchment. For instance, the spatial variability of precipitation can be very high, especially in areas where strong convective storms are frequent (Lobligeois et al., 2014).

iv. Model structure and parameters: models can be subject to over-parametrisation. Careful sensitivity analyses and uncertainty assessment should be considered along with model results to avoid over-interpretation (Beven, 2019; Refsgaard et al., 2007).

v. Model calibration and evaluation: models that are reliable in known conditions may be biased in other periods due to the non-stationarity of the climate (Vaze et al., 2010). The calibration period should encompass the diversity of hydrological observations and meteorological conditions, and be evaluated with relevant performance criteria assessing different aspects of the model (Clark et al., 2021; Gupta et al., 2009; Hartmann et al., 2017; Knoben et al., 2019; Seibert et al., 2018).

Hydrological models applied in karst environments must therefore consider and address these limitations, in addition to the constraints associated with the complexity and the characteristic heterogeneity of karst, which manifest at all levels:

i. Surface: the infiltration of precipitation into the system can be either diffuse or concentrated.

ii. Subsurface/unsaturated zone: varying degrees of porosity and permeability, preferential flow paths, local and temporary water storage, interaction with vegetation.

iii. Underground/saturated zone: dual flow regime with conduit and diffuse flow (through matrix).

Traditional hydrological models are often too simplistic to capture the complexity of karst functioning. The majority of the system is inaccessible to human exploration, and investigations and samplings only provide information about very small areas, at the borehole scale (Parise et al., 2018). This strong heterogeneity makes estimating the parameters of physical models very challenging (Hartmann et al., 2014), especially given that the location of conduits is unknown in most cases (Parise et al., 2018). Anthropogenic influence through pumping also poses a challenge in models as it complicates the interactions between the various compartments of the systems. In some cases, this even causes the drying up of the source (Hamed Ferjani et al., 2020; Maréchal et al., 2013), whose observations are a crucial aspect of hydrodynamic modelling. Despite these high contrasts in porosity and permeability, which can be challenging to evaluate, hydrological models still offer valuable insight into the functioning of karst systems (Parise et al., 2018), particularly in predicting the impact of climate and land use changes (Hartmann et al., 2014).

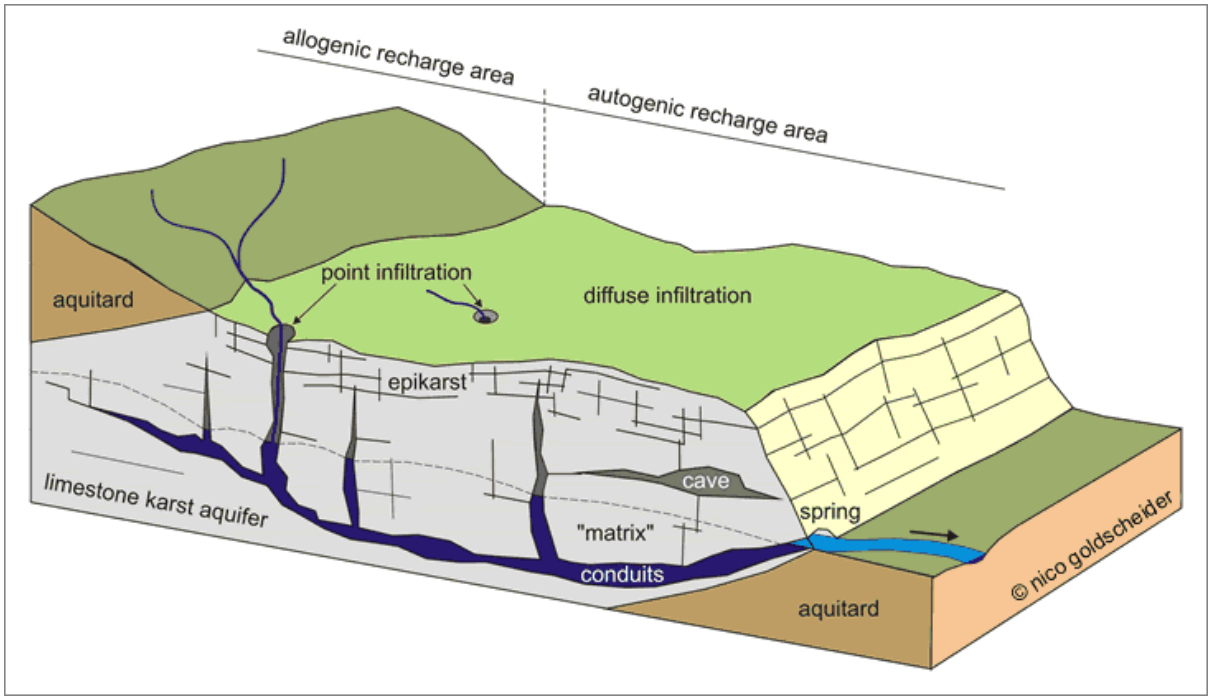


Figure 1.3: Block diagram of a heterogeneous karst aquifer illustrating the duality of recharge (allogenic vs. autogenic), infiltration (point vs. diffuse) and porosity/flow (conduits vs. matrix) (Goldscheider and Drew, 2007).

1.2.2 Modelling approaches

Hydrological models are numerical tools that aim to reproduce key variables of karst systems, such as aquifer piezometric levels or spring discharge. Precipitation and evapotranspiration are essential input variables for all hydrological models, but many other variables can be used to better represent or constrain the physical phenomena and internal processes of the system. Hydrological models can be categorised into lumped parameter, semi-distributed and distributed approaches (Hartmann, 2018; Kovács and Sauter, 2007). While distributed models partition a karst system into a two- or three-dimensional grid, where each cell is assigned suitable hydraulic parameters and system states, lumped parameter models rely on the mathematical analysis of input data (e.g. precipitation, temperature) to simulate spring discharge time series. Lumped parameter models encompass (i) "black-box" models, which do not use any a priori information about a system functioning, and (ii) "conceptual reservoir" models, which are based on a conceptual representation of a system. Some reservoir models are semi-distributed, i.e., their structure is repeated in space to take into account the spatial heterogeneity of land cover, precipitation, and hydrogeological properties. A wide range of these numerical modelling tools currently used in research and operation have been compared by Jeannin et al. (2021) based on the simulation results of the discharge of a karst spring. This section briefly presents the main modelling approaches used in karst hydrology: distributed, black-box, and conceptual reservoir models. Choosing a modelling approach primarily depends on the goal of the study, but is also constrained by the current understanding of the system, available data, and regional or institutional preferences (Addor and Melsen, 2019).

1.2.2.a Distributed models

Distributed, physically-based models typically discretise a karst system into a 2 or 3-dimensional grid. Each cell of the grid has its own physically-based parameters and system states which must be defined or calibrated. There are different approaches to physically-based models which are presented in more detail by Hartmann et al. (2014), Jeannin et al. (2021), Jourde and Wang (In press), and Mazzilli (2011):

i. Equivalent porous medium

ii. Dual continuum

iii. Discrete features (pipe network models)

iv. Combined Discrete-Continuum

These models are particularly of interest for (i) gaining insight into the spatio-temporal distributions of internal flows in a system (Gill et al., 2021; Kavousi et al., 2020), (ii) characterising the geometry and evolution of conduits (Fandel et al., 2021), and (iii) assessing the impacts of changes in boundary condition, system variables, or anthropogenic influence (Mazzilli, 2011). The primary limitation of this approach is related to the lack of knowledge about the physical properties of the system (Hartmann et al., 2014) – most models therefore rely on assumptions or propose multiple conduit and fracture networks. As a result, the calibration of a physically-based model can also be challenging, as there are a large number of parameters that may be subject to equifinality (Beven, 2006).

1.2.2.b Black-box models

Black-box models are empirical models transforming an input signal into an output signal without any consideration of the system structure or physical parameters. They consist in analytical transfer functions or machine learning methods, with the latter seeing significant growth in the past decades (Maier et al., 2010). Black-box approaches include, among others, linear and non-linear transfer functions (Jukić and Denić-Jukić, 2006; Labat et al., 1999), deep recurrent multilayer perceptron, NARX model, convolutional neural networks, long short-term memory networks (Jeannin et al., 2021). Both approaches can be very effective in simulating known functioning of a system, but may be limited in the case of unobserved conditions (e.g. extreme) or disruptive events such as climate change or anthropogenic pressures. Neural networks, in particular, require a learning period to develop the function and its parameters, which can make their use difficult in data-scarce contexts (Jeannin et al., 2021). However, they are quick, reliable, and flexible on input data – the models can quickly assimilate meteorological processes (evapotranspiration, snow accumulation and melting) without necessarily requiring any preprocessing of the input data (Kong-A-Siou et al., 2015). Although the acquisition of knowledge in terms of a system internal functioning remains limited (Jeannin et al., 2021), machine learning approaches can help determine the most relevant input, identify the major contributing areas of a catchment, or even assist in their delineation (Wunsch et al., 2022).

1.2.2.c Conceptual reservoir models

Conceptual reservoir models are a conceptual representation of a hydrosystem, which involves the association of several components (reservoirs) that are thought to be representative of the main processes at stake. Each reservoir is characterised by its water height and a flow equation that translates the variation of water height into discharge. The flow equation is generally function of a specific discharge coefficient and a positive exponent (different from 1 for non-linear flows), which are defined by calibration against observed data. Other flow equations exist such as the infinite characteristic time transfer function, which allows to account for long-term memory effects (Guinot et al., 2015). Many reservoir models have been developed to study the relation between precipitation and discharge in karst systems – some of them being semi-distributed (Bittner et al., 2018; Hartmann et al., 2012a; Hartmann et al., 2014; Jeannin et al., 2021; Ollivier et al., 2020; Sarrazin et al., 2018). They all differ in complexity with respect to the number of reservoirs and parameters, which need to be well-thought-out in order to preserve physical realism and limit equifinality on model parameters (Hartmann, 2018). Reservoir models can be seen as a compromise between simulation performance and insight into the functioning of a system (Mazzilli, 2011).

In strict one-dimensional models, input data generally need preprocessing and thus contribute to model uncertainty: (i) arbitrary decisions on the raw data (e.g. choosing precipitation from one meteorological station rather than another), (ii) interpolation (when data from several meteorological stations over a catchment are available), or (iii) preprocessing (e.g. snow processes, potential evapotranspiration). Even if it still remains a conceptualisation, the underlying structure of the model allows the characterisation of flows dynamics within a system. For example, it allows to assess the water partitioning between the different compartments, as well as internal flow processes (Baudement et al., 2017).

Based on this modelling approach, the French SNO KARST (Jourde et al., 2018) has developed the KarstMod platform, which provides a modular, user-friendly interface for simulating discharge and piezometry of karst systems (Mazzilli et al., 2019). The structure of models built using the KarstMod platform is based on the conceptual model of a karst aquifer with infiltration and saturated zones. The infiltration zone (soil and epikarst) drains water from the surface through a vertical network of fissures and conduits. Water storage can occur in the infiltration zone, as well as local saturation. The saturated zone comprises a dual porosity functioning with a network of high-permeability fractures and conduits, and a low-permeability matrix with a high storage capacity. The KarstMod model structure can include up to four reservoirs. One at the upper level reflects the processes occurring in the soil and epikarst zone (infiltration, storage and drainage). Three at the lower level can be connected with the first one and correspond to the infiltration and/or saturated zones. The discharge can be simulated with (i) several linear and non-linear water level-discharge laws, including a hysteretic water level-discharge function to reproduce the hysteretic functioning observed on the wet-dry cycles in the unsaturated zone (Lehmann et al., 1998; Tritz et al., 2011), (ii) an exchange function that aims to reproduce the interactions between the matrix and conduits, and (iii) an infinite characteristic time transfer function (Guinot et al., 2015). More details on the balance equations, the parameters involved and the KarstMod platform in general can be found in Mazzilli et al. (2019) or in the KarstMod User Guide (Mazzilli et al., 2023).

1.3 Conclusions

Lumped parameter models offer a good compromise between model accessibility and relevance. Their use in karst environments is intricate due to the characteristic heterogeneity of these systems, adding another challenge on top of the typical uncertainties and biases associated with hydrological modelling. Nevertheless, they provide relevant and beneficial insights in the characterisation and prediction of karst systems functioning. This thesis explores the options for (i) characterising the functioning of karst systems, (ii) improving the robustness of lumped parameter models, as well as (iii) exploring the information that can be derived from the analysis of simulations – even in a data-scarce context.

1.4 References for Chapter 1

Addor, N., & Melsen, L. A. (2019). Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. *Water Resources Research*, *55*(1), 378–390. https://doi.org/10.1029/2018WR022958 (cit. on p. 6).

Arfib, B., De Marsily, G., & Ganoulis, J. (2007). Locating the Zone of Saline Intrusion in a Coastal Karst Aquifer Using Springflow Data. *Ground Water*, *45*(1), 28–35. https://doi.org/10.1111/j.1745-6584.2006.00252.x (cit. on p. 2).

Bakalowicz, M. (2015). Karst and karst groundwater resources in the Mediterranean. *Environ Earth Sci*, *74*(1), 5–14. https://doi.org/10.1007/s12665-015-4239-4 (cit. on p. 2).

Baudement, C., Arfib, B., Mazzilli, N., Jouves, J., Lamarque, T., & Guglielmi, Y. (2017). Groundwater management of a highly dynamic karst by assessing baseflow and quickflow with a rainfall-discharge model (Dardennes springs, SE France). *Bull. Soc. géol. Fr.*, *188*(6), 40. https://doi.org/10.1051/bsgf/2017203 (cit. on p. 8).

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007 (cit. on p. 7).

Beven, K. (2019). How to make advances in hydrological modelling. *Hydrol. Res.*, *50*(6), 1481–1494. https://doi.org/10.2166/nh.2019.134 (cit. on p. 5).

Bittner, D., Narany, T. S., Kohl, B., Disse, M., & Chiogna, G. (2018). Modeling the hydrological impact of land use change in a dolomite-dominated karst system. *Journal of Hydrology*, *567*, 267–279. https://doi.org/10.1016/j.jhydrol.2018.10.017 (cit. on p. 8).

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., ... Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrol. Sci. J.*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507 (cit. on p. 4).

Bonacci, O., Pipan, T., & Culver, D. C. (2009). A framework for karst ecohydrology. *Environ Geol*, *56*(5), 891–900. https://doi.org/10.1007/s00254-008-1189-0 (cit. on p. 2).

Chen, L., Tan, L., Zhao, M., Sinha, A., Wang, T., & Gao, Y. (2023). Karst carbon sink processes and effects: A review. *Quaternary International*, *652*, 63–73. https://doi.org/10.1016/j.quaint.2023.02.009 (cit. on p. 2).

Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., & Goldscheider, N. (2017). The World Karst Aquifer Mapping project: Concept, mapping procedure and map of Europe. *Hydrogeol J*, *25*(3), 771–785. https://doi.org/10.1007/s10040-016-1519-3 (cit. on p. 2).

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., & Papalexiou, S. M. (2021). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resour. Res.*, *57*(9), e2020WR029001. https://doi.org/10.1029/2020WR029001 (cit. on p. 5).

De Waele, J., Gutiérrez, F., Parise, M., & Plan, L. (2011). Geomorphology and natural hazards in karst areas: A review. *Geomorphology*, *134*(1-2), 1–8. https://doi.org/10.1016/j.geomorph.2011.08.001 (cit. on p. 4).

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.*, *13*(6), 913–921. https://doi.org/10.5194/hess-13-913-2009 (cit. on p. 5).

Dirnböck, T., Kobler, J., Kraus, D., Grote, R., & Kiese, R. (2016). Impacts of management and climate change on nitrate leaching in a forested karst area. *Journal of Environmental Management*, *165*, 243–252. https://doi.org/10.1016/j.jenvman.2015.09.039 (cit. on p. 4).

Fandel, C., Ferré, T., Chen, Z., Renard, P., & Goldscheider, N. (2021). A model ensemble generator to explore structural uncertainty in karst systems with unmapped conduits. *Hydrogeol J*, *29*(1), 229–248. https://doi.org/10.1007/s10040-020-02227-6 (cit. on p. 7).

Ford, D., & Williams, P. (2007). Karst Hydrogeology. In *Karst Hydrogeology and Geomorphology* (pp. 103–144). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118684986.ch5 (cit. on p. 1).

Gill, L. W., Schuler, P., Duran, L., Morrissey, P., & Johnston, P. M. (2021). An evaluation of semidistributed-pipe-network and distributed-finite-difference models to simulate karst systems. *Hydrogeol J*, *29*(1), 259–279. https://doi.org/10.1007/s10040-020-02241-8 (cit. on p. 7).

Goldscheider, N. (2005). Fold structure and underground drainage pattern in the alpine karst system Hochifen-Gottesacker. *Eclogae Geol. Helv.*, *98*(1), 1–17. https://doi.org/10.1007/s00015-005-1143-z (cit. on p. 2).

Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., & Veni, G. (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeol J*, *28*(5), 1661–1677. https://doi.org/10.1007/s10040-020-02139-5 (cit. on pp. 1, 2, 4).

Goldscheider, N., & Drew, D. (2007). *Methods in karst hydrogeology*. Taylor & Francis. (Cit. on p. 6).

Gu, X., Xiong, K., Zhang, J., & Chen, H. (2023). A Comprehensive Analysis on Integrity Conservation of World Natural Heritage Site and Buffer Zone Tourism Development with an Implication for Karst Heritage Sites. *Geoheritage*, *15*(1), 8. https://doi.org/10.1007/s12371-022-00779-5 (cit. on p. 2).

Guinot, V., Savéan, M., Jourde, H., & Neppel, L. (2015). Conceptual rainfall–runoff model with a two-parameter, infinite characteristic time transfer function. *Hydrol. Process.*, *29*(22), 4756–4778. https://doi.org/10.1002/hyp.10523 (cit. on p. 8).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003 (cit. on p. 5).

Gutiérrez, F., Parise, M., De Waele, J., & Jourde, H. (2014). A review on natural and human-induced geohazards and impacts in karst. *Earth-Science Reviews*, *138*, 61–88. https://doi.org/10.1016/j.earscirev.2014.08.002 (cit. on p. 4).

Hamed Ferjani, A., Guellala, R., Gannouni, S., & Inoubli, M. H. (2020). Enhanced Characterization of Water Resource Potential in Zaghouan Region, Northeast Tunisia. *Nat Resour Res*, *29*(5), 3253–3274. https://doi.org/10.1007/s11053-020-09647-x (cit. on p. 5).

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.*, *52*(3), 218–242. https://doi.org/10.1002/2013RG000443 (cit. on pp. 4, 5, 7, 8).

Hartmann, A., Lange, J., Weiler, M., Arbel, Y., & Greenbaum, N. (2012a). A new approach to model the spatial and temporal variability of recharge to karst aquifers. *Hydrol. Earth Syst. Sci.*, *16*(7), 2219–2231. https://doi.org/10.5194/hess-16-2219-2012 (cit. on p. 8).

Hartmann, A. (2018). Experiences in calibrating and evaluating lumped karst hydrological models. *SP*, *466*(1), 331–340. https://doi.org/10.1144/SP466.18 (cit. on pp. 6, 8).

Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proc. Natl. Acad. Sci.*, *114*(11), 2842–2847. https://doi.org/10.1073/pnas.1614941114 (cit. on p. 5).

Hartmann, A., Lange, J., Aguado, A. V., Mizyed, N., Smiatek, G., & Kunstmann, H. (2012b). A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *J. Hydrol.*, *468–469*, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024 (cit. on p. 4).

Jeannin, P.-Y., Artigue, G., Butscher, C., Chang, Y., Charlier, J.-B., Duran, L., Gill, L., Hartmann, A., Johannet, A., Jourde, H., Kavousi, A., Liesch, T., Liu, Y., Lüthi, M., Malard, A., Mazzilli, N., Pardo-Igúzquiza, E., Thiéry, D., Reimann, T., … Wunsch, A. (2021). Karst modelling challenge 1: Results of hydrological modelling. *J. Hydrol.*, *600*, 126508. https://doi.org/10.1016/j.jhydrol.2021.126508 (cit. on pp. 6–8).

Jourde, H., Massei, N., Mazzilli, N., Binet, S., Batiot-Guilhe, C., Labat, D., Steinmann, M., Bailly-Comte, V., Seidel, J., Arfib, B., Charlier, J., Guinot, V., Jardani, A., Fournier, M., Aliouache, M., Babic, M., Bertrand, C., Brunet, P., Boyer, J., … Wang, X. (2018). SNO KARST: A French Network of Observatories for the Multidisciplinary Study of Critical Zone Processes in Karst Watersheds and Aquifers. *Vadose Zone J.*, *17*(1), 180094. https://doi.org/10.2136/vzj2018.04.0094 (cit. on p. 8).

Jourde, H., & Wang, X. (In press). Advances, challenges and perspective in modelling the functioning of karst systems: A review. *Environ. Earth Sci.* (cit. on p. 7).

Jouves, J., Viseur, S., Arfib, B., Baudement, C., Camus, H., Collon, P., & Guglielmi, Y. (2017). Speleogenesis, geometry, and topology of caves: A quantitative study of 3D karst conduits. *Geomorphology*, *298*, 86–106. https://doi.org/10.1016/j.geomorph.2017.09.019 (cit. on p. 2).

Jukić, D., & Denić-Jukić, V. (2006). Nonlinear kernel functions for karst aquifers. *Journal of Hydrology*, *328*(1-2), 360–374. https://doi.org/10.1016/j.jhydrol.2005.12.030 (cit. on p. 7).

Kavousi, A., Reimann, T., Liedl, R., & Raeisi, E. (2020). Karst aquifer characterization by inverse application of MODFLOW-2005 CFPv2 discrete-continuum flow and transport model. *Journal of Hydrology*, *587*, 124922. https://doi.org/10.1016/j.jhydrol.2020.124922 (cit. on p. 7).

Király, L. (1998). Modelling karst aquifers by the combined discrete channel and continuum approach. *Bull. Hydrogeol.*, *16*, 77–98 (cit. on p. 2).

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019 (cit. on p. 5).

Kong-A-Siou, L., Johannet, A., Borrell Estupina, V., & Pistre, S. (2015). Neural networks for karst groundwater management: Case of the Lez spring (Southern France). *Environ. Earth Sci.*, *74*(12), 7617–7632. https://doi.org/10.1007/s12665-015-4708-9 (cit. on p. 7).

Kovács, A., & Sauter, M. (2007). Modelling karst hydrodynamics. In N. Goldscheider & D. Drew (Eds.), *Methods in Karst Hydrogeology* (pp. 201–222). Taylor & Francis. (Cit. on p. 6).

Labat, D., Ababou, R., & Mangin, A. (1999). Linear and nonlinear input/output models for karstic springflow and flood prediction at different time scales. *Stochastic Environmental Research and Risk Assessment*, *13*(5), 337. https://doi.org/10.1007/s004770050055 (cit. on p. 7).

Lehmann, P., Stauffer, F., Hinz, C., Dury, O., & Flühler, H. (1998). Effect of hysteresis on water flow in a sand column with a fluctuating capillary fringe. *J. Contam. Hydrol.*, *33*(1), 81–100. https://doi.org/10.1016/S0169-7722(98)00066-7 (cit. on p. 8).

Lionello, P., & Scarascia, L. (2018). The relation between climate change in the Mediterranean region and global warming. *Reg Environ Change*, *18*(5), 1481–1493. https://doi.org/10.1007/s10113-018-1290-1 (cit. on p. 4).

Liu, Z., Dreybrodt, W., & Wang, H. (2010). A new direction in effective accounting for the atmospheric CO2 budget: Considering the combined action of carbonate dissolution, the global water cycle and photosynthetic uptake of DIC by aquatic organisms. *Earth-Science Reviews*, *99*(3-4), 162–172. https://doi.org/10.1016/j.earscirev.2010.03.001 (cit. on p. 2).

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrol. Earth Syst. Sci.*, *18*(2), 575–594. https://doi.org/10.5194/hess-18-575-2014 (cit. on p. 5).

Lowenstam, H. A., & Weiner, S. (1989). *On biomineralization*. Oxford University Press. (Cit. on p. 1).

Maier, H. R., Jain, A., Dandy, G. C., & Sudheer, K. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modelling & Software*, *25*(8), 891–909. https://doi.org/10.1016/j.envsoft.2010.02.003 (cit. on p. 7).

Maréchal, J.-C., Vestier, A., Jourde, H., & Dörfliger, N. (2013). L'hydrosystème du Lez : Une gestion active pour un karst à enjeux. *Karstologia*, *62*, 1–6 (cit. on p. 5).

Mazzilli, N. (2011). *Sensibilité et incertitude de modélisation sur les bassins à forte composante karstique* [Doctoral dissertation, Université de Montpellier]. (Cit. on pp. 7, 8).

Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Soglio, L. D., & Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.*, *122*, 103927. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on p. 8).

Mazzilli, N., Sivelle, V., Cinkus, G., Jourde, H., & Bertin, D. (2023). *KarstMod User Guide - version 3.0*. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on p. 8).

Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2021). Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water*, *13*(1), 28. https://doi.org/10.3390/w13010028 (cit. on p. 5).

Nerantzaki, S. D., & Nikolaidis, N. P. (2020). The response of three Mediterranean karst springs to drought and the impact of climate change. *J. Hydrol.*, *591*, 125296. https://doi.org/10.1016/j.jhydrol.2020.125296 (cit. on p. 4).

Ollivier, C., Mazzilli, N., Olioso, A., Chalikakis, K., Carrière, S. D., Danquigny, C., & Emblanch, C. (2020). Karst recharge-discharge semi distributed model to assess spatial variability of flows. *Sci. Total Environ.*, *703*, 134368. https://doi.org/10.1016/j.scitotenv.2019.134368 (cit. on p. 8).

Parise, M., Gabrovsek, F., Kaufmann, G., & Ravbar, N. (2018). Recent advances in karst research: From theory to fieldwork and applications. *SP*, *466*(1), 1–24. https://doi.org/10.1144/SP466.26 (cit. on pp. 4, 5).

Pelletier, P. M. (1988). Uncertainties in the single determination of river discharge: A literature review. *Can. J. Civ. Eng.*, *15*(5), 834–850. https://doi.org/10.1139/l88-109 (cit. on p. 4).

Pinault, J.-L., Doerfliger, N., Ladouche, B., & Bakalowicz, M. (2004). Characterizing a coastal karst aquifer using an inverse modeling approach: The saline springs of Thau, southern France: CHARACTERIZING A COASTAL KARST AQUIFER. *Water Resour. Res.*, *40*(8). https://doi.org/10.1029/2003WR002553 (cit. on p. 2).

Pipan, T., & Culver, D. (2013). Forty years of epikarst: What biology have we learned? *IJS*, *42*(3), 215–223. https://doi.org/10.5038/1827-806X.42.3.5 (cit. on p. 2).

Ravbar, N., & Šebela, S. (2015). The effectiveness of protection policies and legislative framework with special regard to karst landscapes: Insights from Slovenia. *Environmental Science & Policy*, *51*, 106–116. https://doi.org/10.1016/j.envsci.2015.02.013 (cit. on p. 2).

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process – A framework and guidance. *Environ. Model. Softw.*, *22*(11), 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004 (cit. on p. 5).

Ruban, D. (2018). Karst as Important Resource for Geopark-Based Tourism: Current State and Biases. *Resources*, *7*(4), 82. https://doi.org/10.3390/resources7040082 (cit. on p. 2).

Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., & Wagener, T. (2018). V2Karst V1.1: A parsimonious large-scale integrated vegetation–recharge model to simulate

the impact of climate and land cover change in karst regions. *Geosci. Model Dev.*, *11*(12), 4933–4964. https://doi.org/10.5194/gmd-11-4933-2018 (cit. on p. 8).

Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, *32*(8), 1120–1125. https://doi.org/10.1002/hyp.11476 (cit. on p. 5).

Sivelle, V., Jourde, H., Bittner, D., Mazzilli, N., & Tramblay, Y. (2021). Assessment of the relative impacts of climate changes and anthropogenic forcing on spring discharge of a Mediterranean karst system. *J. Hydrol.*, *598*, 126396. https://doi.org/10.1016/j.jhydrol.2021.126396 (cit. on p. 4).

Stevanović, Z. (Ed.). (2015). *Karst Aquifers—Characterization and Engineering*. Springer International Publishing. https://doi.org/10.1007/978-3-319-12850-4 (cit. on pp. 1, 2).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on pp. 2, 4).

Tendil, A. (2018). *Contrôles tectoniques, climatiques et paléogéographiques sur l'architecture stratigraphique de la plateforme carbonatée urgonienne provençale (France) : Approches sédimentologiques, géochimiques et numériques intégrées* [These de doctorat]. Aix-Marseille. (Cit. on p. 2).

Tritz, S., Guinot, V., & Jourde, H. (2011). Modelling the behaviour of a karst system catchment using non-linear hysteretic conceptual model. *J. Hydrol.*, *397*(3-4), 250–262. https://doi.org/10.1016/j.jhydrol.2010.12.001 (cit. on p. 8).

Vaks, A., Bar-Matthews, M., Ayalon, A., Matthews, A., & Frumkin, A. (2018). Pliocene–Pleistocene palaeoclimate reconstruction from Ashalim Cave speleothems, Negev Desert, Israel. *SP*, *466*(1), 201–216. https://doi.org/10.1144/SP466.10 (cit. on p. 2).

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., & Teng, J. (2010). Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, *394*(3-4), 447–457. https://doi.org/10.1016/j.jhydrol.2010.09.018 (cit. on p. 5).

White, A. F., & Blum, A. E. (1995). Effects of climate on chemical_ weathering in watersheds. *Geochimica et Cosmochimica Acta*, *59*(9), 1729–1747. https://doi.org/10.1016/0016-7037(95)00078-E (cit. on p. 2).

Williams, P. (2008). *World Heritage Caves and Karst. Gland, Switzerland: IUCN*. (Cit. on p. 2).

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., & Goldscheider, N. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrol. Earth Syst. Sci.*, *26*(9), 2405–2430. https://doi.org/10.5194/hess-26-2405-2022 (cit. on p. 7).

Xanke, J., Goldscheider, N., Michel, B., Barberá, J. A., Broda, S., Chen, Z., Ghanmi, M., Andreas, G., Hartmann, A., Jourde, H., Liesch, T., Matías, M., Petitta, M., Ravbar, N., & Stevanovic, Z. (2022). *Mediterranean Karst Aquifer Map 1:5,000,000 (MEDKAM)* (Shapefile,PDF). Shapefile,PDF. BGR, KIT, UNESCO. https://doi.org/10.25928/MEDKAM.1 (cit. on p. 4).

Part I

Characterisation of the hydrological functioning of karst systems

Introduction to Part 1

This part aims to address the first research question of the thesis (see i.) and focuses on the characterisation of the hydrological functioning of karst systems through the analysis of discharge time series, with the consideration of data-scarce contexts. The main research questions addressed in this part are as follows:

i. **How to characterise the functioning of a karst system in data-scarce contexts?**

ii. Which discharge time series analyses are relevant for classifying the functioning of karst systems?

iii. Does the hydrological functioning of a karst system exhibit a strong identifiability?

iv. Is the classification consistent irrespective of the length of the time series?

v. How can the classification process be facilitated for researchers and stakeholders?

Chapter 2 investigates different time series analyses that can be effectively applied to karst spring discharge records, even in data-scarce context. Different indicators of karst dynamics are calculated using a core dataset of 10 karst systems. The selection of the most insightful indicators and the proposal of a classification are based on multivariate analyses. The relevance and robustness of the classification are further examined on a larger dataset.

Chapter 3 presents a toolbox that facilitates the completion of discharge time series analyses and the classification of a karst system. KarstID is a freely available, open-source software with ongoing development on a developer community platform. Its user-friendly installation and launch make it especially accessible for those without programming experience.

Chapter 2

Classification of karst hydrological functioning

Classification is a first-line tool for understanding the main characteristics of a natural system's response. We propose a new classification of karst systems hydrological functioning that is based on spring discharge time series and takes profit of spring discharge databases to encompass the high diversity of karst hydrological functioning. It discriminates six different classes based on three relevant indicators of karst hydrological functioning. A core dataset made of 10 karst systems was first considered for the set-up of the classification. The spring discharge time series were investigated according to recession curves, statistical and signal analyses to identify relevant indicators of hydrological functioning. The selection of the most relevant indicators and the proposal of the classification were based on multivariate analyses. The classification was then tested on spring discharge time series of 78 karst systems located worldwide. All the systems homogeneously spread among the six proposed classes, which highlights the relevance of the approach and the representativeness of the various classes of hydrological functioning. Results from the proposed methodology were finally discussed to explore its limitations and define guidelines for its application.

This work has contributed to the KARMA project in the form of a deliverable and has been presented during a progress meeting. It has also resulted in a publication in Journal of Hydrology (Cinkus et al., 2021a) and a presentation during EGU21 (Cinkus et al., 2021b).

Article:

Cinkus, G., Mazzilli, N., and Jourde, H.: Identification of relevant indicators for the assessment of karst systems hydrological functioning: Proposal of a new classification, J. Hydrol., 603, 127006, https://doi.org/10.1016/j.jhydrol.2021.127006, 2021.

Contents

2.1 Introduction

10 % of the world's population is dependent on karst water resources for drinking water (Stevanović, 2018). Karst systems are underground entities that drain recharge water over a catchment towards a main outlet. The water is drained through conduits, fractures and matrix, which originate from the dissolution of the calcite deposits by acidic water from the surface. Understanding the functioning of these complex and heterogeneous systems is therefore a major challenge for long-term water resource management. Over the past century, different methods have been developed to analyse hydrological time series, and subsequently characterise the functioning of karst systems. These methods can be considered as a preliminary step in the development and design of hydrological models of karst functioning for sustainable water resource management.

Classification has been widely used in surface hydrology to characterise hydrosystems. Although the three-dimensional properties of aquifers generate an additional complexity, some authors proposed different methodologies to classify them based on geological, morphological and hydrological functioning analyses (Dahl et al., 2007; Heath, 1982; Heudorfer et al., 2019). In many cases however, these classifications fail to address the complexity of karst systems, which are strongly heterogeneous and correspond to a wide diversity of hydrological functioning. For these reasons, many authors worked on classifications specific to karst systems, either based on geological and morphological analyses (Jouves et al., 2017; Mylroie, 2020; Veress, 2020; Waltham and Fookes, 2003), hydrological response analyses (Bonacci, 1993; Flora, 2004; Kullman, 2000; Malík and Vojtková, 2012; Mangin, 1975; Mangin, 1984; Rashed, 2012; Soulios, 1991; Springer et al., 2008) or even karst groundwater microbiological analyses (Sinreich et al., 2013).

The spring discharge of a karst system is considered as a base information in karst hydrogeology. It results from the combination of flows from the different compartments of the hydrosystem (soil, epikarst, unsaturated and saturated zones). The widely increased development of karst spring discharge monitoring offers the opportunity to study the relations between discharge and hydrological functioning in depth. Some authors proposed classifications of karst systems based on the analysis of spring discharge time series, either with visual interpretations (Bonacci, 1993; Soulios, 1991), by calculating indicators of functioning (Flora, 2004; Mangin, 1975; Mangin, 1984; Rashed, 2012; Springer et al., 2008) or by interpreting the parameters of recession models (Kullman, 2000; Malík and Vojtková, 2012; Mangin, 1975).

However, the aforementioned classifications have been developed by analysing only few karst systems or without considering the high diversity of karst hydrological functioning. Therefore, diversity in karst systems physical properties and hydrological functioning is not fully considered, which impairs the relevance of these classifications and raises the need for a more generic approach. This work aims to provide a new classification of karst systems hydrological functioning with the following key features: (i) a clear methodological basis, (ii) the analysis of a wide diversity of karst systems representative of contrasted hydrodynamics behaviours, (iii) an approach being relevant worldwide and in a data-scarce context (i.e. sites where there is little knowledge of the system, or only discharge monitoring for a few years).

In this paper, we took advantage of the recent release of spring discharge time series in databases such as the French Karst National Observatory Service (SNO KARST, Jourde et al., 2018) or the WoKaS database (Olarinoye et al., 2020) to propose a new classification of karst systems hydrological functioning. The typology describes a system

where one single hydrodynamic response to precipitation impulse is expected. The aim of the classification is to characterise the hydrological functioning of a system, but not to decorrelate the factors that influence the functioning. The paper is organised as follows. In Section 2.2, we define the general characteristics of the karst systems considered in this study. Section 2.3 presents the various tools and analyses considered for the characterisation of karst systems hydrological functioning. The most relevant indicators of karst hydrodynamics are identified and presented in Section 2.4, using the discharge time series of 10 well-known karst systems that cover a wide range of hydrological functioning. Section 2.5 is devoted to multivariate analyses that are considered for the proposal of the new classification described in Section 2.6. The discussion in Section 2.7 aims to evaluate the relevance of the proposed approach applied to 78 karst systems and to highlight some of its limitations. Section 2.8 gives the conclusions.

2.2 Data and study sites

This section presents the data we used to develop and test the classification, which involve in two different datasets: (i) a core dataset for the assessment of the most relevant indicators of functioning and the design of the classification, and (ii) a complementary dataset for assessing the most efficient recession model, testing the classification and identifying its strength and limitations.

2.2.1 Core dataset

To ensure the quality of the study and its relevance to the problem, we performed spring selection on the basis of three criteria: (i) quality of the hydrodynamic monitoring, which is function of time-step, instrumentation, measurement uncertainty and length of the time series, (ii) diversity of the hydrological functioning among the karst systems, meaning that the final dataset should cover a wide range of hydrological functioning (e.g. related to dimensions of the catchment, rainfall, degree of karstification, hydrological functioning, etc.), and (iii) existing knowledge from prior studies, to ensure that the classification is consistent with the actual knowledge on system functioning.

We selected 10 karst systems (Table 2.1), located in France. We retrieved data from several organisations: The SNO KARST, the Parc Naturel Régional des Grands Causses (PNRGC), Suez, and the DREAL of Bourgogne Franche-Comté. Selected systems have been the subject of several comprehensive studies based of methods such as geology, cartography, field observations, tracing, geochemistry, time series analysis and modelling.

2.2.2 Complementary dataset

In order to check the relevance of the method and its capacity to differentiate karst systems functioning, the classification resulting from the analysis of the aforementioned well-known karst systems was tested on a complementary dataset of 68 karst systems with different characteristics (e.g. dimensions of the catchment, meteorological regime, climate, and karstification degree). We worked with springs discharge time series of 23 French karst systems coming from a database provided by the French state (Banque Hydro) and took the other 45 from the WoKaS (World Karst Spring hydrograph) database, which provides details of over 400 karst systems worldwide (Olarinoye et al., 2020).

System	Köppen-Geiger climate classification	Catchment area [km^2]	Discharge time series	
			Length [year]	Time step
Aliou	Cfb	12	45	Daily
Baget	Cfb	13	47.5	Daily
Durzon	Csb	117	9	Daily
Esperelle	Cfb	91	8	Daily
Fontaine-de-Nîmes	Csa	45	18.8	Daily
Fontaine-de-Vaucluse	Csa	1115	52.2	Daily
Lods	Cfb	35	6.4	Daily
Mouline	Cfb	32	9	Daily
Mouthe	Dfb	50	7.3	Daily
Toulon	Cfb	100	5.5	Daily

Table 2.1: General characteristics of the selected karst system (Bakalowicz and Ricard, 1994; Blavoux et al., 1992; Cholet, 2017; Lorette et al., 2018; Mangin, 1975; Maréchal and Ladouche, 2006; Moussu, 2011) and their associated discharge time series.

The quality of these 68 springs discharge time series is appreciated according to low (C2) to very good (A) quality discharge data (Figure 2.1C) as proposed in Olarinoye et al. (2020). The dataset is fairly well distributed throughout the world, with karst springs discharge from 17 countries (Figure 2.1B). The considered karst systems are located in various climatic conditions. According to Köppen-Geiger classification (Peel et al., 2007), these climatic conditions correspond to 12 different climates, the temperate oceanic (Cfb) being the most represented (63.2 %, Figure 2.1A).

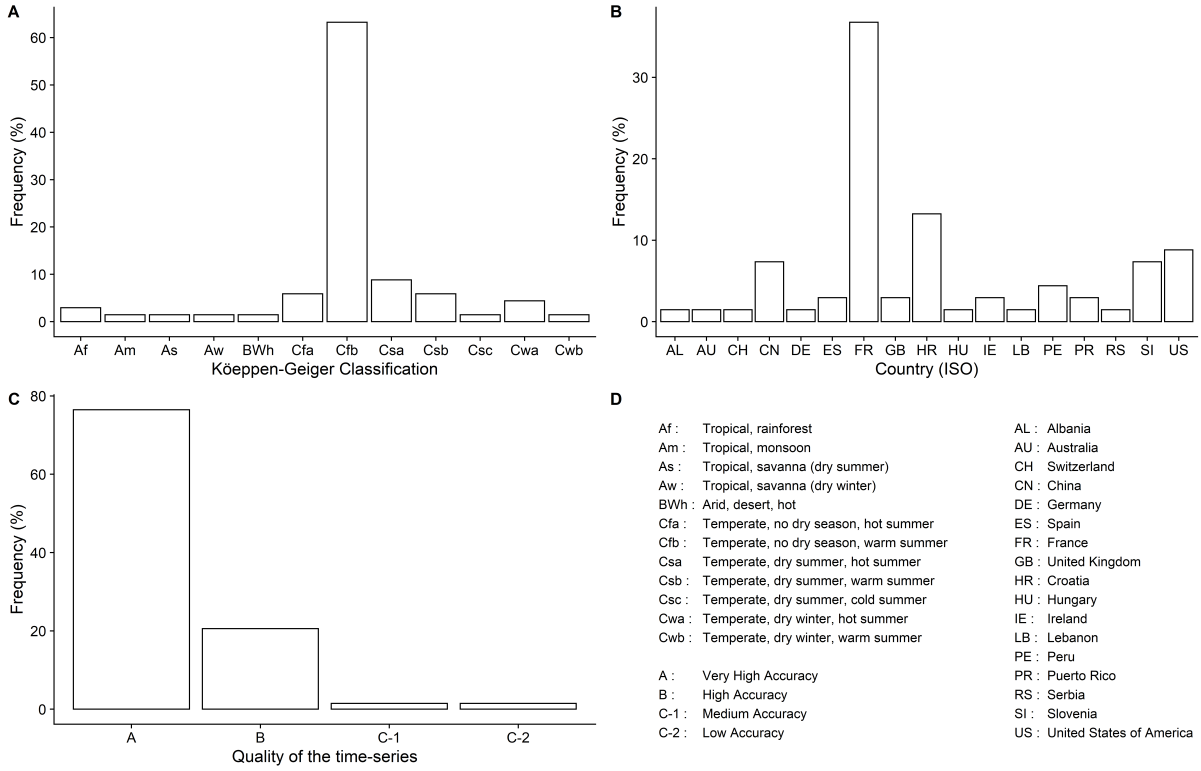


Figure 2.1: Details about the complementary dataset, regarding (A) climate (Köppen-Geiger classification), (B) countries and (C) quality of the time series. (D) is the legend for the different abbreviations.

2.3 Methodology

The discharge is directly related to the recharge and the emptying of the capacitive function of a system, but also depend on the system geometry and hydrodynamic properties (Malík, 2015). This section provides details about four methods for analysing discharge time series and the indicators that can be retrieved from these methods to characterise the functioning of a karst system.

2.3.1 Statistical analyses

Statistical indicators of discharge time series provide basic information about the overall functioning of a system. The most common are the mean, minimum, maximum, standard deviation, and the various quantiles. The mean interannual discharge depends on both the dimensions of the catchment and the mean recharge; it can therefore be used to assess the dimensions of a system. The observed minimum and maximum discharges make it possible to understand the flow amplitude. However, significant uncertainty related to the extrapolation of extreme discharges at springs is associated with this indicator.

The comparison between various karst systems is facilitated when using nondimensionalised indicators, such as the coefficient of variation (CV), which is the ratio between the standard deviation and the mean of the discharge time series. Netopil (1971) proposed the calculation of a "characteristic discharge", which is the ratio between the 0.9 quantile (discharge value that is exceeded 10 % of the time) and the 0.1 quantile. This "characteristic discharge" was referred to as spring variability coefficient (SVC) in Ste-

vanović (2015). The SVC is less affected by extreme values. The specific discharge (QS) corresponds to the ratio between the mean interannual discharge and the catchment area. It allows assessing karst hydrodynamics with an implicit consideration of the dimensions of the catchment. However, the area information is not always available and also depends on the quality of hydrogeological studies over the catchment of interest.

2.3.2 Recession curves analysis

The hydrograph of a flood recession corresponds to the period when the discharge gradually decreases as water is not replenished (Toebes and Strang, 1964). It is possible to distinguish two regimes: (i) the influenced (quickflow) regime, which corresponds to the period when the discharge is influenced by the rapid infiltration of water into conduits in the unsaturated zone; and (ii) the non-influenced (baseflow) regime, which begins when rapid infiltration ends, and corresponds to the draining of the saturated zone and less transmissive compartments of the system (Mangin, 1975). In the literature, analysis of recession curves is mainly used to estimate groundwater reserves (Drogue, 1972; Forkasiewicz and Paloc, 1967; Mangin, 1975), determine the hydrodynamic parameters of the aquifer (Mangin, 1975), and provide information on flows, drainage, and karstification degree (Drogue, 1972; Krešić, 2007; Kullman, 2000; Malík, 2006; Malík and Vojtková, 2012; Mangin, 1975). The karstification degree is an indicator of the karst maturity of a system, which is directly related to groundwater recharge, storage capacity, spring discharge dynamics and system connectivity.

Numerous models for the analysis of recession curves of discharge time series have been proposed (Table 2.2). Boussinesq (1877) and Maillet (1905) made the first proposals. Horton (1933) and Barnes (1939) then developed recession models based on Maillet's equation. Coutagne (1948) proposed an equation that simulates the discharge from a reservoir. Padilla et al. (1994) highlighted that Coutagne's equation struggles to reproduce the recession curves of karst systems, and thus introduced a new parameter Qc. This later parameter corresponds either to the discharge from poorly transmissive zones of the aquifer, or to the discharge from aquitards outside the karst system; it may also have no physical meaning. Drogue (1972) proposed to approach the whole recession curve with a hyperbolic function. Mangin (1975) developed a two-equation recession model, representative of the influenced and non-influenced regimes. Kullman (2000) approached recession curves by fitting a model based on a linear equation (Kullman, 1983) for influenced regime, and Maillet's equation for non-influenced regime. Based on Kullman's work, Malík and Vojtková (2012) proposed a classification of karst systems functioning according to the number of equations required to model the recession and the values of the parameters. Other authors (Griffiths and Clausen, 1997; Ladouche et al., 2006; Samani and Ebrahimi, 1996) proposed models inspired by the aforementioned models. We did not consider physical-based models as we do not have any information on reservoir geometry. Further details about empirical, physical-based models and recession curve analysis can be found in Fiorillo (2014).

In this study, we wanted to assess the variability of the hydrological response of the karst systems, which is only possible when accounting for different recession periods. For this reason, we dismissed the use of a Master Recession Curve (MRC) which aims to overcome the problem of recession variation by combining several recession curves into one and only.

The Mangin's model has been widely used to characterise karst systems, mainly be-

Model	Equation	Comment
Boussinesq (1903)	$Q_t = \frac{Q_0}{(1+\alpha t)^2}$	Non-influenced stage Surface water
Maillet (1905)	$Q_t = Q_0 e^{-\alpha t}$	Non-influenced stage Surface water
Horton (1933)	$Q_t = Q_0 e^{-\alpha t^n}$	More suitable to surface water
Barnes (1939)	$Q_t = \sum_{i=1}^{k} Q_0 e^{-\alpha_i t}$	More suitable to surface water
Coutagne (1948)	$Q_t = Q_0[1 + (n-1)\alpha t]^{\frac{n}{(1-n)}}$	Suitable for karst systems
Padilla et al. (1994)	$Q_t = (Q_0 Q_c)[1 + (n-1)\alpha t]^{\frac{n}{(1-n)}} + Q_c$	Suitable for karst systems Q_c strengthens Coutagne model
Drogue (1972)	$Q_t = \frac{Q_0}{(1+\alpha t)^n}$	Suitable for karst systems
Mangin (1975)	$Q_t = Q_{R0} e^{-\alpha t} + q_0 \frac{1-\eta t}{1+\varepsilon t}$	Suitable for karst systems Associated classification
Kullman (2000)	$Q_t = \sum_{i=1}^{k} Q_{0i} e^{-\alpha_i t} + \sum_{j=1}^{k} (\frac{1}{2} + \frac{\lvert 1-B_j t \rvert}{2(1-B_j t)}) Q_{0j}(1 - B_j t)$	Suitable for karst systems Associated classification

Table 2.2: Summary of the main models developed to analyse recession curves, corresponding equations and comment on their most appropriate usage. Q_t [L^3 T^{-1}] is the discharge at time t [T], Q_0 [L^3 T^{-1}] the discharge at $t = 0$ [T], α the recession coefficient [T^{-1}], n [-] a constant, Q_c [L^3 T^{-1}] a constant discharge, Q_{R0} [L^3 T^{-1}] the baseflow extrapolated at $t = 0$, q_0 [L^3 T^{-1}] the influenced discharge corresponding to the difference between Q_0 and Q_{R0}, η [T^{-1}] a constant characterising the speed of infiltration ($\eta = \frac{1}{t_i}$, with t_i [T] being the duration of the influenced stage), ε [T^{-1}] a constant characterising the concavity of the influenced part of the recession curve and β [T^{-1}] a recession coefficient for the turbulent flow. L and T are the dimensions for the base quantities of length and time, respectively.

cause the author proposed a classification based on two indicators derived from the parameters of the model equation. The parameter α is assumed to characterise the draining of the capacitive function of the karst system, which corresponds in most cases to flow from the saturated zone. In case of a low karstification degree, α can be affected by flows occurring in the unsaturated zone, which may induce a lag in the response in the non-influenced regime (Mudarra and Andreo, 2011). The indicator k is strongly linked to the α recession coefficient of the first component of the Mangin model (Table 2.2), which is applied only on the non-influenced regime. The indicator k is thus focused on the slow depletion of the aquifer and assumed to characterise the ability of a system to store and return recharge water. It is calculated with the following equation:

$$k = \frac{V_{DYN}}{V_{an}} \tag{2.1}$$

With V_{DYN} the dynamic volume and V_{an} the interannual mean yearly volume of water discharged at the spring. The dynamic volume is calculated by integrating the exponential function:

$$V_{DYN} = \int_0^\infty Q_i e^{-\alpha t} dt = \frac{Q_i}{\alpha} \tag{2.2}$$

With Q_i the discharge at the time t_i (t_i being the time at which the flow is considered to be laminar and also the beginning of the non-influenced regime) and α the recession coefficient. In his work, Mangin (1975) suggested to characterise the capacity of dynamic storage with the maximum calculated V_{DYN}, as it tends towards a stable value for a large number of analysed recession curves. The indicator i is used to characterise the capacity of a system to filter and attenuate the precipitation signal. It corresponds to the discharge attributed to the influenced regime (second component of the Mangin model, Table 2.2) two days after the flood peak. This discharge is expressed in relative proportion to the influenced discharge q_0 and is calculated with the following equation:

$$i = \frac{1 - 2\eta}{1 + 2\varepsilon} \tag{2.3}$$

With η a constant characterising the speed of infiltration ($\eta = \frac{1}{t_i}$, with t_i being the duration of the influenced stage) and ε a constant characterising the concavity of the recession curve during the influenced stage.

The classification initially proposed by Kullman (2000) and updated by Malík and Vojtková (2012) differentiate systems by their karstification degree. The methodology consists to reproduce a recession curve by fitting one to several equations (either exponential or linear) and calibrate the α and β parameters of each formula. The karstification degree is then deduced from a table based on the presence of different flow sub-regimes (i.e. the number and nature of the equations) and the value of the α and β parameters. It ranges from 0.5 to 10 and is associated with a description of assumed structure of the system and karst groundwater circulation.

2.3.3 Correlational and spectral analyses

Correlational and spectral analyses are time series analyses that are used to study the frequency content of a signal (referred to as "simple analysis") and relations between signals (referred to as "cross-analyses") (Massei et al., 2006). The simple analysis consists

of calculating the autocorrelation function of a signal and the corresponding spectrum (obtained using a Fourier transformation, the calculations are detailed in Appendix 2.A). The principle is to compare the signal with itself over an increasing time interval or shift (Jeannin and Sauter, 1998). The cross-analyses examine the transformation of the input signal into an output signal (Padilla and Pulido-Bosch, 1995). Signal analyses, mainly developed by Box and Jenkins (1976), Brillinger (1975), Hannan (2008), and Jenkins and Watts (1968), were first applied to karst hydrology by Mangin (1984).

According to Mangin (1984), a karst system can be characterised by its response time to a unitary impulse (precipitation) and its inertia, which depends on both the volume of groundwater reserves and karstification degree of the karst system.

Simple correlational and spectral analyses allow determining three indicators of karst hydrological functioning: (i) the memory effect (ME), which is the shift k for an autocorrelation coefficient r_k of 0.2. It translates variation in discharge over time, and is directly related to the inertia of the karst system (Marsaud, 1997); (ii) the regulation time (RT), which is the inverse of the bandwidth, i.e. the maximum ordinate of the spectrum divided by 2 (value of the integral of the function between 0 and $+\infty$). It provides information on the duration of the influence of a unitary impulse (Kovács, 2003; Larocque et al., 1998), on the volume of groundwater reserves (Marsaud, 1997), and makes it possible to assess the overall organisation of flows in the system (e.g., conduits, fractures, and cracks) (Jeannin and Sauter, 1998); (iii) the cut-off frequency or spectral band breadth (SBB), which corresponds to the frequency f at which the value of the spectrum s_f becomes negligible. Beyond this frequency, the spectrum is equal to zero, and can be assimilated to noise (Jeannin and Sauter, 1998). The cut-off frequency provides information on the ability of the system to filter unitary pulses (Marsaud, 1997). The results provide a general idea of how a karst system works, with an emphasis on the inertia of the system and its capacity to attenuate the recharge signal.

2.3.4 Analysis of classified discharges

The analysis of classified discharges provides information on flow regimes within a system, based on discharge monitoring at the outlet of a karst system (Marsaud, 1997). Based on empirical observations, Mangin (1971) suggested that the distribution of discharges (or logarithm of discharges) from karst springs can be approximated by a half-normal Gaussian distribution (the calculations are detailed in Appendix 2.B). He concluded that the comparison between quantiles of measured discharges and quantiles given by the half-normal distribution should follow a straight line. According to this theory, any discontinuities of the line (corresponding to an inflexion point) indicate inhomogeneity in the functioning of the system, below or above a certain range of discharge. Such changes can occur at low or high discharges, and may be positive or negative. The interpretation proposed by Mangin is based on an extremely strong hypothesis, which is that the statistical half-normal Gaussian distribution properly describes the distribution of discharges from karst springs under a "homogeneous functioning".

The method allows identifying particular events inherent in karst hydrology, e.g. overflow at outlet, leakage to another system, storage and emptying phenomena, time varying extent of the recharge catchment. It also allows assessing the quality of the gauging station (Dörflinger, 2010; Grasso and Jeannin, 1994; Marsaud, 1997).

System	Mean discharge [m^3 s^{-1}]	Minimum discharge [m^3 s^{-1}]	Maximum discharge [m^3 s^{-1}]	CV [%]	SVC [-]	QS [mm d^{-1}]
Aliou	0.45	0.0	28.91	190.4	32.1	3.20
Baget	0.49	0.02	10.10	147.2	14.0	3.22
Durzon	1.63	0.72	16.33	59.8	3.3	1.21
Esperelle	1.11	0.16	14.50	139.8	11.8	1.05
Fdn	0.54	0.0	16.52	221.9	40.4	1.04
Fdv	17.54	2.79	85.0	71.3	5.9	1.36
Lods	1.01	0.15	8.54	127.2	19.6	2.48
Mouline	0.51	0.19	3.15	45.7	2.6	1.37
Mouthe	1.92	0.01	18.28	120.1	24.4	3.32
Toulon	0.46	0.27	0.99	37.6	2.5	0.40

Table 2.3: Results of the statistical analyses for the core systems.

2.4 Analysis and selection of indicators of karst dynamics

Analysis of discharge time series were performed for the 10 karst systems aforementioned (core dataset) based on methods detailed in Section 2.3, with the goal to select the most relevant indicators of karst systems hydrodynamics.

2.4.1 Statistical indicators

The results of the statistical analyses highlight the diversity of hydrological functioning of the core systems (Table 2.3). The mean discharge allows to distinguish systems with a low discharge (Aliou, Baget, Fontaine-de-Nîmes (Fdn), Mouline and Toulon), with a medium discharge (Durzon, Esperelle, Lods, and Mouthe), and with a large discharge (Fontaine-de-Vaucluse (Fdv)). Although the mean discharge is highly correlated with the dimensions of the catchment, it is also dependent of the precipitation and the hydrological functioning of the system. CV and SVC are highly correlated with each other (correlation coefficient $R = 0.925$, p-value $= 0.00012$). Both can be related to the inertia of the system and allow to differentiate reactive systems (Aliou, Baget, Esperelle, Fontaine-de-Nîmes, Lods and Mouthe) from inertial systems (Durzon, Fontaine-de-Vaucluse, Mouline and Toulon). There is no evident relation between QS and the characteristics of the system. As the dimensions of the catchment are bypassed, we suppose that QS may be related to the karstification degree or the specific recharge (volume of water that goes to the aquifer, by unit area). As QS requires the knowledge about the area of the recharge catchment, which is either unknown or with high uncertainty, it is not retained as a relevant indicator for the classification.

2.4.2 Recession indicators

The dynamics of a karst system can be either (i) at infra-day time scale, meaning that the system is reactive with fast variations in discharge of the order of an hour, or (ii) at daily time scale, meaning that the system has high inertia and changes in discharge can be assessed on a daily basis. A comparison of results issued from recession curves

analysis performed on hourly and daily time series showed that, even for reactive karst systems, the daily time scale provides sufficient accuracy.

We selected a total of 93 recession curves (Figure 2.2) from the overall time series of each system of the core dataset, with no distinction between seasonal or event-scale recessions, on the basis of the following conditions: (i) the peak flood discharge must be significantly high regarding the overall dynamics of the system. We suggest at least one tenth of the maximum discharge of the discharge time series, or greater than the interannual mean discharge. However, we do not exclude the eventuality of site-specific thresholds; (ii) there should be little or no disruption during the recession (e.g. precipitation leading to untimely peaks). In cases where the disruption was of short duration, data could be removed and replaced with a blank; and (iii) the recession must be complete, meaning it should include both the influenced regime and the entire non-influenced regime (with some tolerance for high-inertia systems). In the specific case when karst spring behaviour is influenced by a particular hydrological functioning (e.g. the activation of an overflow outlet) that appears on the recession curve as a bending point, our selection concerned only the last, unaffected part of the curve, including the end of the recession limb. This approach ensured that the models, which are not mean to fit curves with bending points other than the one between the influenced and non-influenced regime, were correctly calibrated. Information loss was relatively minor, as discharges that were excluded from the analysis only represented a tiny part of the overall flow.

The bending point required for Mangin's model (Mangin, 1975) was defined manually and corresponds to the time t when the non-influenced regime begins (when flow is considered as part of the non-influenced regime).

Out of the 9 recession models presented in Section 2.3.2, 4 models suit the study's requirements (we identified them as relevant for karst hydrodynamics analysis but also easy to automate). They correspond to the models from Coutagne (1948), Drogue (1972), Mangin (1975), and Padilla et al. (1994), and will be further referred as Hyperbolic, Coutagne, Padilla and Mangin models, respectively. We tested the 4 models by examining their performance in fitting all 390 observed recession curves of both core dataset and complementary dataset (Figure 2.3), and by performing a sensitivity analysis.

We found Hyperbolic, Padilla and Mangin models to be relatively successful in fitting the observed recession curves, although Padilla and Hyperbolic model significantly failed (Nash-Sutcliffe Efficiency (NSE) lower than 0) on 31 and 3 recession curves, respectively. This illustrates that Padilla and Hyperbolic models are not suitable for all the karst systems. The median relative error of Hyperbolic, Padilla and Mangin models are of 7.7 %, 6.7 % and 5.6 %, respectively; with median NSE of 0.986, 0.989 and 0.995. Coutagne model showed a poor performance with a median relative error of 38 % and a median NSE of 0.730. The sensitivity analyses revealed that Coutagne and Padilla models have equifinality issues for the parameters α and n, with only Q_c having an optimum. On the other hand, the parameters of Hyperbolic and Mangin models have systematically an optimum.

We selected Mangin model for the analysis of recession curves, as it provides a consistent and very good fit for all the recession curves (lowest NSE of 0.91), with a limited equifinality. Moreover, this model has been widely used for years and its indicators are well known by the community. Results show that these indicators clearly differentiate among systems of the core dataset (Figure 2.4). The analysis of i values, which allow assessing the capacity of a system to filter and attenuate the precipitation signal, revealed a relationship between the value of this indicator, and the saturation state of the system

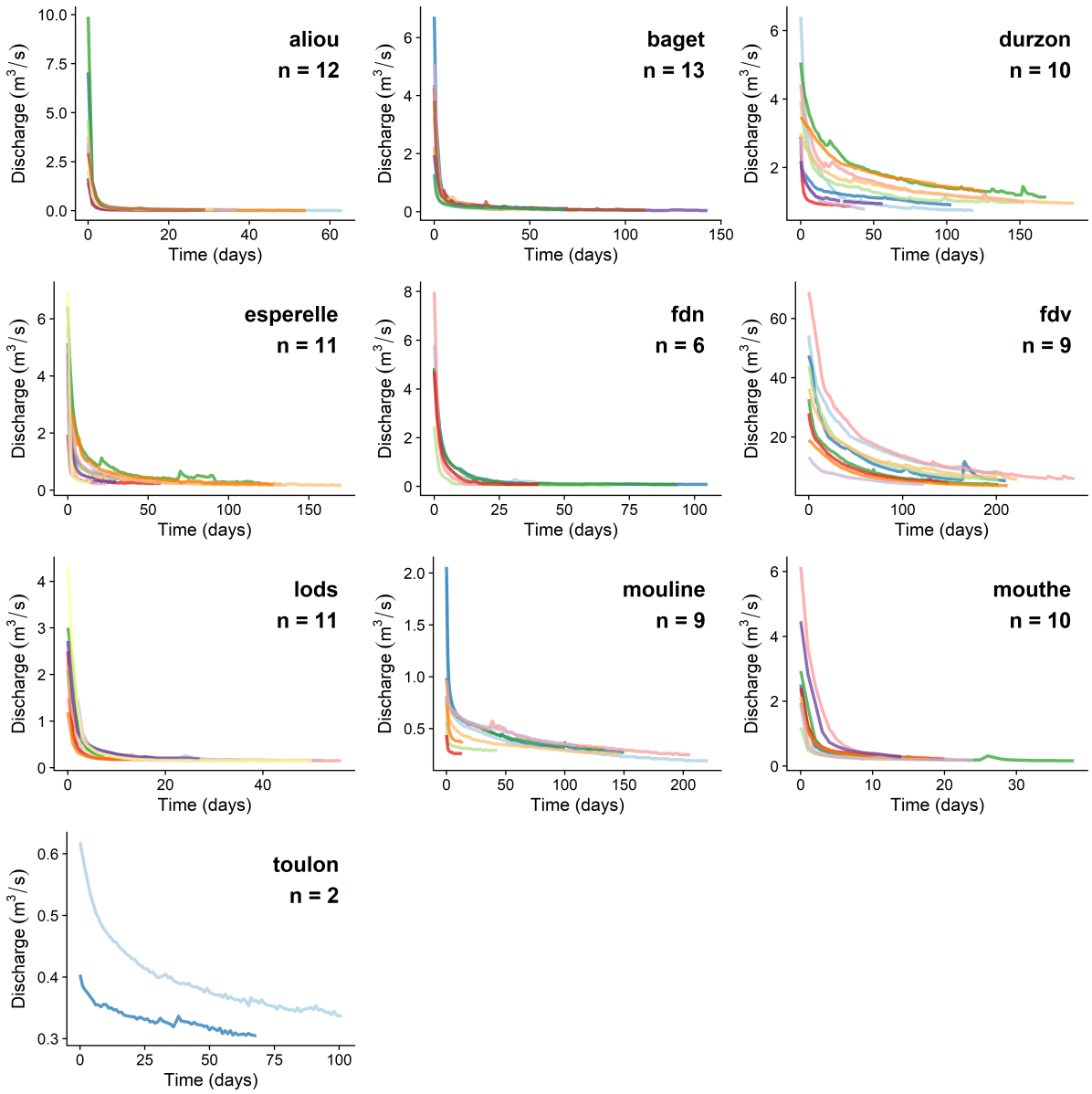


Figure 2.2: Selected recession curves for each of the 10 karst systems. n corresponds to the number of recession curves identified over the available discharge time series of each spring (a total of 93 recession curves is considered).

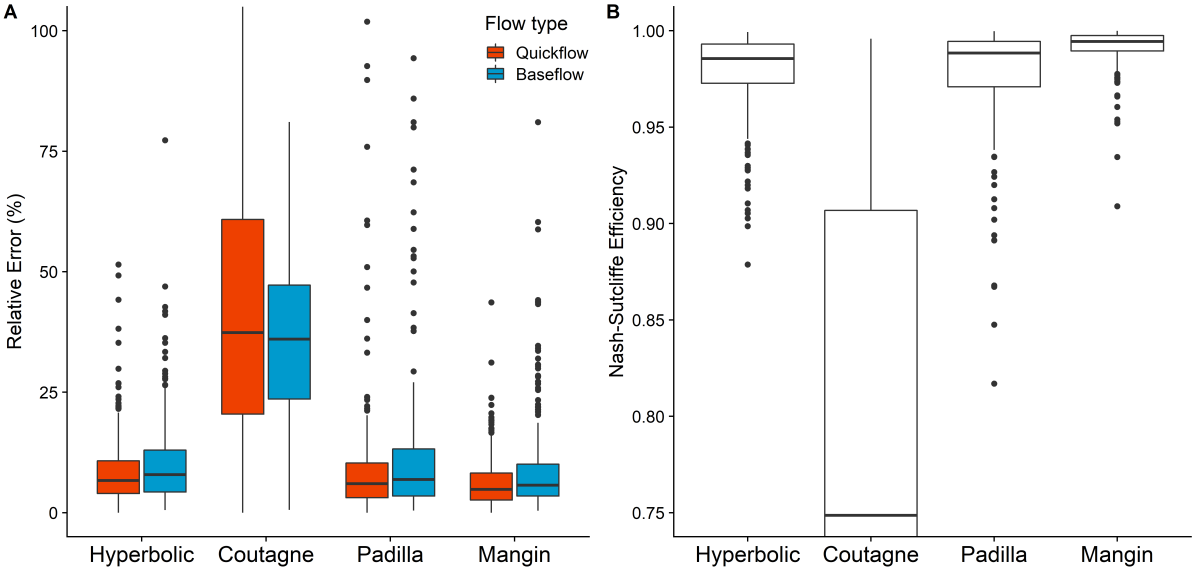


Figure 2.3: Comparison of the performance of the models with respect to observed discharge over all recessions of both core and complementary dataset. (A) Boxplot of relative errors for influenced and non-influenced flow regimes. (B) Boxplot of Nash-Sutcliffe Efficiency.

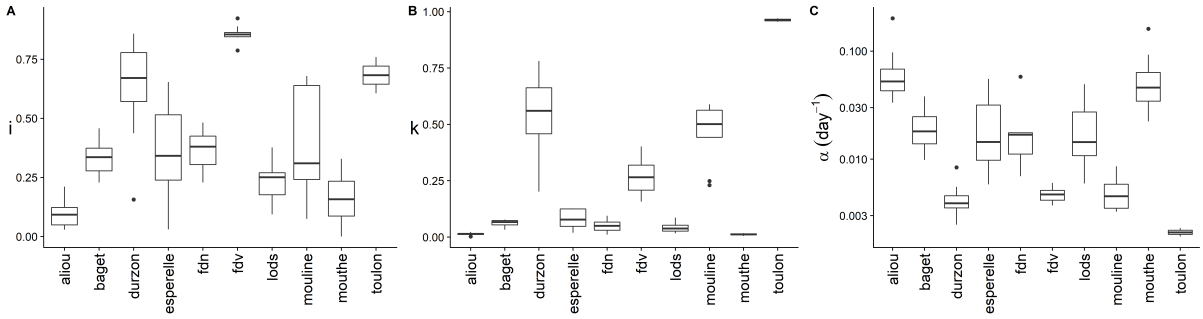


Figure 2.4: Boxplot of the values of the indicators proposed by Mangin (1975) after analysing the spring discharge recession curves for each of the core systems. (A) i, (B) k, and (C) α.

(Figure 2.5), which corresponds to the volume of water stored in both the saturated and unsaturated zones. Variability in i can be a consequence of either (i) the volume of water already stored in the karst system when the recharge occurs (i.e. the influence of the saturation state on the system connectivity), but also (ii) the variability of the overall organisation of flow and groundwater storage between matrix, fracture and conduits in the different compartments of the system. We therefore decided to take account of variability in i values (or lack thereof) as an additional information of karst system hydrological functioning, by including a new indicator based on the variance of i.

2.4.3 Signal indicators

The correlational and spectral analyses were performed with a sampling step of 1 day and a maximum offset m of 125 days, according to the suggestions of Mangin (1984) on short-term and long-term analyses. The short-term analysis can be performed on systems that have at least one continuous year of daily discharge recording. As the 10 core systems are all in the same climatic context according to the Köppen-Geiger climate

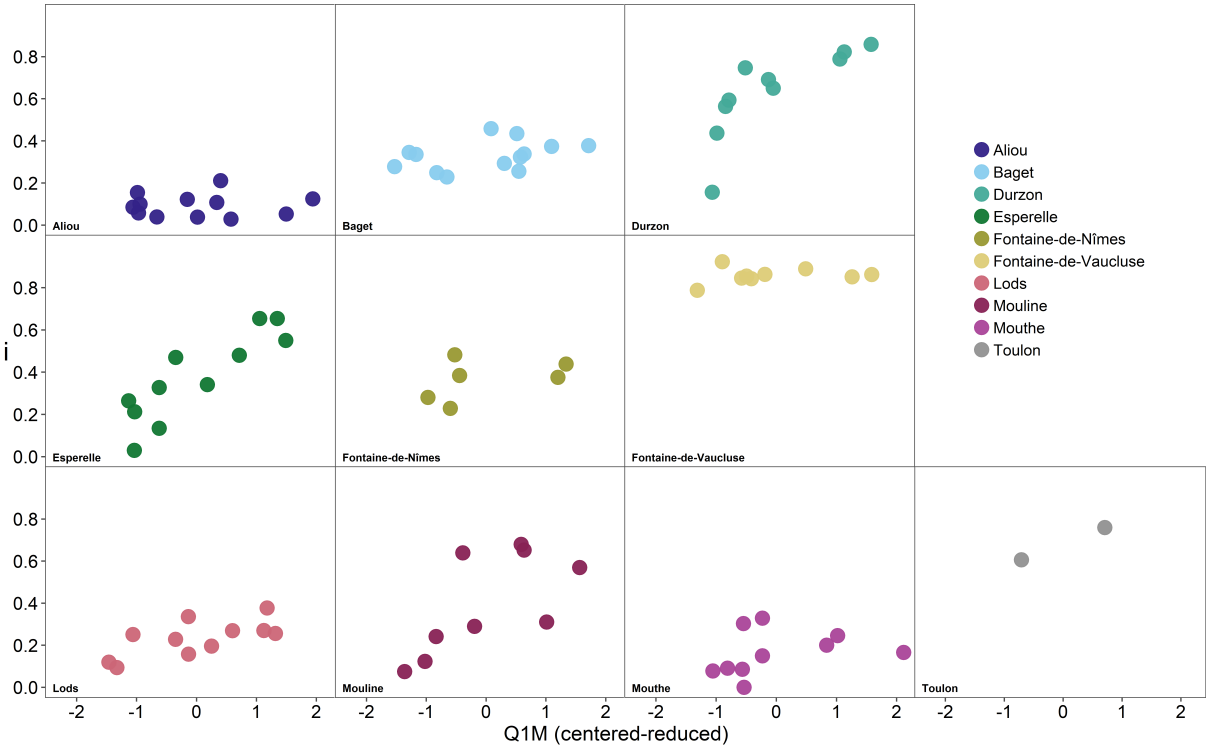


Figure 2.5: Variability in i for the core systems. The standardised ratio between the mean discharge of the month preceding the flood and the mean interannual discharge ($Q1M$) is used as a proxy of the saturation state of the system.

classification, the starting date for the discharge time series was set as the beginning of the hydrological cycle (September 1) to consider the seasonality.

The results highlight the diversity of hydrological functioning of the core systems (Table 2.4). Aliou and Mouthe are very reactive systems that rapidly transfer a response proportional to the intensity and duration of unitary pulses (precipitation), with almost no dampening of the recharge signal. Because they exhibit a fast response to recharge events but are less reactive than the former, Baget, Fontaine-de-Nîmes and Lods are referred as low-inertia systems. Durzon, Esperelle and Mouline correspond to the category of medium-inertia systems, which are able to filter a greater-or-lesser proportion of unitary pulses, and dampen the recharge signal. This category encompasses a wide variety of hydrological functioning, as the medium RT value may translate either (i) a medium inertia or (ii) a high variability of hydrological functioning with both high inertia and low inertia responses resulting in an average RT. Fontaine-de-Vaucluse and Toulon are considered as high-inertia systems, as they have a high filtration capacity and noticeably dampen the recharge signal.

2.4.4 Indicators issued from the analysis of classified discharges

Within the core dataset (Table 2.4), this analysis hints that there is no change in the hydrological functioning of 5 systems (Aliou, Baget, Durzon, Esperelle and Mouline), but that there might be hydraulic or flow properties changes beyond a certain discharge for 4 systems (Fontaine-de-Nîmes, Fontaine-de-Vaucluse, Lods and Toulon). For 4 other systems (Fontaine-de-Nîmes, Lods, Mouthe and Toulon), there might be an overflow outlet, discharge to another system or a temporary storage of water within the karst

System	ME [day]	RT [day]	SBB [day^{-1}]	Interpretation of the classified discharges curve
Aliou	4.6	11.2	0.41	A
Baget	17.6	24.4	0.34	A
Durzon	49.9	41.4	0.24	A
Esperelle	28.3	30.0	0.25	A
Fdn	18.9	17.8	0.33	B
Fdv	81.4	67.8	0.13	B
Lods	13.0	23.5	0.36	B, C
Mouline	57.8	44.6	0.265	A
Mouthe	7.9	11.7	0.415	C
Toulon	101.8	86.1	0.08	B, C

Table 2.4: Results of correlational and spectral simple analyses and interpretation of the classified discharges curve for the core systems. (A) Systems with no apparent, specific functioning; (B) systems in which the hydraulic or flow properties change beyond a certain discharge; (C) systems in which there is an activation of an overflow outlet, a discharge to another system, or a temporary storage of water that occur above a certain discharge.

system when the discharge reaches a certain value. These interpretations are confirmed in the literature for Fontaine-de-Nîmes (Maréchal and Ladouche, 2006), Fontaine-de-Vaucluse (Mangin, 1975), Lods, Mouthe (Cholet, 2017) and Toulon (Lorette et al., 2018).

This analysis, whose interpretation requires prior knowledge of the system or field observations to be fully relevant, does not seem appropriate for the classification of karst systems. The method is based on a strong hypothesis that may not be suitable for all systems. Although some relevant indicators of functioning can be retrieved from this analysis (e.g. activation of an overflow outlet, flow to another system), the interpretation is overly influenced by (i) the quality of the discharge-water height relationship, and (ii) the subjective vision of the operator who is performing the analysis. Moreover, when there is poor or no prior knowledge of the functioning of a system, it is very difficult to identify the specific functioning behind a bending point on the curve (e.g. the difference between a bending point due to the activation of an overflow outlet, or due to uncertainties on ungauged discharges). Thus, we choose to not include the indicators of functioning retrieved from this analysis in the proposed methodology for the classification of karst system hydrological functioning.

2.5 Multivariate analyses

The aim of this section is to gain insights into the dataset and the relations between indicators. We applied two unsupervised techniques on a dataset consisting of 9 variables and 10 observations. The observations correspond to the 10 core karst systems, and the variables are relevant quantitative indicators of functioning resulting from the application of the different methods of discharge time series analysis (excluding the indicators resulting from the analysis of classified discharge). The selected indicators are k_{max}, i_{mean}, IR, α_{mean}, ME, RT, SBB, CV and SVC. Terms "max" and "mean" correspond to, respectively, maximum and average values of the indicator over all the recession analysed.

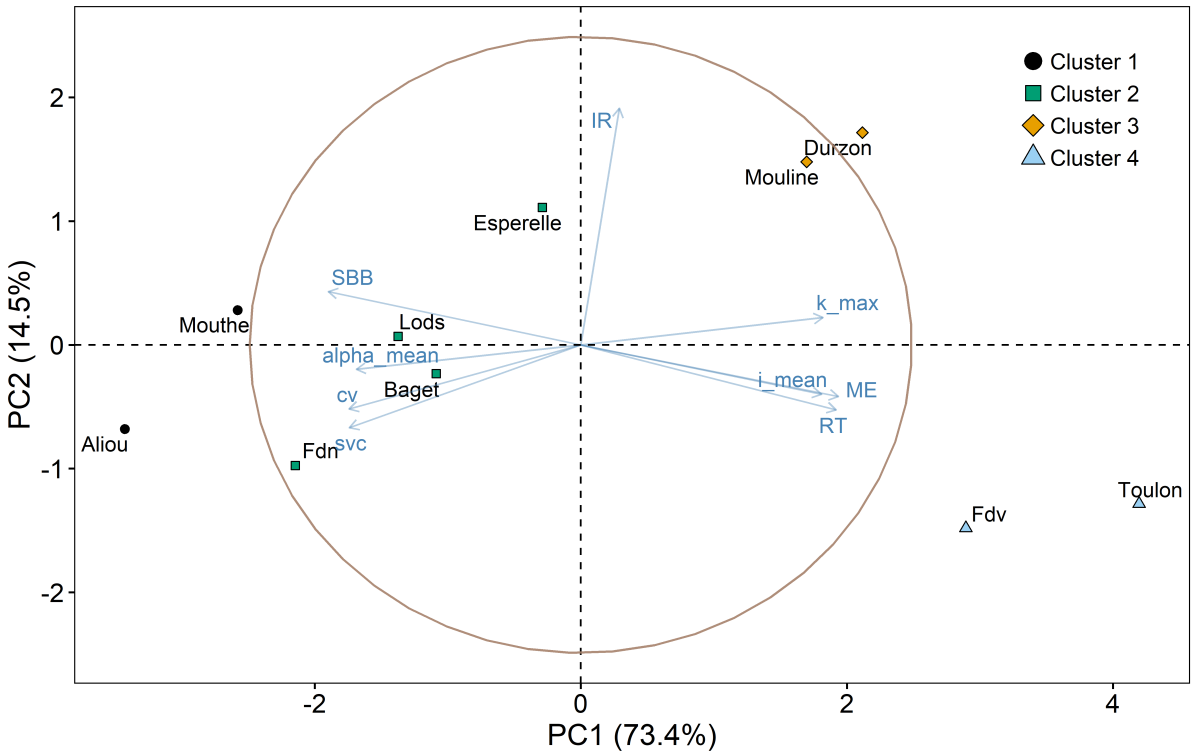


Figure 2.6: Biplot of the first two principal components PC1 and PC2 resulting from the PCA performed on the dataset (9 variables for 10 observations). The points correspond to the PC1 and PC2 scores of each observation (i.e. karst systems). The arrows represent the correlation of the variables (i.e. indicators of functioning) with PC1 and PC2. The brown circle indicates the theoretical maximum extent of the arrows. Clusters (colours) are referring to the results of the hierarchical clustering performed on the dataset (see section 2.5.2).

2.5.1 Principal component analysis

2.5.1.a Principle

Principal component analysis (PCA) is a multivariate method that aims to reduce the dimensions of an observation space by producing "principal components", which are linear combinations of initial variables of a dataset that retain the most possible variation. Principal components are uncorrelated with each other (i.e. orthogonal to the previous one) and are ordered according to the amount of variance explained by the combination (Everitt and Hothorn, 2011). The results of a PCA can be seen as a small number of new variables that contains most of the information of a large number of initial variables. The interpretation of a principal component is realised by looking at the correlation between initial variables and the component, i.e. assessing the contribution of each variable. This analysis provides information on trends in the dataset and allows identifying eventual complementarity between initial variables.

2.5.1.b Results

The results of the PCA are presented as a biplot of the first factorial plane (Figure 2.6), which explains 87.9 % of the total variance of the dataset.

The first principal component (PC1) is strongly correlated with all indicators except

IR: (i) k_{max} informs about the capacity of dynamic storage, (ii) i_{mean}, ME, RT and SBB are related to the capacity to attenuate the precipitation signal, (iii) α_{mean} characterises the draining dynamic of the capacitive function and (iv) CV and SVC are about flow dynamics. Together, these indicators can be seen as characterising the inertia of a system, which would correspond to the first principal component (PC1, 73.4 % of variance explained). The second principal component (PC2, 14.5 % of variance explained) is mainly correlated with IR and thus would reflect the variability of the hydrological response of a system. The third principal component (PC3, 5.8 % of variance explained) is most correlated with α_{mean}, which characterises the draining dynamic of the capacitive function of a system.

It is worth mentioning that the high correlation between several input parameters (e.g. CV and SVC) may have a substantial effect on the PCA results, particularly on the variance explained by the principal components (i.e. PC2 and PC3 could have had a higher variance), by overemphasising the contribution of redundant indicators.

Systems from the core dataset are mostly scattered along PC1 with rather inertial systems on the positive part (e.g. Fontaine-de-Vaucluse, Toulon) and reactive systems on the negative part (e.g. Aliou, Mouthe), and to a lesser extent along PC2 with "highly variable hydrological response" systems on the positive part (e.g. Esperelle, Durzon, Mouline) and "more regular, steady hydrological response" systems on the negative part (e.g. Fontaine-de-Vaucluse, Toulon).

2.5.2 Clustering

2.5.2.a Principle

The purpose of data clustering is to identify clusters that contain observations or objects with similar characteristics (Govender and Sivakumar, 2020; Jain et al., 1999). Clustering analysis can be used to identify archetypes, and offer a better understanding of the structure within a dataset (Halkidi et al., 2001). This technique is considered unsupervised, because it is not based on predefined classes or examples that would give an idea of the structure of the dataset (Berry and Linoff, 1997).

We selected a Ward hierarchical clustering method for performing the analysis, which consists of a succession of binary fusions that minimise between-cluster variance until one cluster remains (Murtagh and Legendre, 2014). The Ward distance is equal to (Tufféry and Riesco, 2011):

$$D(A,B) = \frac{d(a,b)^2}{n_A^{-1} + n_B^{-A}} \tag{2.4}$$

With D the Ward distance between two clusters A and B that have centers of gravity a and b and frequencies n_A and n_B. The analysis was realised with standardised data and Euclidean distance as measure of dissimilarity, which is calculated with the following equation:

$$d_{euc}(x,y) = [\sum_{j=1}^{d}(x_j - y_j)^2]^{\frac{1}{2}} \tag{2.5}$$

With d the distance between two points x and y of a d-dimensional dataset, and x_i and y_i the values of the j^{th} attribute of x and y, respectively (Gan et al., 2007).

This method is suitable for clusters of different sizes and shapes and provides a graphical representation (dendrogram) which helps for understanding the clusters structure and how they are connected. The main advantages over a non-hierarchical method are (i) that it is not necessary to know the number of clusters prior to the analysis, and (ii) that the results do not depend on the choice of initial clusters (Tufféry and Riesco, 2011).

A way to assess whether a variable is relevant to characterise a cluster is to realise a value-test or v-test (Lebart et al., 1984). For a given quantitative variable, it involves in comparing the mean $\bar{x_k}$ for this variable of a particular cluster k to the overall mean $\bar{X}$, with the formula:

$$v = \frac{\bar{x_k} - \bar{X}}{\sqrt{\frac{s^2}{n_k}\left(\frac{N-n_k}{N-1}\right)}} \tag{2.6}$$

With v the result of the test, s^2 the overall standard deviation, N the total number of observations and n_k the number of observations in the cluster k. A value of $|v|$ greater than 1.96 corresponds to a p-value less than 0.05, which rejects the following hypothesis: the mean of the particular cluster is equal to the overall mean (Lebart et al., 1984). In this case, the variable is relevant to describe the group of observations in the cluster.

2.5.2.b Results

Results are presented as a dendrogram. Four relevant clusters (corresponding to two majors clusters) have been identified (Figure 2.7). There is a great similarity between clusters and PCA results (Figure 2.6).

A v-test was performed to assess the indicators that best characterise each cluster (Appendix 2.C). The major clusters A and B are differentiated on the basis of the following indicators: k_{max}, ME, CV, RT, i_{mean}, SBB, SVC and α_{mean} (in order of importance), which are related to the capacity of dynamic storage and inertia of a system. According to the sign of the v-test results, cluster A includes reactive systems, with low to medium capacity of dynamic storage and cluster B includes inertial systems, with higher capacity of dynamic storage. The systems from cluster 1 are characterised by a high α_{mean} value, corresponding to a fast draining of the capacitive function. None of the indicators clearly characterise the systems in cluster 2. The systems in cluster 3 are characterised by a high IR value, corresponding to a high variability of the hydrological response. The systems in cluster 4 are characterised by very high RT, ME and i_{mean} and a very low SBB, corresponding to a high attenuation capacity of the precipitation signal.

2.5.3 Confrontation of the results with the actual knowledge of the functioning of the systems

PCA and clustering gave similar insight into the functioning of the 10 core karst systems. Two major clusters were identified (A and B, Figure 2.7): (i) cluster A is characterised by systems with a highly reactive functioning, which can be divided into two sub clusters (1 and 2) by looking at the draining dynamic of the capacitive function; (ii) cluster B is characterised by systems with inertial functioning, and can be divided into two sub clusters (3 and 4) by looking at the variability of the hydrological functioning (Figure 2.6, Figure 2.7).

The hydrological response of Aliou and Baget corresponds to well karstified systems (Mangin, 1984), which is consistent with their position in the major cluster A.

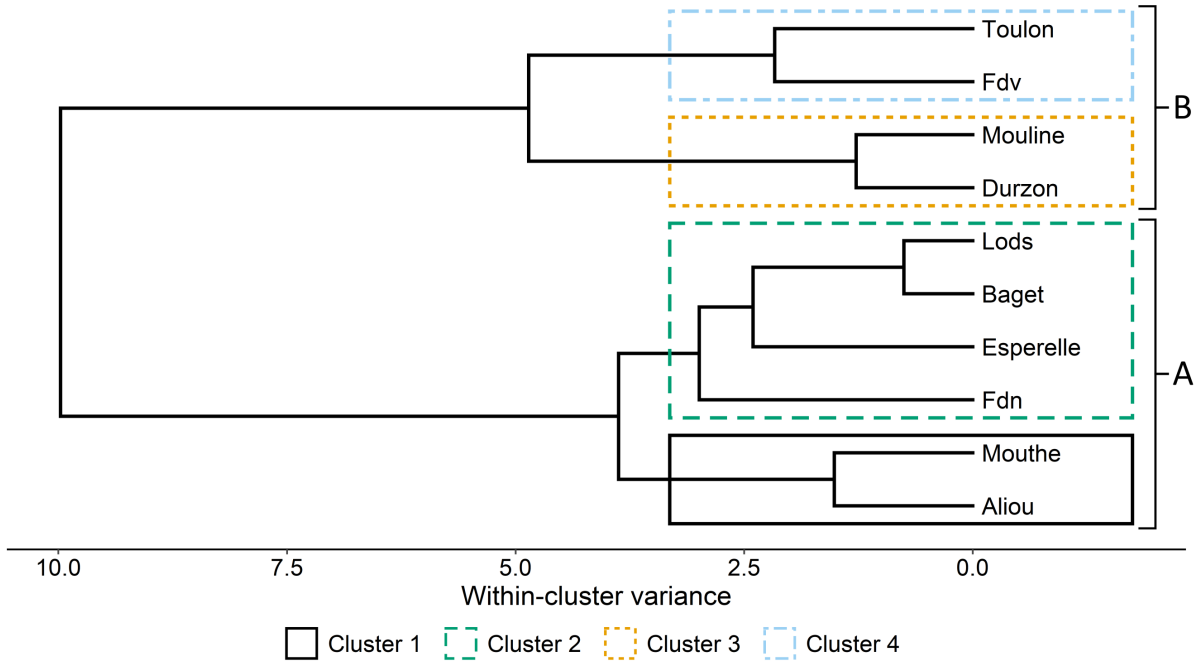


Figure 2.7: Dendrogram resulting from the Ward's hierarchical clustering performed on the dataset (9 variables for 10 observations). The X-axis gives the distance between observations and/or clusters. The nodes (or vertical bars) indicate at which value two observations and/or clusters are merged. Four relevant clusters (1, 2, 3 and 4) are identified with different colours, and "major" clusters A and B correspond to the merging of clusters 1-2 and clusters 3-4, respectively.

The karst conduits network induces floods of short duration, that can have a high amplitude (Mangin, 1975), with a response time estimated to 7 hours after a precipitation event for Aliou and 14 hours for Baget (Sivelle et al., 2019). Sivelle et al. (2019) also found lower transfer coefficients for Baget in their reservoir modelling study. These results on the two systems support the difference between both clusters (1 for Aliou and 2 for Baget), as Aliou is characterised by a faster draining dynamic.

Cross-correlational analyses between precipitation and discharge performed by (Cholet, 2017) on Lods and Mouthe systems showed low response times (17 and 10 hours, respectively) and high peak values (0.3 and 0.41, respectively), which are characteristics of a very reactive functioning. On Mouthe system, the higher draining dynamic of the capacitive function (cluster 1) is consistent with both the lower response time and higher peak value of the cross-correlation function than the ones obtained on Lods system (cluster 2).

Fontaine-de-Nîmes is a reactive karst system with a moderate degree of karstification (Maréchal and Ladouche, 2006). The hydrological response is fast due to a high infiltration rate and a fast water transfer in a well-developed conduits network (Maréchal et al., 2008), which corresponds to the characteristics of cluster A . Fleury et al. (2013) found that the draining of the saturated zone was slow, which is consistent with the position of the system in cluster 2.

Esperelle karst system is described as significantly fractured (Moussu, 2011), and characterised by both a high impulse response height and dampened recession (Pinault et al., 2001). This description is consistent with the one of cluster 2, which consists of

reactive systems with moderate draining of the capacitive function.

Durzon, Mouline, Fontaine-de-Vaucluse and Toulon systems are included in the cluster B. Mouline is considered as a complex karst system, with long response times despite presence of flush flow effect that induces a quick transfer during winter (Pinault et al., 2001). Although there is no explanation for quick transfer in Durzon system, we can suppose that its functioning is similar to Mouline as it is located in the same area. The existence of both slow and fast dynamics on these systems depending on recharge event and/or geometrical structure highlights a high variability of hydrological functioning, which is characteristic of cluster 3. Fontaine-de-Vaucluse and Toulon are differentiated in another cluster (cluster 4) due to the low variability of their hydrological functioning. The high inertia and homogeneous response of the Fontaine-Vaucluse system can be explained by the thickness of the non-saturated zone (800 m in average) and the large area of its catchment, estimated to be about 1160 km^2 (Ollivier et al., 2019). The Toulon system, defined as a complex karst system by Lorette et al. (2018), is a multi-layer system with the discharge of an unconfined, fractured and karstified aquifer that is permanently fed by another confined aquifer. This continuous alimentation can explain the mostly homogeneous response of the system.

2.6 Classification of karst hydrological functioning

In this section, we first present the classification based on the results of multivariate analyses. The classification is then applied on both core and complementary datasets to assess the relevance of the approach, regarding (i) the coherence with the well-known hydrological functioning of the karst systems in the core dataset, and (ii) the distribution of karst systems among the different classes from a worldwide perspective.

2.6.1 Classification of karst systems according to various types of hydrological functioning

The exploration realised with PCA and clustering confirmed the expected relations between the functioning of karst systems and indicator values. In both methods, karst systems are first differentiated from two main aspects: their capacity of dynamic storage and their capacity to attenuate the precipitation signal. The former is expressed through k_{max} and the latter through i_{mean}, RT, ME and SBB (n.b. the absolute values of the Pearson correlation coefficients of each pair are greater than 0.85, suggesting strong correlations). We chose the indicator RT to characterise the capacity to attenuate the precipitation signal (further referred as global inertia, as it considers the overall organisation of flows in the system, the shape and the dimensions of the catchment and the saturation state of the system). Indeed, this indicator is more relevant than (i) i_{mean}, which is biased by the number of available recession curves, and (ii) ME and SBB because their assessments are somehow questionable as they rely on an arbitrary threshold and a subjective evaluation, respectively. A second element of differentiation between karst systems is the draining dynamic of the capacitive function of a system with the α_{mean} indicator. In this case, the mean of α values seems relevant due to the relatively low amplitude of the values for a given system (Figure 2.4C). A third element of differentiation between systems is the variability of the hydrological response by quantifying the variability of i with the IR indicator.

Class	k_{max}	α_{mean}	IR	Capacity of dynamic storage	Draining of the capacitive function	Variability of the hydrological functioning
C1	≤ 0.4	≥ 0.03	≥ 0.25	Poor	Fast	Substantial
C2	≤ 0.4	≥ 0.03	< 0.25	Poor	Fast	Low
C3	≤ 0.4	< 0.03	≥ 0.25	Poor	Moderate	Substantial
C4	≤ 0.4	< 0.03	< 0.25	Poor	Moderate	Low
C5	> 0.4	< 0.03	≥ 0.25	Noticeable	Slow	Substantial
C6	> 0.4	< 0.03	< 0.25	Noticeable	Slow	Low

Table 2.5: Characterisation of the karst systems hydrological functioning for the six defined classes, in terms of capacity of dynamic storage, draining dynamic of the capacitive function, and variability of the hydrological response, with the corresponding indicator values.

The proposed classification thus relies on the following three main characteristics of the hydrological functioning: (i) the capacity of dynamic storage, (ii) the draining dynamic of the capacitive function, and (iii) the variability of the hydrological response. These characteristics are assessed using (i) k_{max}, (ii) α_{mean}, and (iii) IR indicators, respectively. Based on these 3 distinct characteristics of hydrodynamic behaviour, six classes are proposed to discriminate the hydrological functioning of karst systems (Table 2.5, Figure 2.8): (i) C1: Poor capacity of dynamic storage, with fast draining of the capacitive function and substantial variability of hydrological functioning; (ii) C2: Poor capacity of dynamic storage, with fast draining of the capacitive function and low variability of hydrological functioning; (iii) C3: Poor capacity of dynamic storage, with moderate draining of the capacitive function and substantial variability of hydrological functioning; (iv) C4: Poor capacity of dynamic storage, with moderate draining of the capacitive function and low variability of hydrological functioning; (v) C5: Noticeable capacity of dynamic storage, with slow draining of the capacitive function and substantial variability of hydrological functioning; and (vi) C6: Noticeable capacity of dynamic storage, with slow draining of the capacitive function and low variability of hydrological functioning.

We chose to not include the indicator RT in the classification methodology, as it seems that the global inertia of a karst system is relatively independent of its main characteristics of functioning, especially for systems in C3, C4, C5, and C6 (Figure 2.9B).

The proposed classification is based on recession parameters derived from Mangin's model for recession simulation. This model gave good simulations results over the 78 catchments considered in this study. However, the good performance of this model should be checked before any use of the classification over other systems. To gain insights into the functioning of such systems, we recommend the operators to make use of correlational and spectral analyses, or analysis of classified discharges, which can already give relevant information about global inertia and functioning of a system.

2.6.2 Application of the proposed methodology to 78 karst systems

The classification was applied on the complementary dataset presented in section 2.2.2 and composed of 68 karst systems, plus the 10 core systems for a total of 78 karst systems (Figure 2.9A).

The most represented class is C3, which represents systems with a very low to medium

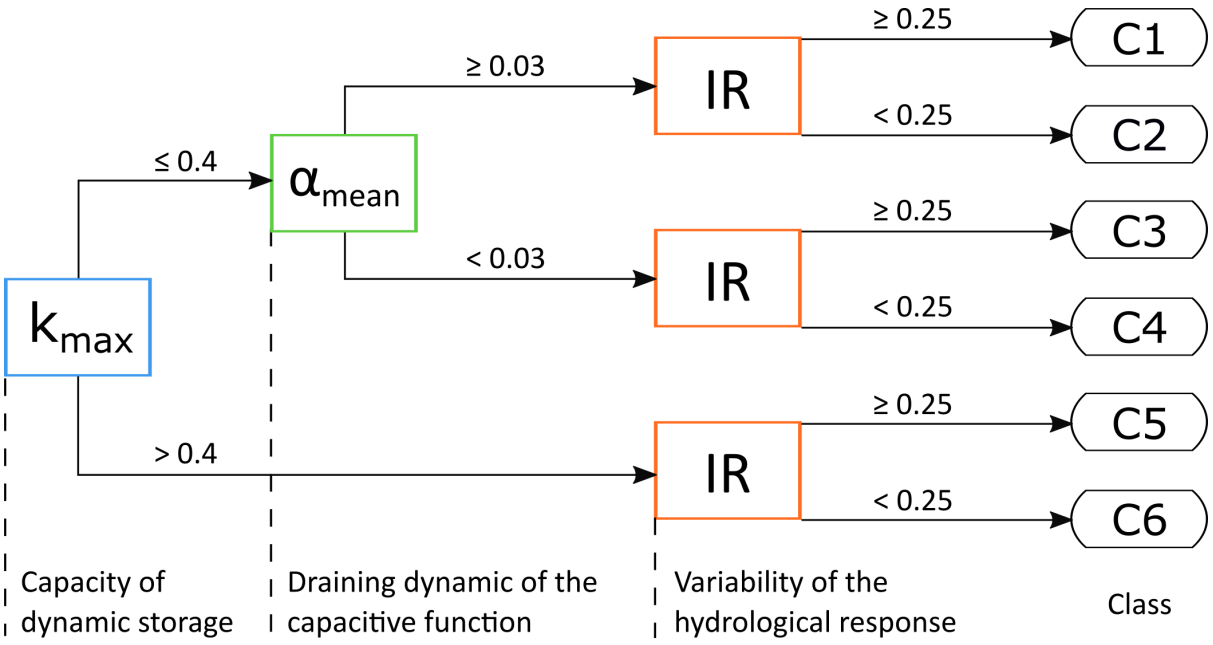


Figure 2.8: Flowchart for the classification of karst systems based on three indicators issued from the recession curves analysis.

capacity of dynamic storage, a moderate to slow draining of the capacitive function and a medium to high variability of the hydrological response. About 74 % of the systems are characterised with a very low to medium capacity of dynamic storage (C1, C2, C3 and C4), 26 % have a fast draining of their capacitive function (C1 and C2) and 67 % have a medium to high variability of the hydrological response (C1, C3 and C5).

RT was calculated for each system (except for 1 system in C3 and 4 systems in C2 with too much gaps in the time series), with means of 19.5, 13.9, 31.8, 41.4, 43.8 and 46.3 days for, respectively, C1, C2, C3, C4, C5 and C6 (Figure 2.9B). The small increase of *RT* throughout the classes is consistent given the functioning description of each class (C1 being the most reactive and C6 the most inertial). The smaller mean in C2 is related to the structure of the classification: a high *IR* in C3 and C5 means that the hydrological response can be more reactive than expected, whereas in C1 it means that the hydrological response can be more inertial than expected. It results in a higher *RT* mean in C1 (over C2) as it includes systems with potential inertial responses. Finally, we observed that *RT* is biased for systems with long dry periods (Saint-Pierre or Lez systems, the latter being under anthropic influence), thus it is suggested to not calculate this indicator for these systems.

The systems are spread across all classes and types of hydrological functioning. It means that, even applied on a wider dataset, there is a relative representativeness of all classes and types of hydrological functioning. The spread of the 10 core systems between the six classes somehow confirms the respect of the second criteria for spring selection (diversity of the hydrological functioning among the karst systems, see section 2.2.1) and the relevance of the proposed classification.

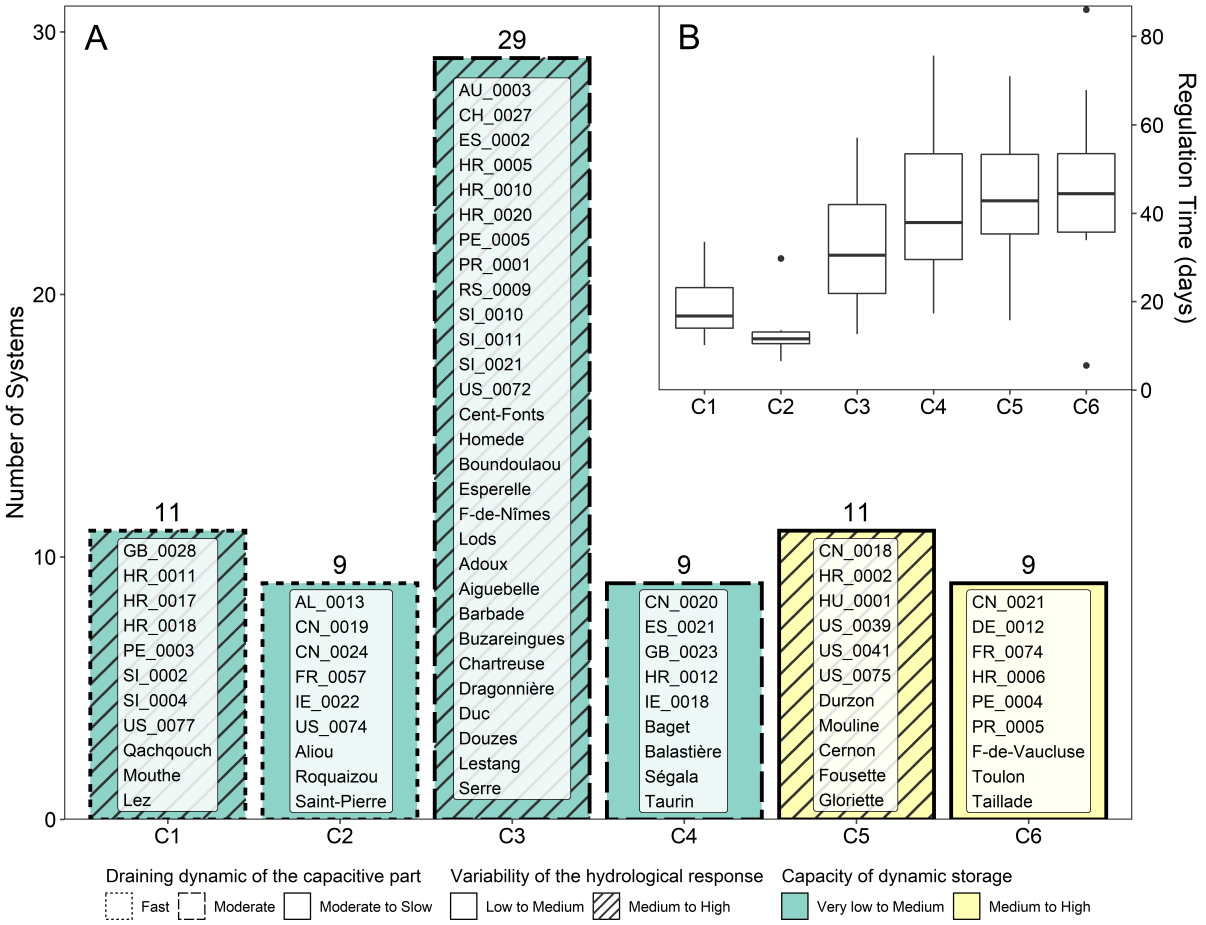


Figure 2.9: Results of the classification for the systems in the core and complementary datasets. (A) The 6 classes are depicted on the x-axis and the number of systems by class on the y-axis. The colour is related to the capacity of dynamic storage, the outline is related to the draining dynamic of the capacitive function and the pattern is related to the variability of the hydrological response. (B) On the right-upper side, the boxplot shows the distribution of RT among the systems.

2.7 Discussion

The aim of this section is to take a step back on the methodology by questioning some of its limitations and assessing their impact on the relevance of the classification.

2.7.1 Influence of the length of the time series on the classification

We performed a sensitivity analysis in order to assess the reliability of the classification regarding discharge time series of short length. For 9 systems of the 10 core systems (Toulon was not considered as there is only 5 years of monitoring), we defined 7 scenarios that range from 1-year (Y1) to 7-years (Y7) length of the discharge time series. The classification methodology was then applied on each scenario in order to compare the results to those obtained with the full-length (FL) discharge time series (Table 2.6).

The accuracy of each indicator increases with the length of the time series (Figure 2.10). k_{max} is the most consistent with a steady decrease throughout the years in the deviation to the k_{max} value obtained for the full-length time series (FL indicator). α_{mean} becomes more stable and probably more reliable from Y5 even though it does not show any significant increase in precision for Y6 and Y7. IR is highly uncertain for the shorter time series (<Y4) but stabilizes from Y5 and become relevant for Y6 and Y7.

Aliou, Baget, Durzon and Mouline were correctly classified from Y2 (Table 2.6), Esperelle and Mouthe from Y5 and Y6, respectively, while Fontaine-de-Nîmes, Fontaine-de-Vaucluse and Lods were not consistently classified at all. The results emphasise the benefits of long-term monitoring (as there are more chances of observing multiple flood events, as well as different meteorological conditions), but also highlight several limitations:

i. The classification is not reliable if only one recession curve is considered, as IR would be 0.

ii. When the indicator is close to the threshold, it can mislead the classification (e.g. Esperelle and Fontaine-de-Nîmes, for which the variability in α_{mean} can be explained by either a highly variable or a complex hydrological functioning, respectively).

iii. As the maximum observed discharge Q_{max} can change over time, this can affect the recession curves selection, since only recessions with a peak flood discharge of at least one tenth of the maximum observed discharge are considered for the analysis (e.g. Fontaine-de-Nîmes with Q_{max} that went from 8.2 to 16.5 m^3 s^{-1} from Y1 to Y3).

iv. As the mean interannual discharge Q_{mean} may vary over years, this can induce an evolution of k_{max} value and thus modify the class of the karst spring when the value is close to the threshold (e.g. Fontaine-de-Vaucluse, for which k_{max} oscillate around the 0.4 threshold with changes in Q_{mean} during the years, despite being issued from the same recession curve).

Based on these results, we suggest working with at least 3-years length discharge time series for the classification. These 3 years should be taken as a guideline and may differ notably depending on the system's dynamics. Indeed, very reactive systems may require only a few years for a definitive classification thanks to their high hydrodynamic variability, while the minimum length of the discharge time series required to reach satisfying

System	Indicator	Y1	Y2	Y3	Y4	Y5	Y6	Y7	FL
Aliou	k_{max}	0.02	0.01	0.02	0.02	0.02	0.02	0.02	0.02
	α_{mean}	0.127	0.127	0.115	0.090	0.084	0.083	0.086	0.067
	IR	0.02	0.02	0.04	0.18	0.18	0.18	0.18	0.18
	Class	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**
	R. num	2	2	3	6	7	8	10	12
Baget	k_{max}	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.08
	α_{mean}	0.032	0.028	0.026	0.026	0.028	0.028	0.027	0.021
	IR	/	0.07	0.07	0.07	0.09	0.09	0.13	0.23
	Class	/	**4**	**4**	**4**	**4**	**4**	**4**	**4**
	R. num	1	2	3	3	5	6	7	13
Durzon	k_{max}	0.54	0.63	0.68	0.72	0.75	0.76	0.75	0.78
	α_{mean}	0.005	0.004	0.004	0.004	0.004	0.004	0.004	0.004
	IR	/	0.51	0.68	0.68	0.68	0.68	0.68	0.70
	Class	/	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	R. num	1	3	5	7	8	9	10	10
Esperelle	k_{max}	0.14	0.09	0.10	0.11	0.12	0.12	0.12	0.13
	α_{mean}	0.030	0.034	0.034	0.034	0.029	0.029	0.027	0.022
	IR	0.55	0.57	0.57	0.57	0.57	0.57	0.57	0.62
	Class	**3**	1	1	1	**3**	**3**	**3**	**3**
	R. num	3	5	7	7	9	9	10	11
Fdn	k_{max}	0.01	0.06	0.04	0.04	0.05	0.05	0.05	0.09
	α_{mean}	0.035	0.023	0.035	0.035	0.031	0.031	0.031	0.021
	IR	/	0.36	0.21	0.21	0.25	0.25	0.25	0.25
	Class	/	**3**	2	2	1	1	1	**3**
	R. num	1	2	2	2	3	3	3	6
Fdv	k_{max}	0.38	0.35	0.37	0.41	0.35	0.35	0.37	0.40
	α_{mean}	0.003	0.004	0.005	0.005	0.005	0.005	0.005	0.005
	IR	/	0.10	0.10	0.10	0.13	0.13	0.13	0.14
	Class	/	4	4	**6**	4	4	4	**6**
	R. num	1	2	3	4	5	6	7	9
Lods	k_{max}	0.03	0.08	0.09	0.08	0.08	0.08	0.07	0.09
	α_{mean}	0.029	0.018	0.018	0.016	0.021	0.020	0.021	0.021
	IR	/	0.02	0.02	0.18	0.18	0.22	0.22	0.28
	Class	/	4	4	4	4	4	4	3
	R. num	1	3	3	6	7	10	10	11
Mouline	k_{max}	0.58	0.49	0.54	0.55	0.57	0.56	0.56	0.59
	α_{mean}	0.003	0.005	0.005	0.005	0.005	0.005	0.005	0.005
	IR	/	0.43	0.58	0.59	0.59	0.59	0.59	0.60
	Class	/	**5**	**5**	**5**	**5**	**5**	**5**	**5**
	R. num	1	3	6	7	8	8	8	9
Mouthe	k_{max}	0.01	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	α_{mean}	0.060	0.060	0.064	0.064	0.064	0.061	0.061	0.059
	IR	0.08	0.11	0.11	0.11	0.11	0.25	0.25	0.33
	Class	2	2	2	2	2	**1**	**1**	**1**
	R. num	3	5	6	6	6	8	8	10

Table 2.6: Results of the classification performed on the different scenarios of discharge time series of different length (from 1 to 7 years, Y1 to Y7). For each scenario, there are details about the results of indicators k_{max}, α_{mean}, IR, the corresponding class, and the number of recession curves considered in the analysis. The full-length (FL) column corresponds to the results on the classification on the whole discharge time series.

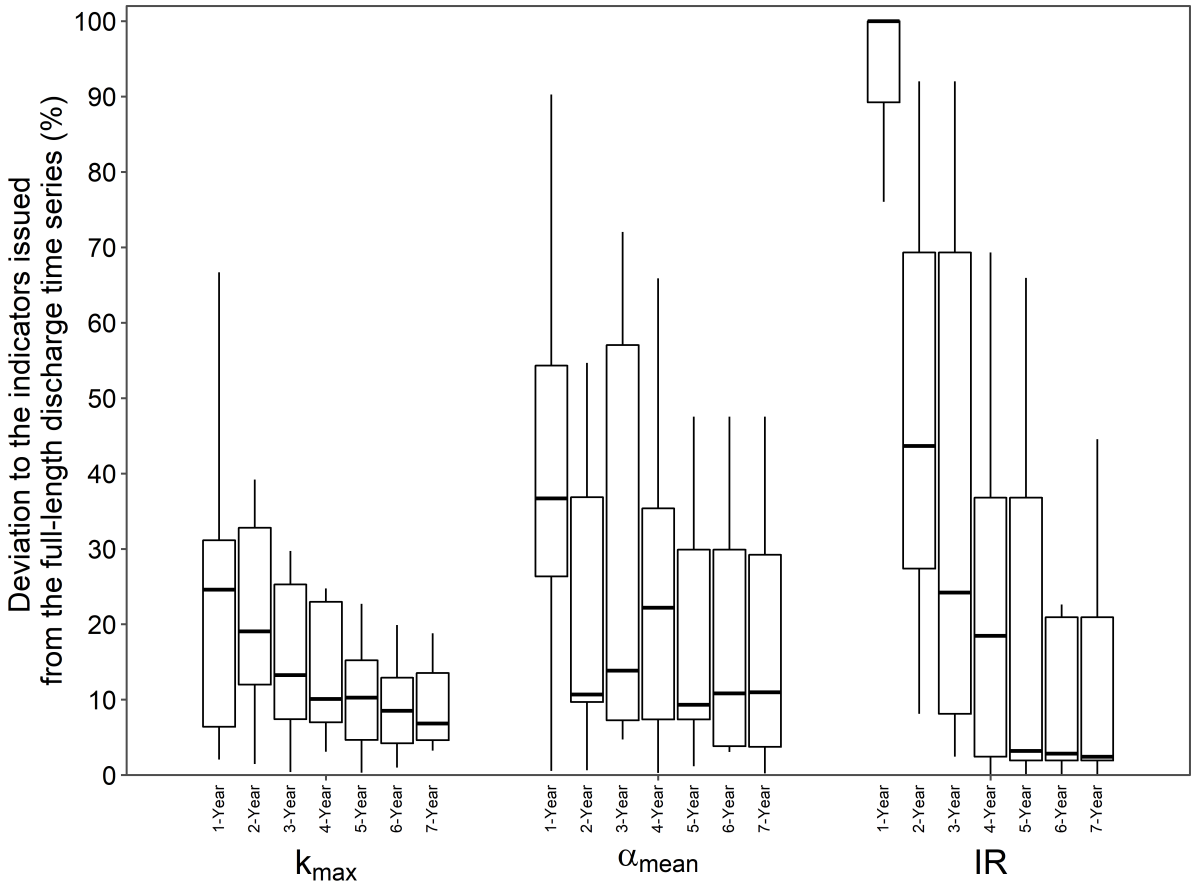


Figure 2.10: Boxplot of the mean relative error of each indicator across 9 core systems and for the different scenarios, regarding the indicators issued from the analysis of the full-length discharge time series. Each indicator on the x-axis is associated with 7 boxes that are sorted from shorter (left) to wider (right) length.

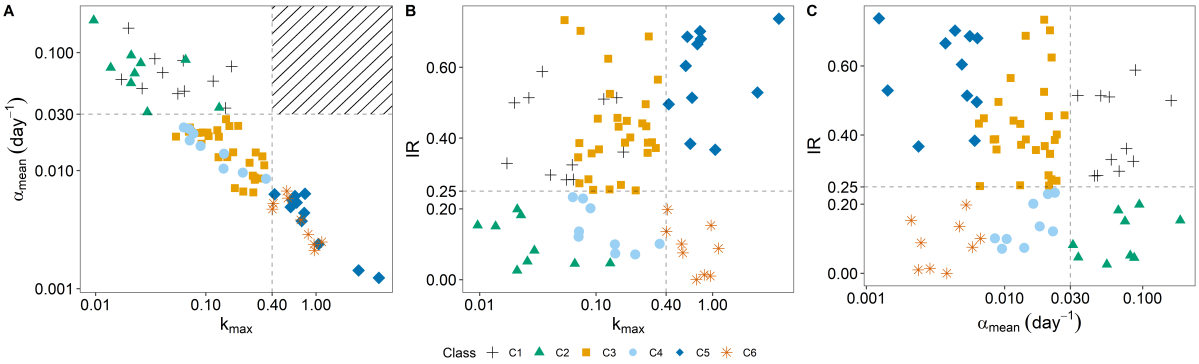


Figure 2.11: Distribution of the 78 systems according to each pair of indicators: (A) α_{mean} vs. k_{max}, (B) IR vs. k_{max}, and (C) IR vs. α_{mean}. The colours correspond to the different classes.

classification may increase for very inertial systems. However, it remains appropriate to work with shorter time series when there is no alternative. Our analysis shows that 7 out of 9 systems are correctly or almost correctly classified at Y2 if we include those that are close to the threshold (Table 2.6).

2.7.2 Evaluation of the distance between a system and other classes

The uncertainties related to either (i) the length of the discharge time series, or (ii) the indicators that are close to the threshold, can be addressed by estimating the distance to the other classes. The distance D_{s-C} of a system s to an adjacent class C is measured with the following equation:

$$D_{s-C} = \frac{|I_{th} - I_{calc}|}{I_{th}} \tag{2.7}$$

Where I_{calc} corresponds to the calculated value of a given indicator and I_{th} corresponds to its threshold value. The indicator to consider in the calculation is the one that is critical for the class definition (i.e. corresponding to the junction in the flowchart). A distance D lower or equal to 0.1 (10 %) can be considered as close to the threshold. As the distance increase, the system is not likely to be related with the involved class.

In relation to the structure of the classification, there is no need to calculate the distance between classes 1-5 and classes 2-6, as it is highly uncertain that a system has both a k_{max} higher than 0.4 and a α_{mean} greater than or equal to 0.03 (hatched area, Figure 2.11A). The formula allows for calculating the distance between (i) classes 1-2, 3-4 and 5-6 with IR, (ii) classes 1-3 and 2-4 with α_{mean}, and (iii) classes 3-5 and 4-6 with k_{max}. The distance between diagonal classes can be calculated by applying the Pythagorean Theorem:

$$D_{s-C} = \sqrt{D^2_{s-Cx} + D^2_{s-Cy}} \tag{2.8}$$

Where x and y correspond to two different classes. The formula allows for calculating distance between (i) classes 1-4 and 2-3 with α_{mean} and IR (Figure 2.11C), and (ii) classes 3-6 and 4-5 with k_{max} and IR (Figure 2.11B). These notions of distances can be more easily appreciated on the Figure 2.12, which shows the 3-dimensional distribution of the karst systems in the k_{max}, α_{mean} and IR space.

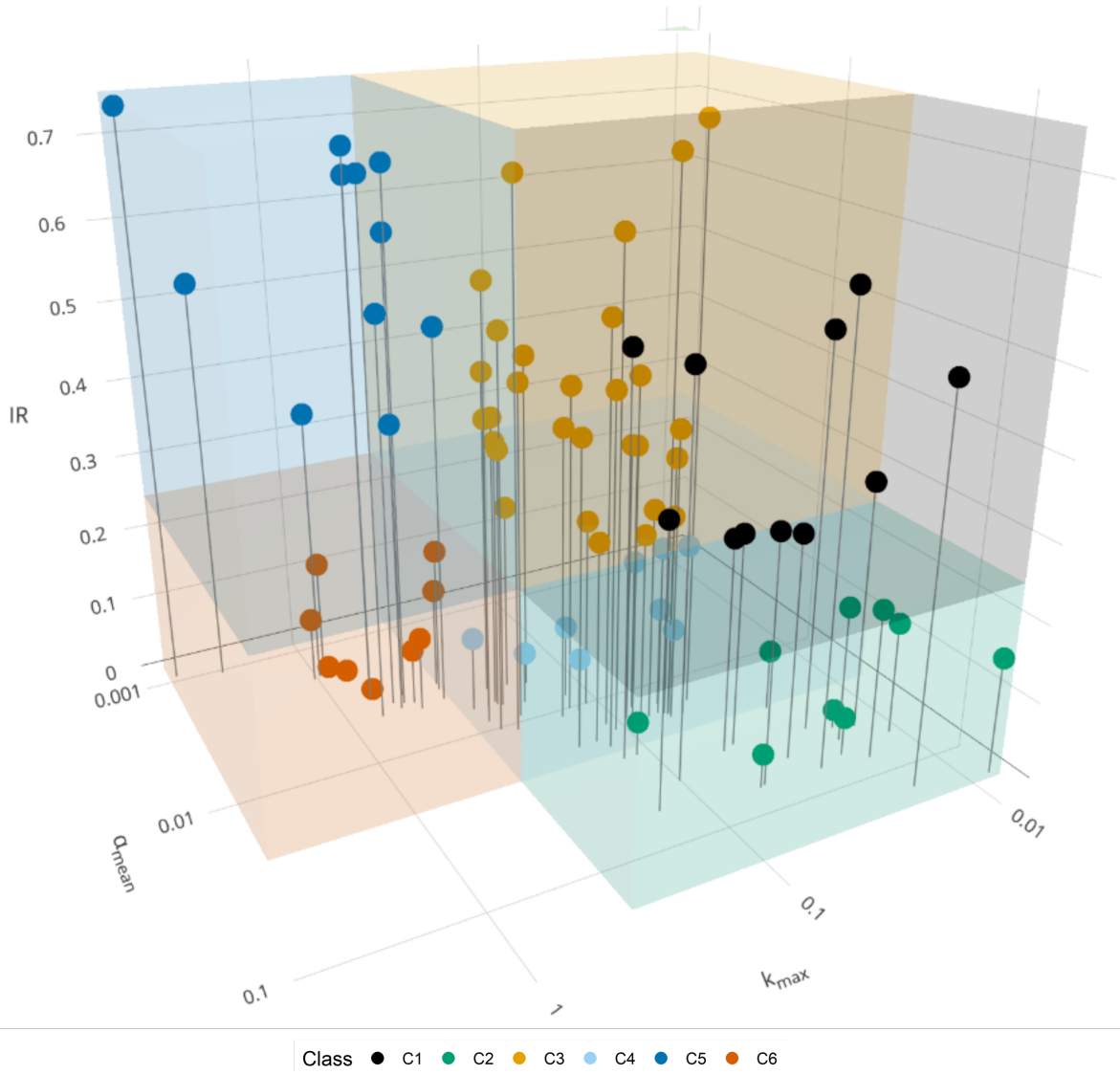


Figure 2.12: 3D distribution of the 78 systems according to each indicator. The associated rectangles show the boundaries of each class. The colours correspond to the different classes.

System	Class	Distance to					
		C1	C2	C3	C4	C5	C6
Aliou	C2	0.27	**0**	1.25	1.22	/	/
Baget	C4	0.32	0.31	0.08	**0**	0.81	0.81
Durzon	C5	/	/	0.95	2.04	**0**	1.81
Esperelle	C3	0.27	1.52	**0**	1.50	0.68	1.64
Fdn	C3	0.30	0.30	**0**	0.01	0.76	0.76
Fdv	C6	/	/	0.46	0.004	0.46	**0**
Lods	C3	0.31	0.33	**0**	0.13	0.78	0.80
Mouline	C5	/	/	0.47	1.49	**0**	1.42
Mouthe	C1	**0**	0.31	0.98	1.03	/	/
Toulon	C6	/	/	1.48	1.43	0.39	**0**

Table 2.7: Distance to the other classes for the 10 core karst systems.

To complement the discussion in Section 2.7.1 related to the "close to the threshold" systems, we calculated the distance to other classes for the 10 core karst systems (Table 2.7). The results highlight that (i) Baget is close to C3, (ii) Fontaine-de-Nîmes is close to C4 and (iii) Fontaine-de-Vaucluse is close to C4. Regarding the 78 systems, only 12 systems are close to a threshold with a distance to another class lower or equal than 0.1.

This distance measure allows reflecting on the classification results by highlighting potential threshold issues. Thus, we can tell if a system is clearly into a said class, or is close to one or two classes.

2.7.3 Beyond the classification

Correlational, spectral and classified discharges analyses can be performed post-classification to really exploit the maximum of information that can be obtained by analysing the discharge time series, in order to get a deeper knowledge of the functioning of a system.

We propose to use RT as an additional indicator to complement the above-described classification and to gain further insights into the global inertia of a karst system. We defined four ranges: (i) lower or equal than 15 days; (ii) greater than 15 days and lower or equal than 30 days; (iii) greater than 30 days and lower or equal than 45 days; and (iv) greater than 45 days, which are referred as (i) low, (ii) medium, (iii) high, and (iv) very high global inertia, respectively.

In a similar manner, we propose to perform the visual interpretation of the curve of classified discharges after the classification to have additional information about the presence or absence of major specific functioning.

For example, HR_0020 system has a k_{max} of 0.07, an α_{mean} of 0.021 day^{-1}, and an IR of 0.34, corresponding to C3 (Appendix 2.D). HR_0020 is thus described as a system with a very low to low capacity of dynamic storage, a fast draining of the capacitive function and a medium to high variability of the hydrological response. The regulation time of 15.5 days indicates that the system has a medium capacity to attenuate the precipitation signal (or medium global inertia). The bending point on the curve of classified discharges at 1.8 m^3 s^{-1} may be due to the activation of an overflow outlet, a discharge to another system, or a temporary storage of water. The one at 30 m^3 s^{-1} may translate the same processes, but it is not excluded that it can be related to uncertainties on ungauged

discharges.

2.8 Conclusion

Our objective was to propose a new classification of karst systems hydrological functioning based on the analysis of spring discharge time series that is representative of a wide diversity of karst systems. Several methods were considered to identify indicators of functioning. Multivariate analyses allowed the identification of relevant indicators that allow distinguishing the analysed karst systems. Three indicators were identified as the most relevant indicators and thus retained to characterise karst systems and propose a classification into six different types of hydrological functioning. The hydrological functioning of karst systems was distinguished according to their capacity of dynamic storage, the draining dynamic of their capacitive function and the variability of their hydrological functioning. The classification can be completed with two additional analyses to characterise the global inertia and highlight the presence or absence of major specific functioning.

The challenge of developing a relevant classification was addressed (i) by considering a core dataset of karst systems with a high diversity of hydrological functioning, (ii) by selecting the most relevant indicators of hydrological functioning and proposing a classification based on multivariate analyses, and (iii) by testing the relevance of the classification on spring discharge time series of 78 karst systems located worldwide.

As the methodology requires only spring discharge time series, which is the most common monitored data, and gives relevant results with only few years of monitoring, the classification can be used in data-scarce contexts. It can thus be seen as a modern tool for the classification of karst systems hydrological functioning, which provides researchers and stakeholders with a first insight into karst system functioning based on accessible and straightforward analyses. We emphasise that the proposed typology first aims to describe the hydrological functioning of a system where one single hydrodynamic response to precipitation impulse is expected, but remains useful even when two or more responses are observed.

One perspective of this work would be to provide a database that allows both comparing the hydrological functioning of several karst systems and then proposing a link between the developed classification and the design of lumped parameter models. Research perspectives include the study of the relation between classification and the relevant structures and parameters of models for rainfall-discharge simulation.

Author contribution

Guillaume Cinkus: Conceptualisation, Methodology, Software, Validation, Formal analysis, Writing – original draft, Visualisation. **Naomi Mazzilli**: Conceptualisation, Methodology, Investigation, Writing - review & editing, Supervision. **Hervé Jourde**: Conceptualisation, Methodology, Investigation, Writing - review & editing, Supervision.

Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the French Ministry of Higher Education and Research for the thesis scholarship of G. Cinkus as well as the European Commission for its support through the Partnership for Research and Innovation in the Mediterranean Area (PRIMA) programme under Horizon 2020 (KARMA project, grant agreement number 01DH19022A). For the data provided, the French Karst National Observatory Service (SNO KARST) and especially D. Labat and A. Probst for the data gathered by A. Mangin at the Baget and Aliou springs, the Parc Naturel Régional des Grands Causses (PNRGC), G. Lorette for data monitored by Suez, the DREAL Provence Alpes-Côtes d'Azur and the DREAL Bourgogne Franche-Comté are also acknowledged.

2.9 Appendix

2.A Calculation details for the correlational and spectral analyses

The autocorrelation function r_k is calculated with the following equation:

$$r_k = \frac{C_k}{C_0} \tag{2.9}$$

with the autocovariance function C_k:

$$C_k = \frac{1}{n}\sum_{1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x}) \tag{2.10}$$

with k the shift (0, 1, 2, ..., m), n the length of the series, m the maximum shift possible (generally $m < \frac{n}{3}$), x_i the i^{th} element of the series, x_{i+t} the $(i+t)^{th}$ element of the series and $\bar{x}$ the mean of the series. The correlogram corresponds to the plot of r_k versus k.

The spectrum s_f is calculated with the following equation:

$$s_f = 2[1 + 2\sum_{k=1}^{m} D_k r_k cos(2\pi f k)] \tag{2.11}$$

with f the frequency ($f = \frac{j}{2m}$ for daily timestep), r_k the autocorrelation function and D_k a weighting function (Tukey-Hanning window) to ensure that the estimated s_f is not biaised (Mangin, 1984):

$$D_k = \frac{1 + cos(\frac{\pi k}{m})}{2} \tag{2.12}$$

The spectrum is represented on a plot of s_f versus f.

2.B Calculation details for the analysis of classified discharges

The procedure to obtain the curve of classified discharges involves in (i) the quantiles calculation of the empirical distribution function, (ii) the calculation of the corresponding variable from the reference distribution function, (iii) the graphical representation of the relation between the quantiles of the empirical and reference distribution functions, and

(iv) the choice of the x-axis scale, either arithmetic or logarithmic (corresponding to a normal or log-normal adaptation of the reference distribution, respectively).

The repartition function corresponding to the cumulative probability density regarding the standard normal distribution is:

$$P(X \leq z) = \frac{1}{2}[1 + erf(\frac{z}{\sqrt{2}})] \tag{2.13}$$

For a half-Gaussian distribution:

$$P(X \leq z) = erf(\frac{z}{\sqrt{2}}) \tag{2.14}$$

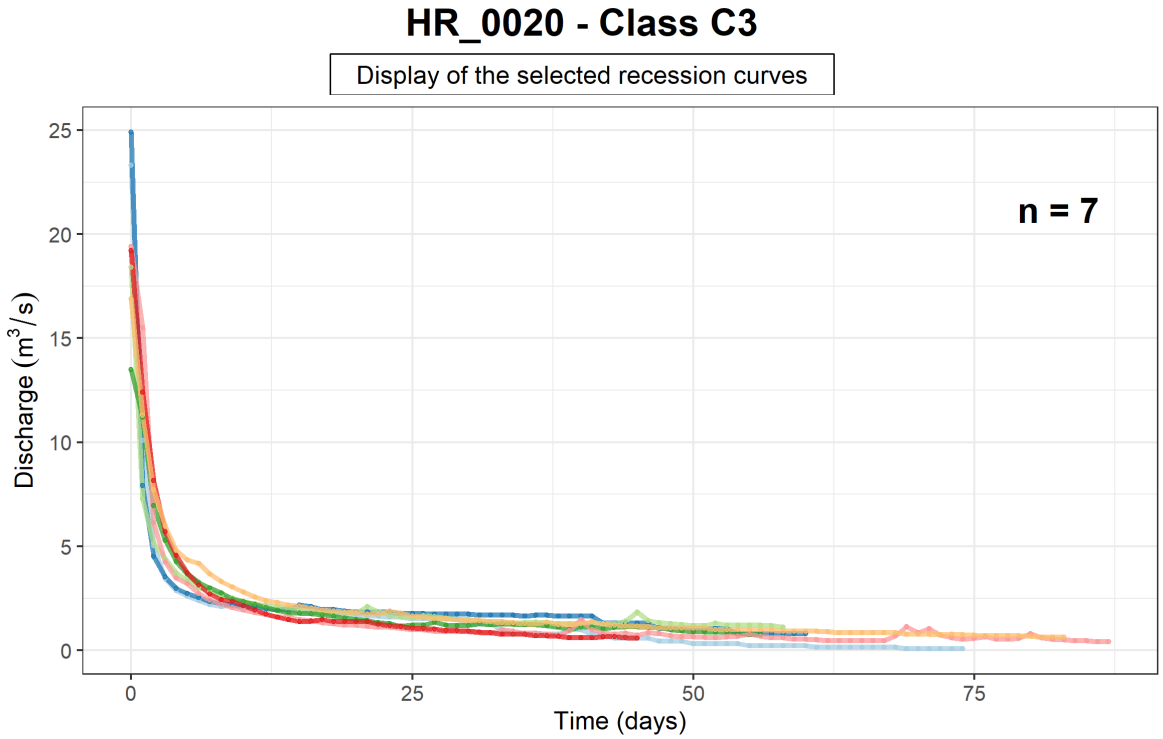
The observed discharges are plotted on the x-axis of the quantile-quantile plot.

2.C Results of the v-test applied on clusters A, B, 1, 2, 3 and 4

Indicator of functioning	v-test value					
	Cluster A	Cluster B	Cluster 1	Cluster 2	Cluster 3	Cluster 4
α_{mean}	**2.00**	**-2.00**	**2.65**	-0.17	-1.18	-1.26
k_{max}	**-2.59**	**2.59**	-1.28	-1.55	1.58	1.59
i_{mean}	**-2.33**	**2.33**	-1.76	-0.89	0.63	**2.22**
IR	-0.57	0.57	-0.67	-0.02	**2.15**	-1.45
CV	**2.47**	**-2.47**	0.95	1.69	-1.53	-1.50
SVC	**2.23**	**-2.23**	1.42	1.07	-1.44	-1.30
ME	**-2.54**	**2.54**	-1.43	-1.37	0.71	**2.41**
RT	**-2.39**	**2.39**	-1.48	-1.18	0.43	**2.50**
SBB	**2.26**	**-2.26**	1.75	0.83	-0.40	**-2.37**

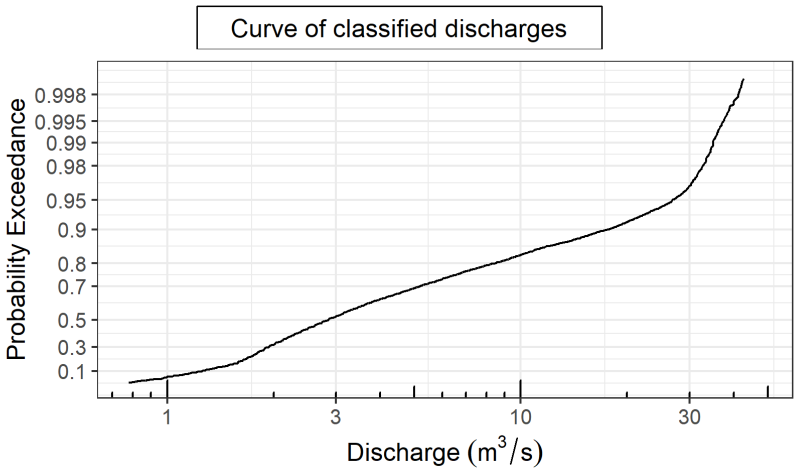
Table 2.8: Results of the v-test applied on clusters A, B, 1, 2, 3 and 4, with values of each indicators of functioning . Bold entries highlight values for which the p-value is lower than 0.05. The sign of the v-test value indicates if the mean of the cluster is lower or greater than the overall mean.

2.D Graphical summary of the typology of HR_0020 karst system



Indicators details

N°	k	i	alpha
1	0.03	0.14	0.037
2	0.07	0.1	0.017
3	0.07	0.18	0.014
4	0.05	0.45	0.016
5	0.04	0.35	0.014
6	0.02	0.34	0.035
7	0.05	0.38	0.016



Characterization of the hydrological functioning

kmax = 0.07 -> Very low to Low capacity of dynamic storage
alpha = 0.021 -> Fast draining dynamic of the capacitive function
IR = 0.34 -> High variability of the hydrological response
RT = 15.53 -> Medium global inertia

Figure 2.13: Graphical summary of the typology of HR_0020 karst system (WoKaS database, Olarinoye et al., 2020).

2.10 References for Chapter 2

Bakalowicz, M., & Ricard, J. (1994). *Etude hydrogéologique de la source de l'Espérelle et de son bassin d'alimentation, en vue de l'établissement des périmètres de protection* (BRGM No. RR-37951-FR). (Cit. on p. 23).

Barnes, B. S. (1939). The structure of discharge-recession curves. *Trans. AGU*, *20*(4), 5. https://doi.org/10.1029/TR020i004p00721 (cit. on pp. 25, 26).

Berry, M. J. A., & Linoff, G. (1997). *Data mining techniques: For Marketing, Sales, and Customer Relationship Management*. John Wiley & Sons. (Cit. on p. 36).

Blavoux, B., Mudry, J., & Puig, J.-M. (1992). Bilan, fonctionnement et protection du système karstique de la Fontaine de Vaucluse (sud-est de la France). *Geodinamica Acta*, *5*(3), 153–172. https://doi.org/10.1080/09853111.1992.11105225 (cit. on p. 23).

Bonacci, O. (1993). Karst springs hydrographs as indicators of karst aquifers. *Hydrol. Sci. J.*, *38*(1), 51–62. https://doi.org/10.1080/02626669309492639 (cit. on p. 21).

Boussinesq, J. (1877). *Essai sur la théorie des eaux courantes*. (Cit. on p. 25).

Boussinesq, J. (1903). Sur un mode simple d'écoulement des nappes d'eau d'infiltration à lit horizontal, avec rebord vertical tout autour lorsqu'une partie de ce rebord est enlevée depuis la surface jusqu'au fond. *C. R. Acad. Sci.*, *137*, 5–11 (cit. on p. 26).

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (Revised Edition). Holden Day. (Cit. on p. 28).

Brillinger, D. (1975). The Identification of Point Process Systems. *Ann. Probab.*, *3*(6), 909–924. https://doi.org/10.1214/aop/1176996218 (cit. on p. 28).

Cholet, C. (2017). *Fonctionnement hydrogéologique et processus de transport dans les aquifères karstiques du Massif du Jura* [Doctoral dissertation, Université de Bourgogne Franche-Comté]. (Cit. on pp. 23, 34, 38).

Cinkus, G., Mazzilli, N., & Jourde, H. (2021a). Identification of relevant indicators for the assessment of karst systems hydrological functioning: Proposal of a new classification. *J. Hydrol.*, *603*, 127006. https://doi.org/10.1016/j.jhydrol.2021.127006 (cit. on p. 19).

Cinkus, G., Mazzilli, N., & Jourde, H. (2021b). *Proposal of a typology of karst systems functioning based on relevant indicators of karst springs hydrodynamics* (Oral presentation). vEGU21, the 23rd EGU General Assembly. online. https://doi.org/10.5194/egusphere-egu21-4009 (cit. on p. 19).

Coutagne, A. (1948). Étude générale des variations de débit en fonction des facteurs qui les conditionnent. *La Houille Blanche*, (2), 134–146. https://doi.org/10.1051/lhb/1949025 (cit. on pp. 25, 26, 30).

Dahl, M., Nilsson, B., Langhoff, J., & Refsgaard, J. (2007). Review of classification systems and new multi-scale typology of groundwater–surface water interaction. *Journal of Hydrology*, *344*(1-2), 1–16. https://doi.org/10.1016/j.jhydrol.2007.06.027 (cit. on p. 21).

Dörflinger, N. (2010). *Guide méthodologique, Les outils de l'hydrogéologie karstique. Avec la collaboration de Ph. Crochet, R. Guerin, N. Jozja, B. Marsaud, P-H. Mondain, Ph. Muet, V. Plagnes*. (BRGM No. RP-58237-FR). BRGM. (Cit. on p. 28).

Drogue, C. (1972). Analyse statistique des hydrogrammes de decrues des sources karstiques statistical analysis of hydrographs of karstic springs. *J. Hydrol.*, *15*(1), 49–68. https://doi.org/10.1016/0022-1694(72)90075-3 (cit. on pp. 25, 26, 30).

Everitt, B., & Hothorn, T. (2011). *An introduction to applied multivariate analysis with R*. Springer. (Cit. on p. 35).
OCLC: ocn732344476.

Fiorillo, F. (2014). The Recession of Spring Hydrographs, Focused on Karst Aquifers. *Water Resour. Manage.*, *28*(7), 1781–1805. https://doi.org/10.1007/s11269-014-0597-z (cit. on p. 25).

Fleury, P., Maréchal, J. C., & Ladouche, B. (2013). Karst flash-flood forecasting in the city of Nîmes (southern France). *Engineering Geology*, *164*, 26–35. https://doi.org/10.1016/j.enggeo.2013.06.007 (cit. on p. 38).

Flora, S. P. (2004). *Hydrogeological Characterization and Discharge Variability of Springs in the Middle Verde River Watershed, Central Arizona* [Doctoral dissertation, Northern Arizona University]. (Cit. on p. 21).

Forkasiewicz, M. J., & Paloc, H. (1967). Régime de tarissement de la foux-de-la-vis (Gard) étude préliminaire. *La Houille Blanche*, (1), 29–36. https://doi.org/10.1051/lhb/1967002 (cit. on p. 25).

Gan, G., Ma, C., & Wu, J. (2007). *Data clustering: Theory, algorithms, and applications*. SIAM, Society for Industrial and Applied Mathematics ; American Statistical Association. (Cit. on p. 36).
OCLC: ocm77831225.

Govender, P., & Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric Pollution Research*, *11*(1), 40–56. https://doi.org/10.1016/j.apr.2019.09.009 (cit. on p. 36).

Grasso, A., & Jeannin, P.-Y. (1994). Etude critique des méthodes d'analyse de la réponse globale des systèmes karstiques. Application au site de Bure (JU, Suisse). *Bulletin d'Hydrogéologie*, *13*, 87–113 (cit. on p. 28).

Griffiths, G. A., & Clausen, B. (1997). Streamflow recession in basins with multiple water storages. *Journal of Hydrology*, *190*(1-2), 60–74. https://doi.org/10.1016/S0022-1694(96)03060-0 (cit. on p. 25).

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, *17*(2/3), 107–145. https://doi.org/10.1023/A:1012801612483 (cit. on p. 36).

Hannan, E. J. (2008). *Multiple time series*. John Wiley & Sons. (Cit. on p. 28).

Heath, R. C. (1982). Classification of Ground-Water Systems of the United States. *Ground Water*, *20*(4), 393–401. https://doi.org/10.1111/j.1745-6584.1982.tb02758.x (cit. on p. 21).

Heudorfer, B., Haaf, E., Stahl, K., & Barthel, R. (2019). Index-Based Characterization and Quantification of Groundwater Dynamics. *Water Resour. Res.*, *55*(7), 5575–5592. https://doi.org/10.1029/2018WR024418 (cit. on p. 21).

Horton, R. E. (1933). The Role of infiltration in the hydrologic cycle. *Trans. AGU*, *14*(1), 446. https://doi.org/10.1029/TR014i001p00446 (cit. on pp. 25, 26).

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Comput. Surv.*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504 (cit. on p. 36).

Jeannin, P.-Y., & Sauter, M. (1998). Analysis of karst hydrodynamic behaviour using global approaches: A review. *Bull. Hydrogeol.*, *16*, 31–48 (cit. on p. 28).

Jenkins, G. M., & Watts, D. G. (1968). Spectral Analysis and its Applications. *Louvain Econ. Rev.*, *36*(5), 554. https://doi.org/10.1017/S0770451800043062 (cit. on p. 28).

Jourde, H., Massei, N., Mazzilli, N., Binet, S., Batiot-Guilhe, C., Labat, D., Steinmann, M., Bailly-Comte, V., Seidel, J., Arfib, B., Charlier, J., Guinot, V., Jardani, A., Fournier, M., Aliouache, M., Babic, M., Bertrand, C., Brunet, P., Boyer, J., ... Wang, X. (2018). SNO KARST: A French Network of Observatories for the Multidisciplinary Study of Critical Zone Processes in Karst Watersheds and Aquifers. *Vadose Zone J.*, *17*(1), 180094. https://doi.org/10.2136/vzj2018.04.0094 (cit. on p. 21).

Jouves, J., Viseur, S., Arfib, B., Baudement, C., Camus, H., Collon, P., & Guglielmi, Y. (2017). Speleogenesis, geometry, and topology of caves: A quantitative study of 3D karst conduits. *Geomorphology*, *298*, 86–106. https://doi.org/10.1016/j.geomorph.2017.09.019 (cit. on p. 21).

Kovács, A. (2003). *Geometry and hydraulic parameters of karst aquifers: A hydrodynamic modeling approach.* [Doctoral dissertation, Neuchâtel University]. (Cit. on p. 28).

Krešić, N. (2007). *Hydrogeology and Groundwater Modeling* (2e edition). CRC Press. (Cit. on p. 25).

Kullman, E. (1983). Rezim podzemnych vod s turbulentnym prudenım v puklinovo-prasovom horninovom prostredı [Groundwaterregime with turbulent flow in fissure-karst rock environment (in Slovak)]. *Geologicky ustav Dionyza Stura*, *79*, 237–262 (cit. on p. 25).

Kullman, E. (2000). Nové metodické prístupy k riešeniu ochrany a ochranných pásiem zdrojov podzemných vôd v horninových prostrediach s krasovopuklinovou a puklinovou priepustnosťou. *Podzemná voda ISSN*, *1335–1052*, 31–41 (cit. on pp. 21, 25–27).

Ladouche, B., Maréchal, J.-C., Dörfliger, N., & Lachassagne, P. (2006). Étude réalisée dans le cadre du projet de recherche EAUR15 COMPLEX'AQUI du BRGM, 277 (cit. on p. 25).

Larocque, M., Mangin, A., Razack, M., & Banton, O. (1998). Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *J. Hydrol.*, *205*(3-4), 217–231. https://doi.org/10.1016/S0022-1694(97)00155-8 (cit. on p. 28).

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis* (Wiley). J. Wiley. (Cit. on p. 37).

Lorette, G., Lastennet, R., Peyraube, N., & Denis, A. (2018). Groundwater-flow characterization in a multilayered karst aquifer on the edge of a sedimentary basin in western France. *J. Hydrol.*, *566*, 137–149. https://doi.org/10.1016/j.jhydrol.2018.09.017 (cit. on pp. 23, 34, 39).

Maillet, E. T. (1905). *Essais d'hydraulique souterraine et fluviale.* A. Hermann. (Cit. on pp. 25, 26).

Malík, P. (2006). Assessment of regional karstification degree and groundwater sensitivity to pollution using hydrograph analysis in the Velka Fatra Mountains, Slovakia. *Environ. Geol.*, *51*(5), 707–711. https://doi.org/10.1007/s00254-006-0384-0 (cit. on p. 25).

Malík, P. (2015). Evaluating Discharge Regimes of Karst Aquifer. In Z. Stevanović (Ed.), *Karst Aquifers—Characterization and Engineering* (pp. 205–249). Springer Inter-

national Publishing. https://doi.org/10.1007/978-3-319-12850-4_7 (cit. on p. 24).

Malík, P., & Vojtková, S. (2012). Use of recession-curve analysis for estimation of karstification degree and its application in assessing overflow/underflow conditions in closely spaced karstic springs. *Environ. Earth Sci.*, *65*(8), 2245–2257. https://doi.org/10.1007/s12665-012-1596-0 (cit. on pp. 21, 25, 27).

Mangin, A. (1971). Etude des débits classés d'exutoires karstiques portant sur un cycle hydrologique. *Ann. spéléol.*, *26*(2), 283–329 (cit. on p. 28).

Mangin, A. (1975). *Contribution à l'étude hydrodynamique des aquifères karstiques* [Doctoral dissertation, Université de Dijon]. (Cit. on pp. 21, 23, 25–27, 30, 32, 34, 38).

Mangin, A. (1984). Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *J. Hydrol.*, *67*(1–4), 25–43. https://doi.org/10.1016/0022-1694(84)90230-0 (cit. on pp. 21, 28, 32, 37, 50).

Maréchal, J. C., Ladouche, B., & Dörfliger, N. (2008). Karst flash flooding in a Mediterranean karst, the example of Fontaine de Nîmes. *Engineering Geology*, *99*(3), 138–146. https://doi.org/10.1016/j.enggeo.2007.11.013 (cit. on p. 38).

Maréchal, J.-C., & Ladouche, B. (2006). *Fonctionnement hydrogéologique du système karstique de la Fontaine de Nîmes en crue* (RP-54723-FR). BRGM. (Cit. on pp. 23, 34, 38).

Marsaud, B. (1997). *Structure et fonctionnement de la zone noyée des karsts à partir des résultats expérimentaux.* [Doctoral dissertation, Université Paris XI Orsay]. (Cit. on p. 28).

Massei, N., Dupont, J., Mahler, B., Laignel, B., Fournier, M., Valdes, D., & Ogier, S. (2006). Investigating transport properties and turbidity dynamics of a karst aquifer using correlation, spectral, and wavelet analyses. *J. Hydrol.*, *329*(1-2), 244–257. https://doi.org/10.1016/j.jhydrol.2006.02.021 (cit. on p. 27).

Moussu, F. (2011). *Prise en compte du fonctionnement hydrodynamique dans la modélisation pluie débit des systèmes karstiques* [Doctoral dissertation, Université Pierre et Marie Curie]. (Cit. on pp. 23, 38).

Mudarra, M., & Andreo, B. (2011). Relative importance of the saturated and the unsaturated zones in the hydrogeological functioning of karst aquifers: The case of Alta Cadena (Southern Spain). *Journal of Hydrology*, *397*(3-4), 263–280. https://doi.org/10.1016/j.jhydrol.2010.12.005 (cit. on p. 27).

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z (cit. on p. 36).

Mylroie, J. (2020). Hydrologic classification of caves and karst. *Groundwater as a Geomorphic Agent*, 157–172. https://doi.org/10.4324/9781003028833-7 (cit. on p. 21).

Netopil, R. (1971). Ke Klasifikaci pramenu podle variability vydatnasti (The classification of water springs based on the basis of the variability of yields). *Sbornik-Hydrological Conference, Papers*, *22*, 145–150 (cit. on p. 24).

Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., Allocca, V., Andreo, B., Apaéstegui, J., Apolit, C., Arfib, B., Auler, A., Bailly-Comte, V., Barberá, J. A., Batiot-Guilhe, C., Bechtel, T., Binet, S., Bittner, D., Blatnik, M., Bolger, T., ... Hartmann, A. (2020). Global karst springs hydrograph dataset for research

and management of the world's fastest-flowing groundwater. *Sci. Data*, *7*(1), 59. https://doi.org/10.1038/s41597-019-0346-5 (cit. on pp. 21–23, 52).

Ollivier, C., Chalikakis, K., Mazzilli, N., Kazakis, N., Lecomte, Y., Danquigny, C., & Emblanch, C. (2019). Challenges and Limitations of Karst Aquifer Vulnerability Mapping Based on the PaPRIKa Method—Application to a Large European Karst Aquifer (Fontaine de Vaucluse, France). *Environments*, *6*(3), 39. https://doi.org/10.3390/environments6030039 (cit. on p. 39).

Padilla, A., Pulido-Bosch, A., & Mangin, A. (1994). Relative Importance of Baseflow and Quickflow from Hydrographs of Karst Spring. *Ground Water*, *32*(2), 267–277. https://doi.org/10.1111/j.1745-6584.1994.tb00641.x (cit. on pp. 25, 26, 30).

Padilla, A., & Pulido-Bosch, A. (1995). Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis. *Journal of Hydrology*, *168*(1-4), 73–89. https://doi.org/10.1016/0022-1694(94)02648-U (cit. on p. 28).

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Ko¨ppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, *11*(5), 1633–1644. https://doi.org/10.5194/hess-11-1633-2007 (cit. on p. 23).

Pinault, J.-L., Plagnes, V., Aquilina, L., & Bakalowicz, M. (2001). Inverse modeling of the hydrological and the hydrochemical behavior of hydrosystems: Characterization of Karst System Functioning. *Water Resources Research*, *37*, 2191. https://doi.org/10.1029/2001WR900018 (cit. on pp. 38, 39).

Rashed, K. A. (2012). Assessing degree of karstification: A new method of classifying karst aquifers. *Sixteenth International Water Technology Conference (IWTC)* (cit. on p. 21).

Samani, N., & Ebrahimi, B. (1996). Analysis of spring hydrographs for hydrogeological evaluation of karst aquifer system. *Journal of Theoretical and Applied Karstology*, *8*, 97–112 (cit. on p. 25).

Sinreich, M., Pronk, M., & Kozel, R. (2013). Microbiological monitoring and classification of karst springs. *Environ Earth Sci*, *71*(2), 563–572. https://doi.org/10.1007/s12665-013-2508-7 (cit. on p. 21).

Sivelle, V., Labat, D., Mazzilli, N., Massei, N., & Jourde, H. (2019). Dynamics of the Flow Exchanges between Matrix and Conduits in Karstified Watersheds at Multiple Temporal Scales. *Water*, *11*(3), 569. https://doi.org/10.3390/w11030569 (cit. on p. 38).

Soulios, G. (1991). Contribution à l'étude des courbes de récession des sources karstiques: Exemples du pays Hellénique. *J. Hydrol.*, *124*(1-2), 29–42. https://doi.org/10.1016/0022-1694(91)90004-2 (cit. on p. 21).

Springer, A. E., Stevens, L. E., Anderson, D. E., Parnell, R. A., Kreamer, D. K., Levin, L., & Flora, S. P. (2008). A comprehensive springs classification system: Integrating geomorphic, hydrogeochemical, and ecological critera. Aridland Springs in North America. *Ecol. Conserv.*, 49–75 (cit. on p. 21).

Stevanović, Z. (Ed.). (2015). *Karst Aquifers—Characterization and Engineering*. Springer International Publishing. https://doi.org/10.1007/978-3-319-12850-4 (cit. on p. 24).

Stevanović, Z. (2018). Global distribution and use of water from karst aquifers. *Geological Society, London, Special Publications*, *466*(1), 217–236. https://doi.org/10.1144/SP466.17 (cit. on p. 21).

Toebes, C., & Strang, D. D. (1964). On recession curves - Recession Equations. *J. Hydrol.*, *3*(2), 2–15 (cit. on p. 25).

Tufféry, S., & Riesco, R. (2011). *Data Mining and Statistics for Decision Making*. Wiley. https://doi.org/10.1002/9780470979174 (cit. on pp. 36, 37).

Veress, M. (2020). Karst Types and Their Karstification. *J. Earth Sci.*, *31*(3), 621–634. https://doi.org/10.1007/s12583-020-1306-x (cit. on p. 21).

Waltham, A., & Fookes, P. (2003). Engineering classification of karst ground conditions. *Quarterly Journal of Engineering Geology and Hydrogeology*, *36*(2), 101–118. https://doi.org/10.1144/1470-9236/2002-33 (cit. on p. 21).

Chapter 3

KarstID, a software for the analysis of discharge time series

Karst spring discharge time series analyses are often used to gain preliminary insights into the hydrological functioning of a karst system. KarstID is an R Shiny application that facilitates the completion of such analyses and allows the identification of karst system hydrological functioning. The application permits (i) to perform statistical, recession curves, classified discharges and signal (simple correlational and spectral) analyses; (ii) to calculate relevant indicators representative of distinct hydrological characteristics of karst systems, (iii) to classify karst systems hydrological functioning; and (iv) to compare the results to a database of 78 karst systems. The KarstID software is free, open source, and actively developed on a developer community platform. The user-friendly installation and launch make it especially accessible even for non-programmers, therefore KarstID can be used for both research and educational purposes. The application and its user manual are both available on the French SNO KARST website.

This work has resulted in a publication in Environmental Earth Sciences (Cinkus et al., 2023), several presentations during a progress meeting and a national workshop on karst, as well as a poster during Eurokarst 2022 (Cinkus et al., 2022).

Article:

Cinkus, G., Mazzilli, N., and Jourde, H.: KarstID: an R Shiny application for the analysis of karst spring discharge time series and the classification of karst system hydrological functioning, Environ Earth Sci, 82, 6, https://doi.org/10.1007/s12665-023-10830-5, 2023.

Contents

3.1 Introduction

Around 9 % of the world's population is dependent on karst water resources for drinking water (Stevanović, 2019). Karst systems are heterogeneous media with high contrasts in porosity and permeability, inducing a high variability in infiltration and internal flow processes (Bakalowicz, 2005; Ford and Williams, 2007). With the increasing demand for water, the characterisation of the functioning of karst aquifers become a major challenge for water resource management (Bakalowicz, 2011). Among the numerous methods to study karst aquifers, analyses of spring discharge time series (recession curves, signal, statistics) are the most accessible as they only require the monitoring of spring discharge. Therefore, they are generally used as a preliminary step for characterising the hydrological functioning of karst systems, and subsequently for developing and designing hydrological models. Also, many authors declined these analyses into classifications for differentiating karst systems, with recession curves (Bonacci, 1993; Cinkus et al., 2021; Dewandel et al., 2003; Fiorillo, 2014; Malík and Vojtková, 2012; Mangin, 1975; Soulios, 1991), signal (Gárfias-Soliz et al., 2010; Mangin, 1984), hydrograph (Kovács, 2021; Zhang et al., 2020) and statistical indicators (Flora, 2004; Hakoun et al., 2022; Rashed, 2012; Springer et al., 2008).

Recession curves analysis has been widely developed over the past century (Barnes, 1939; Boussinesq, 1903; Coutagne, 1948; Drogue, 1972; Horton, 1933; Maillet, 1905; Mangin, 1975; Padilla et al., 1994). It consists in calibrating numerical models on a selection of recession curves from a hydrograph and interpreting the parameters of the equations. Signal analyses –developed by Box and Jenkins (1976), Brillinger (1975), and Jenkins and Watts (1968) –were introduced in karst hydrology by Mangin (1984). Their purpose is to characterise the temporal structure of hydrological signals, which allows deducing information on the inertia of a karst system (Jeannin and Sauter, 1998; Kovács, 2003; Larocque et al., 1998). Statistical analyses include distribution indicators such as mean, standard deviation and quantiles, but also cumulative frequency curves (Malík, 2015; Mangin, 1971). Numerous studies across the world are based on these analyses for characterising karst systems properties, determining the hydrodynamic parameters of aquifers and providing information on flow dynamics (Guo et al., 2021; Lorette et al., 2018; Malík et al., 2021; Nurkholis et al., 2019; Sağır et al., 2020; Vrsalović et al., 2022; Zerouali et al., 2021).

The completion of these analyses often requires a meticulous reading of the literature and appropriate programming skills. The application of statistical and signal analyses (e.g. simple correlational and spectral analyses) is generally done using or writing specific code functions. The recession curves analysis requires to (i) select and isolate several parts of the discharge time series, (ii) calibrate a recession model over each recession curve, and (iii) calculate indicators of functioning from the model's parameters values. These operations can be tedious for long time series –and are subjects to errors in selection, calibration and indicators calculation. For these reasons, several authors proposed powerful toolboxes and software to facilitate the completion of the recession curves analysis (Arciniega-Esparza et al., 2017; Carlotto and Chaffe, 2019; Gregor and Malík, 2016; Posavec et al., 2017), one of them also including statistical and signal analyses (BRGM, 2022).

This paper presents an application (KarstID) that provides the user a toolbox for both the analysis of karst spring discharge time series and the characterisation of karst systems hydrological functioning. KarstID is distinguishable from other software because (i) it

supports multiple analyses of discharge time series (statistical, recession curves, simple correlational and spectral, classified discharges) and automatic calibration of recession model; (ii) it proposes a classification of karst systems hydrological functioning (according to the proposal of Cinkus et al., 2021) and a comparison of the results to a database of 78 karst systems; and (iii) it is free, open source and actively developed on a developer community platform. KarstID is built with the R Shiny framework (Chang et al., 2021) and is embedded into an R package (R Core Team, 2021), which make the installation and launch easy even for non-programmers.

3.2 Software overview

The links to the user guide, the source code and the git repository are available on the French SNO KARST (Service National d'Observation du Karst) website. The user guide provides guidelines for the installation and launch, as well as a technical, in-depth and visual description of all the features of the application. The source code includes the data and functions used (i) for applying the analyses, (ii) for generating the plots, (iii) for managing the application and (iv) for building the R package. Users can start discussions or raise issues in the git repository, as well as propose new code or modify existing code with Pull Requests.

3.2.1 Workflow

First, the user has to load an appropriate dataset using the "Data import" tab. The second step is to apply four different methods for analysing the hydrological functioning of the system (Figure 3.1). Two of these analyses (statistical and classified discharges analyses) do not need any actions from the user and the results of these analyses are directly displayed in their respective tabs. Two complementary analyses (recession curves and signal analyses) require the user to select curves and/or define parameters for the functions (i.e. recession model, autocorrelation function). The completion of the recession curves analysis automatically launches the third step which is the classification of the hydrological functioning of the system according to the methodology proposed by Cinkus et al. (2021). The user can then appreciate the results of the classification and compare the various hydrological characteristics of the analysed karst system with the ones of 78 karst systems located worldwide.

3.2.2 Data import

The "Data import" tab allows the user to load a karst spring discharge time series into KarstID. The raw data can be either a plain text or an Excel file, and must have only two columns referring to date and discharge, respectively. KarstID supports date and date-time format for the date column, and numeric format for the discharge column. The application proposes several features to minimise the preprocessing of the data: it is thus possible to (i) skip rows, (ii) select a specific Excel sheet, (iii) use a header or not, (iv) define the decimal mark, (v) define the delimiter and (vi) specify the date format. The user can give a name to the dataset, which will be used when displaying or downloading results.

The user can choose to interpolate missing discharge values and specify the maximum gap that will be covered. The interpolation is performed with the `na.spline(method`

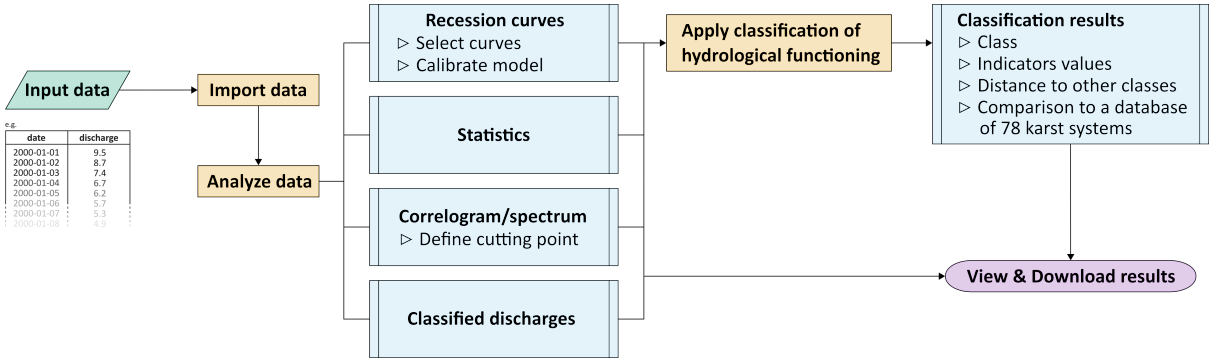


Figure 3.1: Synthetic workflow of the KarstID application. Green, yellow, blue, and purple boxes represent, respectively, (i) input data, (ii) action within KarstID, (iii) available analyses within KarstID, and (iv) output data.

= "`monoH.FC`") R function from the *zoo* package (Zeileis and Grothendieck, 2005). The method is particularly well suited for the interpolation of small gaps, but users must be careful when using it for large gaps. Critical gap length cannot be specified a priori since it depends on both the time step on the time series and the hydrological behaviour of the investigated system. The user can also choose to either (i) keep all missing values in the time series or (ii) keep only the longest part of the time series without missing values.

After defining the import options and starting the importation, the application will (i) look for missing date entries and fill the blanks if necessary (adapted to the time step of the time series), (ii) interpolate missing discharge values, (iii) perform a daily or hourly mean over the discharge time series, and (iv) display a hydrograph on the same page. The interpolation and daily/hourly mean are realised according to the user-defined options. Note that the hourly mean can only be applied if the initial time step of the time series is at an hourly time step or less.

3.2.3 Methods

Four different methods are proposed in KarstID for analysing karst spring discharge time series. The methods can be applied independently of each other in their respective tabs.

3.2.3.a Statistical analyses

Statistical analyses of spring discharge provide fundamental information about the hydrological functioning of a system. In KarstID, the following indicators are automatically calculated over the discharge time series when a dataset is imported: mean, maximum, minimum, standard deviation, 10^{th} percentile ($Q10$), 90^{th} percentile ($Q90$), Coefficient of Variation (CV) and Spring Variability Coefficient (SVC). The coefficient of variation corresponds to the ratio between the standard deviation σ and the mean μ of the values:

$$CV = \frac{\sigma}{\mu} \tag{3.1}$$

The SVC, which corresponds to the proposal of a "characteristic discharge" by Netopil (1971), is the ratio between $Q90$ (value that is exceeded 10 % of the time) and $Q10$:

$$SVC = \frac{Q90}{Q10} \tag{3.2}$$

The statistical indicators appear in a table below the hydrograph in the "Data import" tab. The number of missing discharge values are given in the last column of the table. The statistical analyses can be performed even if there are missing discharge values in the discharge time series.

3.2.3.b Recession curve analysis

Recession curves correspond to the periods when the discharge gradually decreases without replenishment of water (Toebes and Strang, 1964). The analysis of recession curves can be used to assess groundwater storage and gain insights into the hydrological functioning of an aquifer (Drogue, 1972; Forkasiewicz and Paloc, 1967; Kovács, 2003; Krešić, 2007; Kullman, 2000; Malík, 2006; Malík and Vojtková, 2012; Mangin, 1975). Generally, a recession curve can be divided into (i) an influenced regime or quickflow component, and (ii) a non-influenced regime or baseflow component. Usually, the influenced regime results from the fast infiltration of the precipitation through large fractures and conduits, while the non-influenced regime results from slow infiltration through a less transmissive media such as a porous matrix (Mangin, 1975). Numerous recession models exist whose indicators of functioning and interpretation differ.

To date, after analyses of the various aforementioned methods, Mangin's recession model (Mangin, 1975) was identified as the most informative model (Cinkus et al., 2021). Accordingly, KarstID only propose Mangin's recession model to identify relevant indicators necessary for classifying karst systems hydrological functioning (see section 3.2.4). Mangin's model is a two-equations recession model that requires the manual definition of an inflexion point for distinguishing between influenced and non-influenced regimes:

$$Q_t = Q_{R0}e^{-\alpha t} + q_0\frac{1-\eta t}{1+\varepsilon t} \tag{3.3}$$

with Q_t the discharge at time t, α the recession coefficient, Q_{R0} the baseflow extrapolated at $t = 0$, q_0 the influenced discharge corresponding to the difference between Q_0 (discharge at $t = 0$) and Q_{R0}, η a constant characterising the speed of infiltration ($\eta = \frac{1}{t_i}$, with t_i the beginning of the non-influenced regime) and ε a constant characterising the concavity of the influenced part of the recession curve.

The Mangin's recession model is widely used as several indicators can be calculated for characterising the hydrological functioning of a karst system. The indicator k gives information about the capacity of a system to store and release recharge water, and is calculated as follows:

$$k = \frac{V_{DYN}}{V_{an}} \tag{3.4}$$

with V_{an} the yearly mean volume of water discharged at the spring. The dynamic volume V_{DYN} corresponds to the integral of the exponential function of the recession model:

$$V_{DYN} = \int_0^\infty Q_i e^{-\alpha t}dt = \frac{Q_i}{\alpha} \tag{3.5}$$

The indicator i can be used to characterise the capacity of a system to dampen the precipitation signal, and corresponds to the discharge generated by the influenced regime two days after the flood peak:

$$i = \frac{1 - 2\eta}{1 + 2\varepsilon} \tag{3.6}$$

The "Recession curves analysis" tab allows to perform the recession curves analysis. The selection of recession curves is done with the cursor using the graphical interface. The retained recession curves appear in a recap table where they can be selected to apply Mangin's model. The user has to define the inflexion point of the curves, based on his knowledge and experience. The recession model is calibrated with the nonlinear least squares `nlsLM()` function from the *minpack.lm* package (Elzhov et al., 2016), which minimises the squared sum of the residuals between observed and simulated discharges. The Root Mean Square Error (RMSE) between observed and simulated discharges is displayed below the recession model plot and helps to appreciate the performance of the model. Once a recession model is calibrated and validated, the indicators of functioning are calculated. They appear in the recap table when saved by the user.

The user can choose to remove spikes on the recession curves, which usually correspond to the system's response to small precipitation events and can be considered as noise for the modelling. Recession curves analysis can be performed even if there are missing discharge values in the discharge time series.

3.2.3.c Simple correlational and spectral analyses

Simple correlational and spectral analyses are used to study the frequency content of a signal (Massei et al., 2006) by calculating the autocorrelation function and the associated spectrum with a Fourier transform. Mangin (1984) first applied these signal analyses to karst hydrology and proposed three indicators of karst hydrological functioning: the memory effect, the regulation time and the cut-off frequency. These three indicators mainly help to characterise the inertia of a karst system and its capacity to filter unitary impulse (Larocque et al., 1998; Marsaud, 1997; Massei et al., 2006). The autocorrelation r_k and autocovariance C_k functions are calculated as follows:

$$r_k = \frac{C_k}{C_0} \tag{3.7}$$

$$C_k = \frac{1}{n} \sum_{1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x}) \tag{3.8}$$

with n the length of the series, m the maximum shift possible (usually $m < \frac{n}{3}$), k the shift (between 0 and m), $\bar{x}$ the mean of the series and x_i and x_{i+t} the i^{th} and $(i+t)^{th}$ elements of the series, respectively. The spectrum s_f is derived from the autocorrelation function:

$$s_f = 2[1 + 2 \sum_{k=1}^{m} D_k r_k cos(2\pi f k)] \tag{3.9}$$

with f the frequency ($f = \frac{j}{2m}$ at daily time step) and D_k a weighting function to ensure that s_f is not biaised (Mangin, 1984):

$$D_k = \frac{1 + cos(\frac{\pi k}{m})}{2} \tag{3.10}$$

The correlogram is represented as the plot of r_k against k, and the spectrum of s_f against f. The memory effect corresponds to the value of k for a r_k of 0.2, which can be read on the correlogram or calculated from the data. The regulation time corresponds to the value of the integral of the spectrum between 0 and $+\infty$, i.e. the maximum value of the the spectrum divided by 2.

The "Simple correlational and spectral analyses" tab displays the results of the simple correlational and spectral analyses. The user can define the cutting point m, which correspond to the maximum shift possible for the calculation. The cut-off frequency is not displayed as it results from a visual, subjective assessment of the spectrum. Simple correlational and spectral analyses cannot be performed if there are any missing discharge values in the discharge time series. Appendix 3.A presents a comparison of the results obtained by Mangin (1984) and those calculated with KarstID, although the databases are different as the ones used in Mangin (1984) are unavailable.

3.2.3.d Analysis of classified discharges

The analysis of classified discharges provides information on flow dynamics within a system by analysing the distribution of the discharges at the spring. For most authors, classified discharges are equivalent to the empirical cumulative function of discharge (Stevanović, 2015). Mangin (1971) proposed a variant based on the assumption that the distribution of the discharges can be approximated by a half-normal distribution. From this perspective, classified discharges refer to the quantile-quantile graph of observed discharges quantiles against quantiles of the half-normal distribution. Homogeneous hydrological functioning should be outlined by a straight line in the classified discharge plot. Interpretation of Mangin's classified discharges thus consists of assessing the discontinuities of the curve and to relate them to changes in the hydrological functioning (e.g. activation of overflow springs, storage and release of water, leakage to another aquifer or miscalibration of the gauging station). The repartition function corresponding to the cumulative probability density regarding the standard normal distribution is calculated as follows:

$$P(X \leq z) = \frac{1}{2}[1 + erf(\frac{z}{\sqrt{2}})] \tag{3.11}$$

For a half-Gaussian distribution:

$$P(X \leq z) = erf(\frac{z}{\sqrt{2}}) \tag{3.12}$$

The "Analysis of classified discharges" tab displays the results of both analyses of classified discharges. No user action is needed for the calculation. Analyses of classified discharges can be performed even if there are missing discharge values in the discharge time series.

3.2.4 Classification

In KarstID, it is possible to characterise a karst system after the methodology proposed by Cinkus et al. (2021) and compare the results with 78 karst systems located worldwide, the discharge of which being extracted from different database: Banque Hydro (Banque Hydro, 2021), SNO KARST (Jourde et al., 2018), and World Karst Spring

Class	k_{max}	α_{mean}	IR	Capacity of dynamic storage	Draining of the capacitive function	Variability of the hydrological functioning
C1	≤ 0.4	≥ 0.03	≥ 0.25	Poor	Fast	Substantial
C2	≤ 0.4	≥ 0.03	< 0.25	Poor	Fast	Low
C3	≤ 0.4	< 0.03	≥ 0.25	Poor	Moderate	Substantial
C4	≤ 0.4	< 0.03	< 0.25	Poor	Moderate	Low
C5	> 0.4	–	≥ 0.25	Noticeable	Slow	Substantial
C6	> 0.4	–	< 0.25	Noticeable	Slow	Low

Table 3.1: Indicator thresholds and corresponding characterisation of hydrological functioning for each class.

hydrograph (Olarinoye et al., 2020). This dataset covers a wide diversity of karst hydrological functioning (from very reactive to inertial responses) with data from 17 countries in 12 different climatic conditions, according to the Köppen-Geiger classification (Cinkus et al., 2021). The classification allows characterising karst systems hydrological functioning according to 6 classes based on 3 indicators of functioning (Table 3.1). The indicators are derived from the results of the analysis of at least two recession curves. The draining of the capacitive function α_{mean} is calculated by averaging the α parameters of the recession models. The capacity of dynamic storage k_{max} corresponds to the maximum value of k among the analysed recession curves. The variability of the hydrological functioning IR corresponds to the difference between the maximum and minimum of the i distribution:

$$IR = i_{max} - i_{min} \tag{3.13}$$

The "Classification" tab highlights the results obtained for the analysed karst system and summarises the values of the various indicators considered for the classification (Figure 3.6). A flowchart thus indicates how the system is classified according to the values of the indicators of functioning. The associated text section (i) describes the hydrological functioning of the system according to its class, (ii) displays the indicators values and (iii) shows the distance to other classes. A 3D scatter plot shows the investigated system (highlighted in red) alongside 78 other karst systems, with each axis corresponding to one indicator of functioning. Results from statistical, recession curves, simple correlational and spectral analyses, as well as indicators of functioning of all 78 systems also appear in a recap table. By default, the systems in the table are ordered by increasing distance to the investigated system. The user can select a system in the table to highlight (in yellow) its position on the 3D scatter plot.

3.3 Test case

Fontaine de Vaucluse is a karst spring located South-East of France. Its recharge area is estimated to be about 1160 km^2 (Ollivier et al., 2019), resulting in one of the highest karst spring interannual mean discharge in Europe (17.5 m^3 s^{-1} over the 1966–2018 period).

Fontaine de Vaucluse's daily discharges over the 2013–2019 period (amounting to 1923 observations) are provided in KarstID as a test dataset. After importation using the "load test dataset" button, the hydrograph is loaded on the import page (Figure 3.2). The statistical indicators and number of missing discharge values are displayed

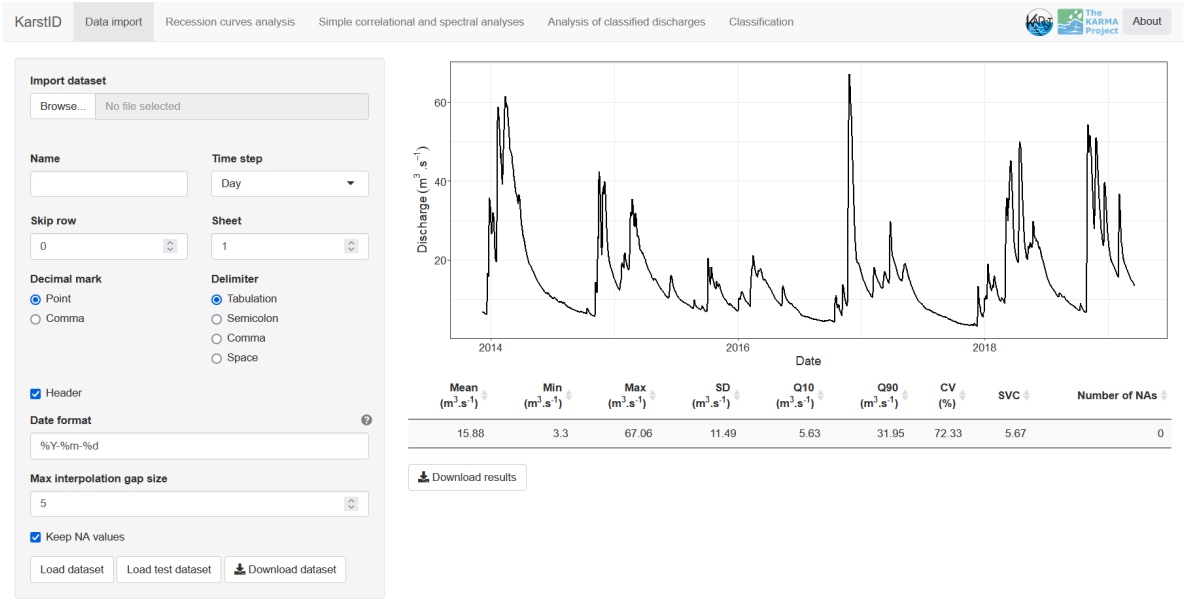


Figure 3.2: Import and statistical analyses tab. Left pane is dedicated to data import (section 3.2.2). Right part presents the hydrograph and the results of statistical analyses (section 3.2.3.a).

in the table below the plot. For this period, Fontaine de Vaucluse's interannual mean discharge is about 15.9 m^3 s^{-1} with no missing discharge values. The maximum observed discharge (about 67.1 m^3 s^{-1}) and the 90^{th} percentile of observed discharges (about 32.0 m^3 s^{-1}) show that the discharge at the spring can be and stay very high during wet periods. The minimum observed discharge and the 10^{th} percentile of observed discharges are relatively close (about 3.3 and 5.6 m^3 s^{-1}, respectively), highlighting a slow and consistent release of water from storage during dry periods. The coefficient of variation (72.3 %) and SVC (5.7) are average and correspond to a "moderate" and "balanced" discharge variability, respectively (Flora, 2004; Springer et al., 2008). The moderate discharge variability and the fact that the discharge can attain very high values can be related to a strong karstification of a part of the system. Using cross-correlation analyses between precipitation and discharge, (Ollivier et al., 2015) found a transfer time between 1 and 6 days, indicating a somewhat reactive response of the system to precipitation events.

The autocorrelation function of discharge (Figure 3.3) declines slowly and steadily, reaching 0 at 117 days. The memory effect and the regulation time are of about 56.0 and 44.0 days, respectively. These values testify of a significant capacity of filtration of the precipitation signal, which relates to the overall organisation of flows in the system (Jeannin and Sauter, 1998). The noticeable dampening of the recharge in the Fontaine de Vaucluse karst system can be related to the very large dimensions of its recharge area and unsaturated zone (Ollivier et al., 2019) or the characteristics of the Urgonian limestones (Carrière et al., 2016).

The analysis of classified discharges (Figure 3.4) according to the methodology proposed by Mangin (1975) hints that there are flow properties changes beyond 20 m^3 s^{-1} (less steep slope following the inflexion point). This discontinuity reflects the overflow threshold of the upper spring pool (Mangin, 1975). The other inflexion point, occurring at 57.5 m^3 s^{-1}, can be related to several hydrological processes: activation of an overflow, temporary storage of water or leakage to another aquifer. It can be also due to a miscal-

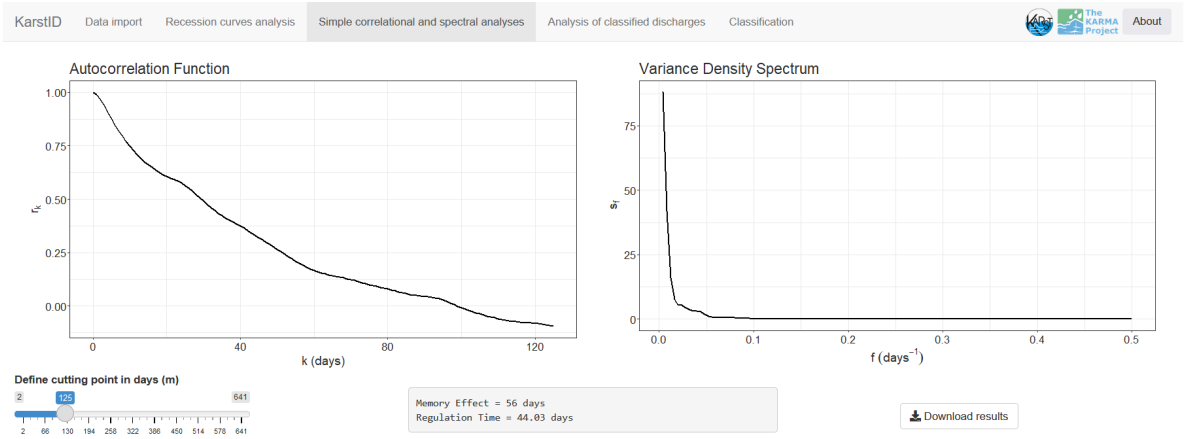


Figure 3.3: Simple correlational and spectral analyses tab. Left and right graphs present the autocorrelation function and the variance density spectrum, respectively (section 3.2.3.c).

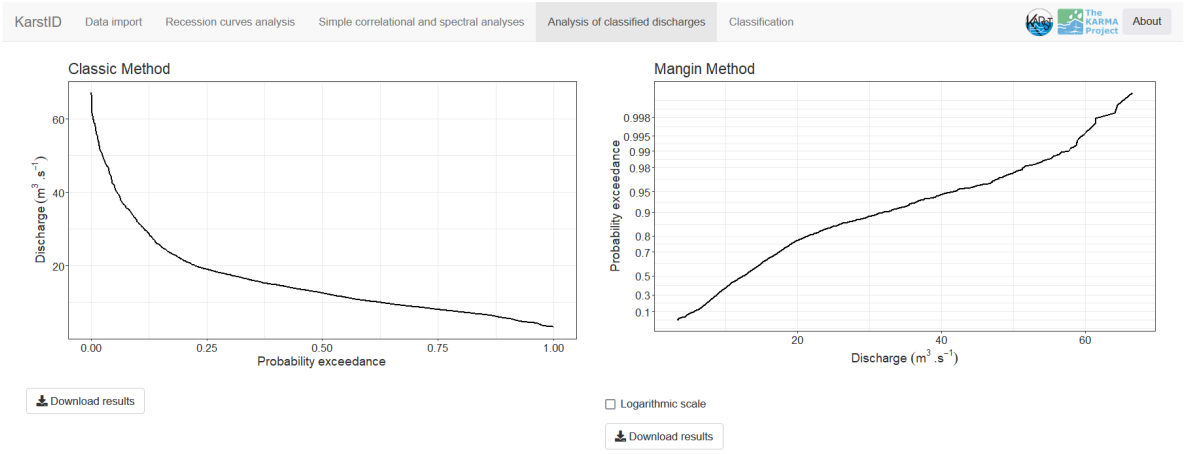


Figure 3.4: Analysis of classified discharges tab. Left and right graphs present the empirical cumulative function of discharge and the Mangin classified discharges, respectively (section 3.2.3.d).

ibration of the gauging station or uncertainties on the water level-discharge calibration curve.

Three recession curves were selected and applied Mangin's recession model to identify relevant indicators (Figure 3.5). The recession curves were chosen according to the following criteria to ensure a maximum relevance of the analysis and its results: (i) the peak discharge must be at least one tenth of the maximum discharge of the time series, (ii) there must be little or no untimely peaks during the recession, and (iii) the recession must include both influenced and non-influenced regimes. The inflexion points (n.b. "breakpoint" in the application) were defined manually based on expert knowledge and RMSE values. The indicators k, i, and α are then calculated for each recession curve and appear in the recap table.

Fontaine de Vaucluse is classified C6 with a k_{max} of 0.403, an α_{mean} of 0.006 and an IR of 0.022 (Figure 3.6). This class characterises a system with noticeable capacity of dynamic storage, slow draining of the capacitive function and low variability of hydrological functioning. Fontaine de Vaucluse is considered close to the C4 class with a distance of about 0.8 % (normalised Euclidean distance in the three-dimensional criteria space), meaning that C4 characteristics can also be considered in the interpretation. It

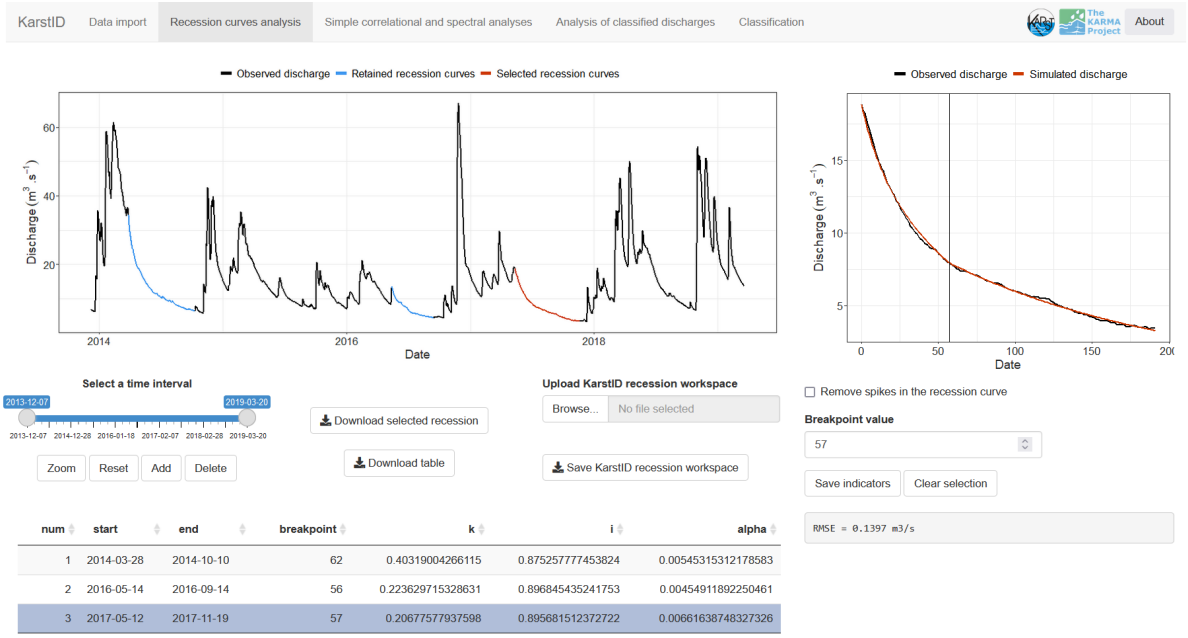


Figure 3.5: Recession curves and modelling tab. The left graph presents the studied time series and retained/selected recession curves. The right graph displays the selected recession with the Mangin recession model. The table shows the details of each recession curves and their corresponding indicators values (section 3.2.3.b).

highlights that the system may have a capacity of dynamic storage and a draining of the capacitive function in-between C4 and C6 characteristics. Fontaine de Vaucluse's class is also far from the classes C3 and C5 with distances of about 91.4 %, which is due to the very low IR (0.022). The low variability of hydrological functioning and noticeable capacity of dynamic storage assigned to this system can be also related to the large extent of the recharge area and the thick unsaturated zone. Local variability of hydrological functioning may thus be mitigated as a consequence of the spatial averaging (indirectly inducing a strong filtration of the precipitation signal). The dampening of the rainfall-discharge relationship and the noticeable capacity of dynamic storage may also be related to particular hydrological behaviour of Urgonian limestones (Carrière et al., 2016). By looking at the page 1 of the database table, users can find other karst systems with similar hydrological functioning: e.g. Taillade, PR_0005 and IE_0018. These systems are highlighted in yellow on the 3D scatter plot, alongside the investigated system highlighted in red. Studying the characteristics of other similar systems may help to support the interpretation of the investigated system. Note that Fontaine de Vaucluse also appears in the table but here corresponds to the permanent entry of the database, which results from the analysis of the whole discharge time series.

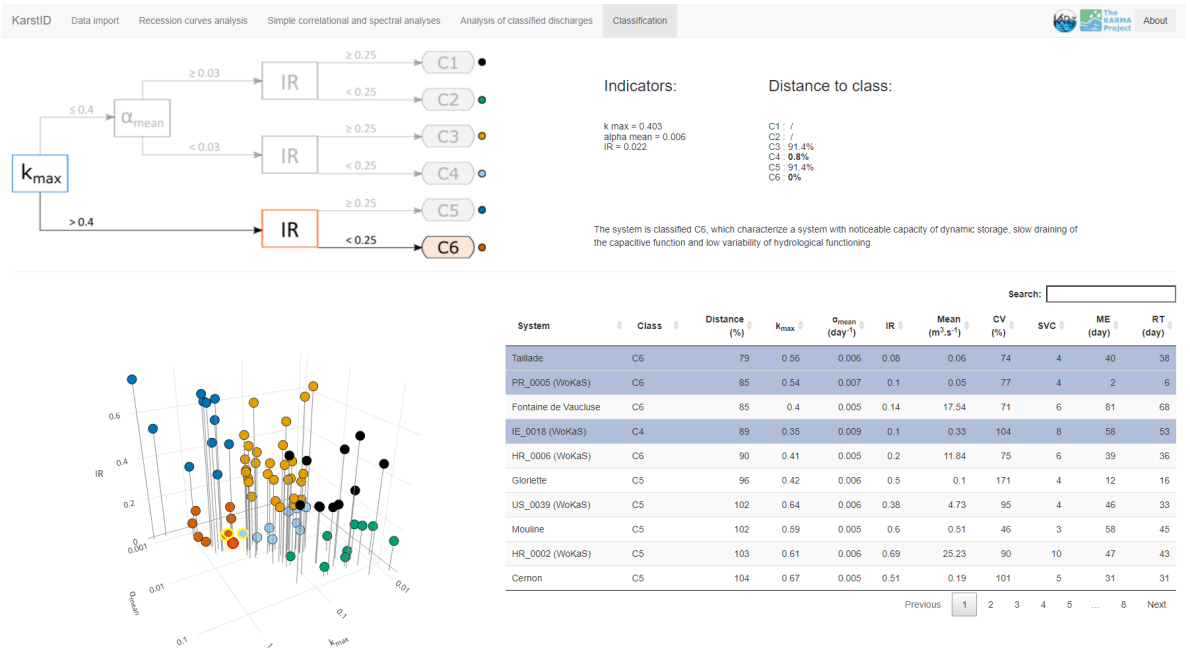


Figure 3.6: Classification tab. The top part shows the classification flowchart and its associated text description: indicator values and distance to other classes (section 3.2.4).

3.4 Conclusion

KarstID can be seen as a useful tool for gaining preliminary insights into the hydrological functioning of a karst system. The application supports different methods for analysing discharge time series and proposes a classification of karst systems hydrological functioning. It is also possible to compare the results with a database of 78 karst systems located worldwide. KarstID is free, open source, and available on a developer community platform, which allow potential interaction between users and developers for improving software efficiency or adding new features. Other than the installation of R and R packages, no programming skills are required to use the application. KarstID could therefore also be relevant for occasional users or educational purpose. Future developments of the application include (i) a continuous consideration of feature requests and bug reports to improve user experience, (ii) the proposition of additional recession models (Drogue, 1972; Kullman, 2000; Padilla et al., 1994), and (iii) the addition of other discharge time series analyses (e.g. wavelet analyses).

Acknowledgements

The French Ministry of Higher Education and Research is thanked for the thesis scholarship of G. Cinkus. This application is developed within the framework of (i) the KARST observatory network (www.sokarst.org) initiated by the INSU/CNRS, which aims to strengthen knowledge-sharing and promote cross-disciplinary research on karst systems; and (ii) the KARMA project (Karst Aquifer Resources availability and quality in the Mediterranean Area, http://karma-project.org/). The DREAL Provence Alpes-Côtes d'Azur (PACA) is also acknowledged for providing Fontaine de Vaucluse's data.

The manuscript was written with the Rmarkdown framework (Allaire et al., 2021; Xie et al., 2018; Xie et al., 2020), using R (R Core Team, 2021) and knitr (Xie, 2021). KarstID software uses functions from the following packages: data.table (Dowle and

Srinivasan, 2021), DT (Xie et al., 2021), minpack.lm (Elzhov et al., 2016), padr (Thoen, 2021), plotly (Sievert, 2020), shiny (Chang et al., 2021), shinyFeedback (Merlino and Howard, 2020), shinyhelper (Mason-Thom, 2019), shinyjs (Attali, 2020), waiter (Coene, 2021), zoo (Zeileis and Grothendieck, 2005), and readxl, readr, dplyr, stringr, magrittr, ggplot2, lubridate (Wickham et al., 2019). The KarstID package was developed and structured using devtools (Wickham et al., 2021), usethis (Wickham and Bryan, 2021) and testthat (Wickham, 2011) packages.

Author contribution

GC, NM and HJ worked on the methodology and underlying sciences; GC programmed, coded and developed the application; GC wrote the original draft; GC, NM and HJ reviewed and edited the original draft.

Data availability

KarstID is a free R Shiny application embedded into an R package (4.6 Mo), developed by Guillaume Cinkus (guillaume.cinkus@umontpellier.fr) and first made available in 2021. The software requires R version 4.0.0 or later, can be run on any web browser, and is licensed under Creative Commons Attribution 4.0 International. The code repository is available on GitHub: https://github.com/busemorose/KarstID.

Competing Interest

The authors have no relevant financial or non-financial interests to disclose.

3.5 Appendix

3.A Comparison of indicators values from the correlational and spectral analyses

System	Indicator	Mangin (1984)	Cinkus et al. (2021)
Aliou	Memory Effect (days)	4	5
	Regulation time (days)	14	11
Baget	Memory Effect (days)	15	18
	Regulation time (days)	22.5	24

Table 3.2: Comparison of indicators values from the correlational and spectral analyses between the original publication of Mangin (1984) and recent work of Cinkus et al. (2021) – calculated with KarstID. The exact time series used by Mangin (1984) are unavailable so different – and more recent – time series were used by Cinkus et al. (2021).

3.6 References for Chapter 3

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2021). *Rmarkdown: Dynamic documents for r.* (Version 2.22, R package). (Cit. on p. 71).

Arciniega-Esparza, S., Breña-Naranjo, J. A., Pedrozo-Acuña, A., & Appendini, C. M. (2017). HYDRORECESSION: A Matlab toolbox for streamflow recession analysis. *Comput. Geosci.*, *98*, 87–92. https://doi.org/10.1016/j.cageo.2016.10.005 (cit. on p. 61).

Attali, D. (2020). *Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds.* (Version 2.1.0, R package). (Cit. on p. 72).

Bakalowicz, M. (2005). Karst groundwater: A challenge for new resources. *Hydrogeol. J.*, *13*(1), 148–160. https://doi.org/10.1007/s10040-004-0402-9 (cit. on p. 61).

Bakalowicz, M. (2011). Management of Karst Groundwater Resources. In P. E. van Beynen (Ed.), *Karst Management* (pp. 263–282). Springer Netherlands. https://doi.org/10.1007/978-94-007-1207-2_12 (cit. on p. 61).

Banque Hydro. (2021). French Ministry of Ecology, Energy, Sustainable Development, Archive of hydrological data. Available online: http://hydro.eaufrance.fr/. (Cit. on p. 66).

Barnes, B. S. (1939). The structure of discharge-recession curves. *Trans. AGU*, *20*(4), 5. https://doi.org/10.1029/TR020i004p00721 (cit. on p. 61).

Bonacci, O. (1993). Karst springs hydrographs as indicators of karst aquifers. *Hydrol. Sci. J.*, *38*(1), 51–62. https://doi.org/10.1080/02626669309492639 (cit. on p. 61).

Boussinesq, J. (1903). Sur un mode simple d'écoulement des nappes d'eau d'infiltration à lit horizontal, avec rebord vertical tout autour lorsqu'une partie de ce rebord est enlevée depuis la surface jusqu'au fond. *C. R. Acad. Sci.*, *137*, 5–11 (cit. on p. 61).

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control* (Revised Edition). Holden Day. (Cit. on p. 61).

BRGM. (2022). *XLKarst : une application Excel pour la caractérisation hydrodynamique des systèmes karstiques. https://www.brgm.fr/en/software/xlkarst-excel-application-hydrodynamic-characterisation-karst-systems.* Retrieved March 30, 2022, from https://www.brgm.fr/fr/logiciel/xlkarst-application-excel-caracterisation-hydrodynamique-systemes-karstiques (cit. on p. 61).

Brillinger, D. (1975). The Identification of Point Process Systems. *Ann. Probab.*, *3*(6), 909–924. https://doi.org/10.1214/aop/1176996218 (cit. on p. 61).

Carlotto, T., & Chaffe, P. L. B. (2019). Master Recession Curve Parameterization Tool (MRCPtool): Different approaches to recession curve analysis. *Comput. Geosci.*, *132*, 1–8. https://doi.org/10.1016/j.cageo.2019.06.016 (cit. on p. 61).

Carrière, S. D., Chalikakis, K., Danquigny, C., Davi, H., Mazzilli, N., Ollivier, C., & Emblanch, C. (2016). The role of porous matrix in water flow regulation within a karst unsaturated zone: An integrated hydrogeophysical approach. *Hydrogeol. J.*, *24*(7), 1905–1918. https://doi.org/10.1007/s10040-016-1425-8 (cit. on pp. 68, 70).

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *Shiny: Web Application Framework for R.* (Version 1.6.0, R package). (Cit. on pp. 62, 72).

Cinkus, G., Mazzilli, N., & Jourde, H. (2021). Identification of relevant indicators for the assessment of karst systems hydrological functioning: Proposal of a new classification. *J. Hydrol.*, *603*, 127006. https://doi.org/10.1016/j.jhydrol.2021.127006 (cit. on pp. 61, 62, 64, 66, 67, 72).

Cinkus, G., Mazzilli, N., & Jourde, H. (2022). *KarstID: An interactive R application for karst spring hydrograph analysis that also provides a classification of karst hydrological functioning* (Poster). Eurokarst 2022 – The European Conference on Karst Hydrogeology and Carbonate Reservoirs. Málaga, Spain. (Cit. on p. 59).

Cinkus, G., Mazzilli, N., & Jourde, H. (2023). KarstID: An R Shiny application for the analysis of karst spring discharge time series and the classification of karst system hydrological functioning. *Environ Earth Sci*, *82*(136), 6. https://doi.org/10.1007/s12665-023-10830-5 (cit. on p. 59).

Coene, J. (2021). *Waiter: Loading Screen for 'Shiny'.* (Version 0.2.3, R package). (Cit. on p. 72).

Coutagne, A. (1948). Étude générale des variations de débit en fonction des facteurs qui les conditionnent. *La Houille Blanche*, (2), 134–146. https://doi.org/10.1051/lhb/1949025 (cit. on p. 61).

Dewandel, B., Lachassagne, P., Bakalowicz, M., Weng, P., & Al-Malki, A. (2003). Evaluation of aquifer thickness by analysing recession hydrographs. Application to the Oman ophiolite hard-rock aquifer. *J. Hydrol.*, *274*(1-4), 248–269. https://doi.org/10.1016/S0022-1694(02)00418-3 (cit. on p. 61).

Dowle, M., & Srinivasan, A. (2021). *Data.table: Extension of 'data.frame'.* (Version 1.14.0, R package). (Cit. on p. 71).

Drogue, C. (1972). Analyse statistique des hydrogrammes de decrues des sources karstiques statistical analysis of hydrographs of karstic springs. *J. Hydrol.*, *15*(1), 49–68. https://doi.org/10.1016/0022-1694(72)90075-3 (cit. on pp. 61, 64, 71).

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., & Bolker, B. (2016). *Minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds.* (Version 1.2.1, R package). (Cit. on pp. 65, 72).

Fiorillo, F. (2014). The Recession of Spring Hydrographs, Focused on Karst Aquifers. *Water Resour. Manage.*, *28*(7), 1781–1805. https://doi.org/10.1007/s11269-014-0597-z (cit. on p. 61).

Flora, S. P. (2004). *Hydrogeological Characterization and Discharge Variability of Springs in the Middle Verde River Watershed, Central Arizona* [Doctoral dissertation, Northern Arizona University]. (Cit. on pp. 61, 68).

Ford, D., & Williams, P. (2007). Karst Hydrogeology. In *Karst Hydrogeology and Geomorphology* (pp. 103–144). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118684986.ch5 (cit. on p. 61).

Forkasiewicz, M. J., & Paloc, H. (1967). Régime de tarissement de la foux-de-la-vis (Gard) étude préliminaire. *La Houille Blanche*, (1), 29–36. https://doi.org/10.1051/lhb/1967002 (cit. on p. 64).

Gárfias-Soliz, J., Llanos-Acebo, H., & Martel, R. (2010). Time series and stochastic analyses to study the hydrodynamic characteristics of karstic aquifers. *Hydrol. Process.*, *24*(3), 300–316. https://doi.org/10.1002/hyp.7487 (cit. on p. 61).

Gregor, M., & Malík, P. (2016). User manual for HydroOffice RC 4.0 tool. Online only, 35 pp. http://hydrooffice.org. (Cit. on p. 61).

Guo, Y., Wang, F., Qin, D.-j., Zhao, Z.-f., Gan, F.-p., Yan, B.-k., Bai, J., & Muhammed, H. (2021). Hydrodynamic characteristics of a typical karst spring system based on time series analysis in northern China. *China Geol.*, *4*(3), 433–445. https://doi.org/10.31035/cg2021049 (cit. on p. 61).

Hakoun, V., Bailly-Comte, V., Charlier, J.-B., Ladouche, B., & Maréchal, J.-C. (2022). Definition of new indicators for the characterization and classification of karst aquifers using discharge time series (cit. on p. 61).

Horton, R. E. (1933). The Role of infiltration in the hydrologic cycle. *Trans. AGU*, *14*(1), 446. https://doi.org/10.1029/TR014i001p00446 (cit. on p. 61).

Jeannin, P.-Y., & Sauter, M. (1998). Analysis of karst hydrodynamic behaviour using global approaches: A review. *Bull. Hydrogeol.*, *16*, 31–48 (cit. on pp. 61, 68).

Jenkins, G. M., & Watts, D. G. (1968). Spectral Analysis and its Applications. *Louvain Econ. Rev.*, *36*(5), 554. https://doi.org/10.1017/S0770451800043062 (cit. on p. 61).

Jourde, H., Massei, N., Mazzilli, N., Binet, S., Batiot-Guilhe, C., Labat, D., Steinmann, M., Bailly-Comte, V., Seidel, J., Arfib, B., Charlier, J., Guinot, V., Jardani, A., Fournier, M., Aliouache, M., Babic, M., Bertrand, C., Brunet, P., Boyer, J., ... Wang, X. (2018). SNO KARST: A French Network of Observatories for the Multidisciplinary Study of Critical Zone Processes in Karst Watersheds and Aquifers. *Vadose Zone J.*, *17*(1), 180094. https://doi.org/10.2136/vzj2018.04.0094 (cit. on p. 66).

Kovács, A. (2003). *Geometry and hydraulic parameters of karst aquifers: A hydrodynamic modeling approach.* [Doctoral dissertation, Neuchâtel University]. (Cit. on pp. 61, 64).

Kovács, A. (2021). Quantitative classification of carbonate aquifers based on hydrodynamic behaviour. *Hydrogeol. J.*, *29*(1), 33–52. https://doi.org/10.1007/s10040-020-02285-w (cit. on p. 61).

Krešić, N. (2007). *Hydrogeology and Groundwater Modeling* (2e edition). CRC Press. (Cit. on p. 64).

Kullman, E. (2000). Nové metodické prístupy k riešeniu ochrany a ochranných pásiem zdrojov podzemných vôd v horninových prostrediach s krasovopuklinovou a puklinovou priepustnosťou. *Podzemná voda ISSN*, *1335–1052*, 31–41 (cit. on pp. 64, 71).

Larocque, M., Mangin, A., Razack, M., & Banton, O. (1998). Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *J. Hydrol.*, *205*(3-4), 217–231. https://doi.org/10.1016/S0022-1694(97)00155-8 (cit. on pp. 61, 65).

Lorette, G., Lastennet, R., Peyraube, N., & Denis, A. (2018). Groundwater-flow characterization in a multilayered karst aquifer on the edge of a sedimentary basin in western France. *J. Hydrol.*, *566*, 137–149. https://doi.org/10.1016/j.jhydrol.2018.09.017 (cit. on p. 61).

Maillet, E. T. (1905). *Essais d'hydraulique souterraine et fluviale.* A. Hermann. (Cit. on p. 61).

Malík, P. (2006). Assessment of regional karstification degree and groundwater sensitivity to pollution using hydrograph analysis in the Velka Fatra Mountains, Slovakia. *Environ. Geol.*, *51*(5), 707–711. https://doi.org/10.1007/s00254-006-0384-0 (cit. on p. 64).

Malík, P. (2015). Evaluating Discharge Regimes of Karst Aquifer. In Z. Stevanović (Ed.), *Karst Aquifers—Characterization and Engineering* (pp. 205–249). Springer International Publishing. https://doi.org/10.1007/978-3-319-12850-4_7 (cit. on p. 61).

Malík, P., Švasta, J., Bajtoš, P., & Gregor, M. (2021). Discharge recession patterns of karstic springs as observed in Triassic carbonate aquifers of Slovakia. *Hydrogeol. J.*, *29*(1), 397–427. https://doi.org/10.1007/s10040-020-02276-x (cit. on p. 61).

Malík, P., & Vojtková, S. (2012). Use of recession-curve analysis for estimation of karstification degree and its application in assessing overflow/underflow conditions

in closely spaced karstic springs. *Environ. Earth Sci.*, *65*(8), 2245–2257. https://doi.org/10.1007/s12665-012-1596-0 (cit. on pp. 61, 64).

Mangin, A. (1971). Etude des débits classés d'exutoires karstiques portant sur un cycle hydrologique. *Ann. spéléol.*, *26*(2), 283–329 (cit. on pp. 61, 66).

Mangin, A. (1975). *Contribution à l'étude hydrodynamique des aquifères karstiques* [Doctoral dissertation, Université de Dijon]. (Cit. on pp. 61, 64, 68).

Mangin, A. (1984). Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *J. Hydrol.*, *67*(1–4), 25–43. https://doi.org/10.1016/0022-1694(84)90230-0 (cit. on pp. 61, 65, 66, 72).

Marsaud, B. (1997). *Structure et fonctionnement de la zone noyée des karsts à partir des résultats expérimentaux.* [Doctoral dissertation, Université Paris XI Orsay]. (Cit. on p. 65).

Mason-Thom, C. (2019). *Shinyhelper: Easily Add Markdown Help Files to 'shiny' App Elements.* (Version 0.3.2, R package). (Cit. on p. 72).

Massei, N., Dupont, J., Mahler, B., Laignel, B., Fournier, M., Valdes, D., & Ogier, S. (2006). Investigating transport properties and turbidity dynamics of a karst aquifer using correlation, spectral, and wavelet analyses. *J. Hydrol.*, *329*(1-2), 244–257. https://doi.org/10.1016/j.jhydrol.2006.02.021 (cit. on p. 65).

Merlino, A., & Howard, P. (2020). *shinyFeedback: Display User Feedback in Shiny Apps.* (Version 0.3.0, R package). (Cit. on p. 72).

Netopil, R. (1971). Ke Klasifikaci pramenu podle variability vydatnasti (The classification of water springs based on the basis of the variability of yields). *Sbornik-Hydrological Conference, Papers*, *22*, 145–150 (cit. on p. 63).

Nurkholis, A., Adji, T. N., Haryono, E., Cahyadi, A., & Suprayogi, S. (2019). Time series analysis application for karst aquifer characterisation in Pindul Cave karst system, Indonesia. *Acta Carsologica*, *48*(1). https://doi.org/10.3986/ac.v48i1.6745 (cit. on p. 61).

Olarinoye, T., Gleeson, T., Marx, V., Seeger, S., Adinehvand, R., Allocca, V., Andreo, B., Apaéstegui, J., Apolit, C., Arfib, B., Auler, A., Bailly-Comte, V., Barberá, J. A., Batiot-Guilhe, C., Bechtel, T., Binet, S., Bittner, D., Blatnik, M., Bolger, T., ... Hartmann, A. (2020). Global karst springs hydrograph dataset for research and management of the world's fastest-flowing groundwater. *Sci. Data*, *7*(1), 59. https://doi.org/10.1038/s41597-019-0346-5 (cit. on p. 67).

Ollivier, C., Danquigny, C., Mazzilli, N., & Barbel-Perineau, A. (2015). Contribution of Hydrogeological Time Series Statistical Analysis to the Study of Karst Unsaturated Zone (Rustrel, France). In B. Andreo, F. Carrasco, J. J. Durán, P. Jiménez, & J. W. LaMoreaux (Eds.), *Hydrogeological and Environmental Investigations in Karst Systems* (pp. 27–33). Springer. https://doi.org/10.1007/978-3-642-17435-3_4 (cit. on p. 68).

Ollivier, C., Chalikakis, K., Mazzilli, N., Kazakis, N., Lecomte, Y., Danquigny, C., & Emblanch, C. (2019). Challenges and Limitations of Karst Aquifer Vulnerability Mapping Based on the PaPRIKa Method—Application to a Large European Karst Aquifer (Fontaine de Vaucluse, France). *Environments*, *6*(3), 39. https://doi.org/10.3390/environments6030039 (cit. on pp. 67, 68).

Padilla, A., Pulido-Bosch, A., & Mangin, A. (1994). Relative Importance of Baseflow and Quickflow from Hydrographs of Karst Spring. *Ground Water*, *32*(2), 267–277. https://doi.org/10.1111/j.1745-6584.1994.tb00641.x (cit. on pp. 61, 71).

Posavec, K., Giacopetti, M., Materazzi, M., & Birk, S. (2017). Method and Excel VBA Algorithm for Modeling Master Recession Curve Using Trigonometry Approach. *Groundwater*, *55*(6), 891–898. https://doi.org/10.1111/gwat.12549 (cit. on p. 61).

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved June 6, 2023, from https://www.R-project.org/ (cit. on pp. 62, 71).

Rashed, K. A. (2012). Assessing degree of karstification: A new method of classifying karst aquifers. *Sixteenth International Water Technology Conference (IWTC)* (cit. on p. 61).

Sağır, Ç., Kurtuluş, B., & Razack, M. (2020). Hydrodynamic Characterization of Mugla Karst Aquifer Using Correlation and Spectral Analyses on the Rainfall and Springs Water-Level Time Series. *Water*, *12*(1), 85. https://doi.org/10.3390/w12010085 (cit. on p. 61).

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida, 2020. https://cran.r-project.org/package=plotly.* (Cit. on p. 72).

Soulios, G. (1991). Contribution à l'étude des courbes de récession des sources karstiques: Exemples du pays Hellénique. *J. Hydrol.*, *124*(1-2), 29–42. https://doi.org/10.1016/0022-1694(91)90004-2 (cit. on p. 61).

Springer, A. E., Stevens, L. E., Anderson, D. E., Parnell, R. A., Kreamer, D. K., Levin, L., & Flora, S. P. (2008). A comprehensive springs classification system: Integrating geomorphic, hydrogeochemical, and ecological critera. Aridland Springs in North America. *Ecol. Conserv.*, 49–75 (cit. on pp. 61, 68).

Stevanović, Z. (Ed.). (2015). *Karst Aquifers—Characterization and Engineering.* Springer International Publishing. https://doi.org/10.1007/978-3-319-12850-4 (cit. on p. 66).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on p. 61).

Thoen, E. (2021). *Padr: Quickly get datetime data ready for analysis.* (Version 0.6.2, R package). (Cit. on p. 72).

Toebes, C., & Strang, D. D. (1964). On recession curves - Recession Equations. *J. Hydrol.*, *3*(2), 2–15 (cit. on p. 64).

Vrsalović, A., Andrić, I., Buzjak, N., & Bonacci, O. (2022). Karst Lake's Dynamics Analysis as a Tool for Aquifer Characterisation at Field Scale, Example of Cryptodepression—Red Lake in Croatia. *Water*, *14*(5), 830. https://doi.org/10.3390/w14050830 (cit. on p. 61).

Wickham, H. (2011). Testthat: Get Started with Testing. *R Journal*, *3*(1), 5–10 (cit. on p. 72).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *J. Open Source Softw.*, *4*(43), 1686. https://doi.org/10.21105/joss.01686 (cit. on p. 72).

Wickham, H., & Bryan, J. (2021). *Usethis: Automate Package and Project Setup.* (Version 2.0.1, R package). (Cit. on p. 72).

Wickham, H., Hester, J., & Chang, W. (2021). *Devtools: Tools to Make Developing R Packages Easier.* (Version 2.4.2, R package). (Cit. on p. 72).

Xie, Y. (2021). *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* (Version 1.33, R package). (Cit. on p. 71).

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Chapman and Hall/CRC. (Cit. on p. 71).

Xie, Y., Cheng, J., & Tan, X. (2021). *DT: A Wrapper of the JavaScript Library 'DataTables'.* (Version 0.19, R package). (Cit. on p. 72).

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook.* Chapman and Hall/CRC. (Cit. on p. 71).

Zeileis, A., & Grothendieck, G. (2005). Zoo: S3 Infrastructure for Regular and Irregular Time Series. *J. Stat. Softw.*, *14*, 1–27. https://doi.org/10.18637/jss.v014.i06 (cit. on pp. 63, 72).

Zerouali, B., Chettih, M., Alwetaishi, M., Abda, Z., Elbeltagi, A., Augusto Guimarães Santos, C., & E. Hussein, E. (2021). Evaluation of Karst Spring Discharge Response Using Time-Scale-Based Methods for a Mediterranean Basin of Northern Algeria. *Water*, *13*(21), 2946. https://doi.org/10.3390/w13212946 (cit. on p. 61).

Zhang, R., Chen, X., Zhang, Z., & Soulsby, C. (2020). Using hysteretic behaviour and hydrograph classification to identify hydrological function across the "hillslope–depression–stream" continuum in a karst catchment. *Hydrol. Process.*, *34*(16), 3464–3480. https://doi.org/10.1002/hyp.13793 (cit. on p. 61).

Part II

Improvement and evaluation of the performance of lumped parameter models for the simulation of spring discharge and water level of karst systems

Introduction to Part 2

This part aims to address the second, third and fourth research questions of the thesis (see ii., iii. and iv.) and is dedicated to the improvement and evaluation of the performance of lumped parameter models in karst hydrology. The main research questions addressed in this part are as follows:

i. **On which aspect can one-dimensional models be further improved?**

ii. Are elaborate objective functions appropriate for karst systems and do they improve model calibration?

iii. How to evaluate a model simulation in a purposeful and insightful way?

iv. **Can performance criteria be trusted for the calibration and evaluation of hydrological models?**

v. **What are the advantages and drawbacks of ANN and reservoir models in karst hydrology?**

Chapter 4 introduces newly developed features implemented in the KarstMod platform version 3.0. Improvements in one-dimensional lumped parameter models are achieved through better-suited input data and more sophisticated objective functions for calibration. New evaluation methods are proposed to assess various relevant aspects of the simulations.

Chapter 5 proposes a critical evaluation of several performance criteria including the Kling-Gupta Efficiency (KGE) and its variants. It shows how performance criteria, when based on relative parameters, can be misleading about the performance of a simulation. The aim is to raise awareness among modellers and promote a careful use of performance criteria for the evaluation of hydrological model performance.

Chapter 6 presents a comparison of the ANN and reservoir modelling approaches in karst hydrology. Considering several karst systems in different contexts, the study details the advantages and drawbacks of each approach with respect to input data, system characterisation, method limitations and model results.

Chapter 4

Improving model relevance: input data and performance criteria

We propose an updated version of KarstMod, an adjustable platform dedicated to lumped parameter rainfall-discharge modelling of karst aquifers. KarstMod provides a modular, user-friendly modelling environment for educational, research and operational purposes. It also includes numerical tools for time series analysis, model evaluation and sensitivity analysis. The modularity of the platform facilitates common operations related to lumped parameter rainfall-discharge modelling, such as (i) set up and parameter estimation of a relevant model structure, and (ii) evaluation of internal consistency, parameter sensitivity and hydrograph characteristics. The updated version now includes (i) external routines to better consider the input data and their related uncertainties, i.e. evapotranspiration and solid precipitation, (ii) enlargement of multi-objective calibration possibilities, allowing more flexibility in terms of objective functions as well as observation type, and (iii) additional tools for model performance evaluation including further performance criteria and tools for model errors representation.

This work has resulted in the updated version 3.0 of KarstMod (available online on the SNO KARST website) and in a scientific article, currently under review in Hydrology and Earth System Sciences (Sivelle et al., 2023).

Important note: The preprint of the article presented in this chapter was written by Vianney Sivelle. As second author, my contributions include module development, as well as the review and editing of various parts of the manuscript.

Article:

Sivelle, V., Cinkus, G., Mazzilli, N., Labat, D., Arfib, B., Massei, N., Cousquer, Y., Bertin, D., and Jourde, H.: Improvement of the KarstMod modeling platform for a better assessment of karst groundwater resources, Hydrol. Earth Syst. Sci., 1–26, https://doi.org/10.5194/hess-2023-17, 2023.

Contents

4.1 Introduction

Karst systems consist of heterogeneous aquifers characterised with the co-existence of three types of porosity: (i) inter-granular porosity, (ii) fracture porosity and (iii) large voids and conduits (Palmer, 1991) characterised by contrasted hydrodynamic properties. The existence of surface karst features such as shaft or swallow hole often leads to concentrated point-source recharge towards karst conduits in addition to the more common homogeneous diffuse recharge over the catchment. It also implies that flow regimes can be either laminar or turbulent. Karst aquifers constitute an essential source of drinking water for about 9.2 % of the world population (Stevanović, 2019) and it is estimated that one-quarter of the world population depends on freshwater from karst aquifers (Ford and Williams, 2007). Karst aquifers contain an important volume of freshwater while only 1 % of its annually renewable water is used for drinking water supply (Stevanović, 2019). Karst groundwater thus represents a unique opportunity to limit the increasing imbalance between growing demand and limited freshwater resource (Bierkens and Wada, 2019; Wada et al., 2016) in the present context of global change. However, karst aquifers are also particularly vulnerable to potential source of contamination, including emergent contaminants (Lukač Reberski et al., 2022), residues of phyto-sanitary products (Lorette et al., 2022) and wastewater (Doummar et al., 2022). Understanding the functioning of karst aquifers and developing operational tools to predict the evolution of freshwater resources is therefore a major challenge for the hydrological science community (Blöschl et al., 2019). Such tools are also required for a better assessment of groundwater vulnerability as well as sustainable management of the groundwater resources (Elshall et al., 2020).

KarstMod is an adjustable modelling platform (Mazzilli et al., 2019) dedicated to lumped parameter rainfall-discharge modelling allowing for (i) simulation of spring discharge, piezometric head and surface discharge, (ii) hydrodynamic analysis of the internal fluxes considered in the model, (iii) model performance evaluation and parametric sensitivity analysis. In this paper, we present the new features incorporated in KarstMod: (i) external routines to better consider the input data and their related uncertainties, i.e. evapotranspiration and solid precipitation, (ii) enlargement of multi-objective calibration possibilities, allowing more flexibility in terms of objective functions as well as observation type with the possibility to include surface water discharge in the calibration procedure and (iii) model performance evaluation, including additional performance criteria as well as additional tools for model errors representation such as the diagnostic efficiency plot (Schwemmle et al., 2021). Also, we present two cases studies to illustrate how KarstMod is useful in the framework of the assessment of karst groundwater resources and its sensitivity to groundwater abstraction. Section 4.2 is devoted to the presentation of the background and motivations to improve the functionalities of the platform while section Section 4.3 presents the main features of KarstMod. Section 4.4 illustrates the application of rainfall-discharge modelling using KarstMod within the Touvre (western France) and the Lez (southern France) karst systems, which both constitute strategic fresh water resources and ensure drinking water supply.

4.2 Background and motivations

4.2.1 Challenges in karst groundwater resources

Karst aquifers are affected by the combination of different components of global change such as (i) effects of climate change which are particularly pronounced in the Mediterranean area (Dubois et al., 2020; Nerantzaki and Nikolaidis, 2020), (ii) increasing groundwater abstraction (Labat et al., 2022), as well as (iii) changes in land cover land use (Bittner et al., 2018; Sarrazin et al., 2018). Therefore, the assessment of karst groundwater resources vulnerability in the present context requires operational tools for estimating the sustainable yield of karst aquifers but also to predict the impacts of climatic or anthropogenic forcing on groundwater resources in the long term (Sivelle et al., 2021). In order to address these issues, different modelling approaches have been developed (Jeannin et al., 2021) such as, among others, fully-distributed models (Chen and Goldscheider, 2014), semi-distributed models (Doummar et al., 2012; Dubois et al., 2020; Ollivier et al., 2020), and lumped parameter models (Mazzilli et al., 2019) including semi-distributed recharge (Bittner et al., 2018; Sivelle et al., 2022a). Among these, lumped parameter models are recognised as major tools to explore the ability of conceptual representations to explain observations in karst systems (Duran et al., 2020; Frank et al., 2021; Poulain et al., 2018; Sivelle et al., 2019) and for managing karst groundwater resources (Cousquer and Jourde, 2022; Labat et al., 2022; Sivelle and Jourde, 2021; Sivelle et al., 2021).

4.2.2 Challenges in lumped parameter modelling in karst hydrology

Lumped parameter models consist of a functional approach that analyses a hydrogeological system at the catchment scale and describes the transformation from rainfall into discharge using empirical or conceptual relationships. Therefore, parameter values or distributions cannot be determined directly from catchment physical characteristics or *in-situ* measurements, excepted the discharge coefficient to the spring that can be estimated on the basis of recession curve analysis. Instead, model parameters values must be estimated by history-matching. In a general way, rainfall-discharge models in karst hydrology are calibrated considering spring discharge measurements.

Former studies have shown the interest of considering various type of observations such as natural hydro-chemical tracers: NO_3 and SO_4 concentrations (Hartmann et al., 2013), electrical conductivity (Chang et al., 2021) or excess air (Sivelle et al., 2022b). Indeed, the consideration of complementary observation data in groundwater model calibration appears relevant in many applications (Schilling et al., 2019) but requires additional investigations before a suitable implementation in KarstMod. Therefore, in this paper, we will focus on the use of hydrodynamics observations only. Indeed, considering piezometric head variations in lumped parameter rainfall-discharge models may lead to better model performance (Cousquer and Jourde, 2022; Mazzilli et al., 2011). Nonetheless, the information content of the piezometric head time series (directly measured, or derived from ground-based gravity measurements) for lumped parameter rainfall-discharge models calibration purpose can be disputable when the available data is not adequate to characterize the whole catchment due to the important heterogeneity in karst aquifers (Mazzilli et al., 2013; Sivelle and Jourde, 2021). Also, Cousquer and Jourde (2022) accounted for the surface runoff in a lumped parameter rainfall-discharge model calibration procedure allowing

to reduce the parametric uncertainties.

Another key point in lumped parameter rainfall-discharge modelling concerns the evaluation of the meteorological forcing, i.e. precipitation (P) and evapotranspiration (ET). The transformation of precipitation into recharge and finally into discharge includes several processes with characteristic time covering several orders of magnitude (Blöschl and Sivapalan, 1995). Thus, the temporal resolution of the hydrological model must be suitable in the range of time and space scale where the physical phenomenons take place. Coupling hydrological models at multiple temporal resolutions can provide a better model consistency (Sivelle et al., 2019) since the transfer function in karst aquifers may present short response time. Also, errors in rainfall time series can significantly affect model parameters and structure (Oudin et al., 2006). Finally, the response of karst spring discharge is sensitive to energy and water fluxes within the soil-vegetation-atmosphere continuum as well as changes in climatic conditions (Hartmann et al., 2017). Bittner et al. (2021) computed several models to evaluate the fluxes related to interception, evapotranspiration and snow process. The results show significant uncertainties related to input data as well as potential compensation between the various uncertain processes. In some cases, snow melt is a controlling factor in the water balance (Doummar et al., 2018a; Liu et al., 2021), thus a suitable snow melt estimation is required to improve hydrological model performance (Çallı et al., 2022). Therefore, two meteorological modules have been added to KarstMod: (i) a "snow routine" and (ii) a "PET routine" allowing to better account for snow and evapotranspiration processes.

4.3 Implementation

The updated version of KarstMod implements additional features to enhance the rainfall-discharge modelling practices. First, we describe the additional modules (snow and PET routines) for a better meteorological forcing estimation. Then, we introduce the additional tools proposed for (i) set up and calibration of the model structure, (ii) model performance evaluation as well as (iii) uncertainties consideration.

4.3.1 Meteorological modules

4.3.1.a Snow routine

KarstMod allows using either observation-based precipitation time series P [L T^{-1}] or estimated precipitation time series P_{sr} [L T^{-1}] using a snow routine. The latter is similar to the one used by Chen et al. (2018) – without the radiation components – which has been successfully used for improving the simulation of karst spring discharge in snow-covered karst systems (Chen et al., 2018; Cinkus et al., 2023b). It consists of a modified HBV-snow routine (Bergström, 1992) for simulating snow accumulation and melt over different sub-catchments based on altitude ranges (Appendix 4.A). The estimated precipitation P_{sr} gives the water leaving the snow routine, equivalent to the recharge into the first compartment of the model (compartment E in KarstMod). P^*_{sr} for each sub-catchment is proportional to its surface regarding the complete catchment area. The snow routine workflow requires both air temperature T [°C] and precipitation P [L T^{-1}] time series. P is considered as snow when T in the sub-catchment is lower than the temperature threshold T_s. Snow melt starts when the temperature overpasses the threshold according to a degree-day expression. The snow melt is a function of the melt coefficient MF and

the degrees above the temperature threshold T_s. Runoff starts when the liquid water holding capacity of snow CWH is exceeded. The refreezing coefficient (CFR) stands for refreezing liquid water in the snow when snow melt is interrupted (Bergström, 1992). The output of the snow routine consists of a redistributed precipitation time series P_{sr}. The four parameters of the snow routine (i.e. T_s, MF, CWH and CFR) can be considered in the parameter estimation procedure as well as sensitivity analysis.

4.3.1.b PET routine

Evapotranspiration in KarstMod can be tackle in four different ways (Figure 4.1):

(a) Effective precipitation time series (P_{eff}) can be preprocessed by user (Eq. 4.1) and the evapotranspiration flux is not activated in the model structure selection window in KarstMod. Therefore, P_{eff} is given through the P time series in the input data file.

$$P_{eff} = P - ETa \tag{4.1}$$

where P_{eff} is effective precipitation [L T^{-1}], P is precipitation [L T^{-1}] and ETa is user-defined actual evapotranspiration [L T^{-1}] computed by observation-based data or external model.

(b) User defined potential evapotranspiration (PET) can be given as input in KarstMod for the evapotranspiration time series. Compartment E stands for a soil and epikarst storage zone, where water is available for actual evapotranpiration (ETa), flows to lower level of the model structure or outflow as surface discharge losses. Using E_{min}, user can simulate water holding capacity and non-linear behavior of karst recharge.

(c) User-defined actual evapotranspiration (ETa) can be given as input data in KarstMod for evapotranspiration time series instead of potential evapotranspiration. KarstMod computes effective precipitation by limiting the evapotranspiration to water content available in compartment E; calculated actual evapotranspiration can then be lower than user's input ETa.

(d) The new feature in KarstMod is the PET routine which estimates the potential evapotranspiration based on the Oudin's formula (Oudin et al., 2005) (Eq. 4.2). It needs a temperature time series and two parameters to be estimated, which can be considered in the parameter estimation procedure as well as sensitivity analysis.

$$PET = \frac{R_e}{\lambda.\rho} \times \frac{T + K2}{K1} \quad if \quad T + K2 > 0 \quad else \quad PET = 0 \tag{4.2}$$

where R_e is the extraterrestrial radiation [MJ L^{-2} T^{-1}] depending only on latitude Lat and Julian day, λ is the latent heat flux (taken equal to 2.45 MJ M^{-1}), ρ is the density of water [M L^{-3}] and T is the mean daily air temperature [°C], which is therefore a single function of the Julian day for a given location. $K1$ [°C] and $K2$ [°C] are constants to adjust over the catchment for rainfall-discharge model, which both can be considered in the parameter estimation procedure and sensitivity analysis.

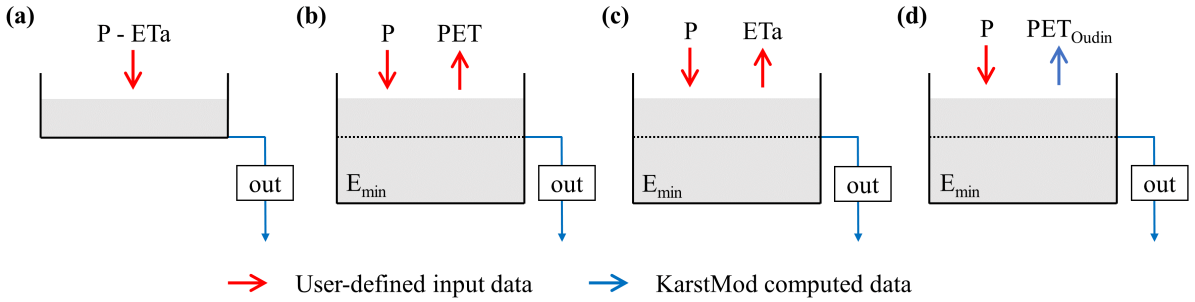


Figure 4.1: The four ways to account for evapotranspiration in KarstMod (P is precipitation, ETa is actual evapotranspiration, PET is potential evapotranspiration, PET_{Oudin} is KarstMod's computed potential evapotranspiration with Oudin's formula).

4.3.2 Set up and calibration of the model structure

The modular structure proposed in KarstMod is based on a widely used conceptual model which separates karst aquifers into an infiltration zone and a saturated zone. Based on this conceptual representation, the platform offers four compartments organised as a two-level structure: (i) compartment E (higher level) and (ii) compartments L, M and C (lower level). The modular structure proposed in KarstMod is based on a widely used conceptual model which separates karst aquifers into an infiltration zone and a saturated zone, or low and quick flows through the unsaturated and saturated zones. Based on this conceptual representations, the platform offers four compartments organised as a two-level structure: (i) compartment E (higher level) and (ii) compartments L, M and C (lower level). A priori, the higher-level stands for the infiltration zone or the soil and epikarst. At the lower level, compartments L, M, and C stand for the different subsystems of the saturated zone, or for low and quick flows of the whole hydrosystem. The various model structures and their governing equations are presented in Mazzilli et al. (2019) and Mazzilli et al. (2023).

The user can activate (or deactivate) the various compartments (E, L, M and C), the fluxes and their activation threshold as well as the exponent of the discharge law α (in case of non-linear discharge law such $\alpha \neq 1$). Figure 4.2 gives an example of model structure in KarstMod where the solid and faded colors represent the activated and the inactivated features respectively. The user must provide the warm-up, calibration and validation periods. The warm-up period must be set in order to be independent from initial conditions to avoid bias in the parameter estimation procedure (Mazzilli et al., 2012). Then, a calibration period (i.e. the period in which the parameter are estimated to reduce the predictive errors) and a validation period (i.e. period separated from the calibration period) can be defined to run the split sample test procedure (Klemeš, 1986). For calibration purpose, KarstMod proposes several widely used performance criteria ϕ: the Pearson's correlation coefficient r_p (Freedman et al., 2007), the Spearman rank correlation coefficient r_s (Freedman et al., 2007), the Nash-Sutcliffe Efficiency NSE (Nash and Sutcliffe, 1970), the volumetric error VE (Criss and Winston, 2008), the modified balance error BE (Perrin et al., 2001), the Kling-Gupta Efficiency KGE (Gupta et al., 2009) and a non-parametric variant of the Kling-Gupta Efficiency KGE_{NP} (Pool et al., 2018). To compute a multi-objective calibration procedure the user can create his own objective function Φ as a weighted sum of several objective functions:

$$\Phi = \sum_{i=1}^{N} \omega_i \times \phi_i(U) \tag{4.3}$$

where ω is the weight affected to the objective function $\phi(U)$ with $\sum w_i = 1$ and U a general notation for the observations used for parameter estimation purpose. In the KarstMod modelling platform U corresponds to either spring discharge Q_s, piezometric head measurements Z (available for compartments E, L, M and C) or surface water discharge Q_{loss} from compartment E. Also, the objective function ϕ can be computed on transformed U to avoid high water level bias on quadratic error. The following transformation are available in KarstMod: $1/U$, $\sqrt{U}$, $1/\sqrt{U}$. Therefore, the user can use any combination of the objective function ϕ, observations U and variable transformations. Depending on the modelling purpose, the user must refer to the literature to define the suitable objective function (Bennett et al., 2013; Ferreira et al., 2020; Hauduc et al., 2015; Jackson et al., 2019).

The model is calibrated using a quasi Monte-Carlo sampling procedure with a Sobol sequence sampling of the parameter space (Sobol, 1976). The procedure consists in finding an ensemble of parameter set providing an objective function ϕ greater than the user defined value. The calibration procedure is stopped when either the user-defined maximum duration t_{max} is reached or the user-defined number of parameter set n_{max} are collected. KarstMod offers a "run" option allowing to run the model for user-defined parameter set, without calibration procedure, and so allowing to investigate "by-hand" the parameter space and the sensitivity of the model to specific parameters.

4.3.3 Model evaluation

The model performance can be evaluated for both the calibration and validation periods. It allows (i) to ensure the robustness of model predictions, even under changing conditions (which is a key point for the assessment of climate change impact) and (ii) to avoid model over-fitting within a specific range of hydro-climatic conditions observed during the calibration period. KarstMod allows the computation of the above mentioned performance criteria for both calibration and validation periods. Even though the notation "validation" is disputable such procedure is required to evaluate both explanatory and predictive dimensions of the model structure (Andréassian, 2022). Then, KarstMod offers an ensemble of numerical tools devoted to (i) check the model consistency, i.e. explanatory dimension of the model (Beven, 2001; Shmueli, 2010), (ii) evaluate the model performance, i.e. predictive dimension of the model structure.

To check the model consistency, the simulation based on the parameter set that provides the highest objective function value can be analysed through an ensemble of graphs such as (i) internal and external fluxes as a function of time, (ii) cumulative volumes for both observed and simulated time series for spring discharge Q_s and surface water discharge Q_{loss}, (iii) simulated mass-balance as function of time, (iv) comparison of observations and simulations for either Q_s or Q_{loss} with probability function plots, auto-correlogram of the spring discharge time series, cross-correlogram of precipitation-discharge time series.

To evaluate the model performance, KarstMod offers a "Model evaluation" panel that includes several sub-panels, from the left to the right (see the KarstMod Graphical User Interface screenshot figure, Figure 4.4):

- The diagnostic efficiency DE (Schwemmle et al., 2021) which consists of a diagnostic polar plot that facilitates the model evaluation process as well as the comparison of multiple simulations. The DE accounts for constant, dynamics and timing errors, and their relative contribution to the model errors. Also, the decomposition of the errors between the periods of high flows and low flows allows to better investigate the model bias, as well as to provide critical evaluation for impact studies, particularly for the assessment of climate change impacts. Indeed, the accurate evaluation of low flow periods (in terms of frequency, intensity and duration) becomes more and more crucial for groundwater resource variability assessment.

- The available objective functions ϕ are presented as a radar chart which consists of a polygon where the position of each point from the center gives the value of the performance criteria. The closer the point is to the outside of the radar chart, the better the model performs. The radar chart is made for both calibration and validation periods and for each of the calibration variables considered in the modelling (Q_s, Z^A_{obs} with A for either E, M, C or L compartments and Q_{loss}).

- The KGE (Gupta et al., 2009) consists of a diagonal decomposition of the NSE (Nash and Sutcliffe, 1970) to separate Pearson's correlation coefficient r_p, representation of bias β_{KGE}, and variability α_{KGE}. Thus, the KGE is comparable to multi-objective criteria for calibration purpose (Pechlivanidis et al., 2013). The sub-panel offers (i) a bi-plot of the three KGE's components and (ii) a radar plot visualisation of the KGE's components, allowing to identify potential counterbalancing errors according to these different components (Cinkus et al., 2023a). The two above mentioned plots also include the decomposition of the $\mathrm{KGE_{NP}}$ (Pool et al., 2018) in terms of Spearman's rank correlation coefficient r_s, representation of bias $\beta_{KGE_{NP}}$ and non-parametric variability $\alpha_{KGE_{NP}}$.

4.3.4 Dealing with uncertainties

Moges et al. (2021) summarise the various source of uncertainties in hydrological models including structural and parametric uncertainties as well as uncertainties related to input data and observations. The latter concern both the input (i.e. precipitation and evapotranspiration) and the output (i.e. discharge) of the modeled systems. Many references are devoted to the uncertainties related to input data and observations. As an example, Westerberg et al. (2022) include information about the discharge uncertainty distribution in the objective function and perform better discharge simulation. Also, the precipitation error can be dependent on the data time step (McMillan et al., 2011) and could impact the hydrological model performance (Ficchì et al., 2016). KarstMod allows to perform hydrological modelling on both daily and hourly temporal resolutions, allowing to account for uncertainty related to the data time step. Lumped parameter hydrological models generally consider meteorological time series representative of a whole catchment, which may require some preprocessing, particularly for snow processes since it can have a strong influence on flow dynamics. Thus, KarstMod includes variables related to both the snow routine (i.e. the redistributed precipitation time series P_{sr}) and the PET routine (i.e. estimated potential evapotranspiration PET) in the parameter estimation procedure. This allows to investigate the sensitivity of the flow simulation to these input data, when using snow and PET routines. Nonetheless, KarstMod does not

include features to investigate the impact of observation uncertainties on the parameter estimation.

As many environmental problems, parameter estimation in rainfall-discharge modelling consists mainly in ill-posed problems, i.e. the modelling encounters issues about the unicity, identifiability and stability of the problem solution (Ebel and Loague, 2006). As a consequence, several representations of the modeled catchment may be considered as equally acceptable (Beven, 2006). Knoben et al. (2020) evaluate the performance of 36 daily lumped parameter models over 559 catchments and show that between 1 and up to 28 models can show performance close to the model structure with the highest performance criteria. Such results are widely covered in catchment hydrology (Dakhlaoui and Djebbi, 2021; Darbandsari and Coulibaly, 2020; Gupta and Govindaraju, 2019; Pandi et al., 2021; Zhou et al., 2021) but still poorly investigated in karst hydrology. Indeed, the structural uncertainty impacts on rainfall-discharge modelling in karst hydrology is not properly evaluated whereas many studies consider several hydrological model structure to include structural uncertainty in flow simulation (Hartmann et al., 2012; Jiang et al., 2007; Jones et al., 2006; Sivelle et al., 2021). KarstMod includes more than 50 combinations of the various compartments as well as various compartments' model (i.e. compartment with linear or non-linear discharge law and compartment with infinite characteristic time) and allows a quick implementation of the various model structures. The user can easily manage to start the modelling with one single compartment and gradually move to more complex model structure with up to 4 compartments, 5 fluxes connected to the spring, 4 internal fluxes and 1 flux running out of the system.

Considering each model structure, parametric equifinality can be investigated using (i) dotty plots of the values of the objective function against the parameter values, (ii) dotty plot of the values of the performance criteria used to define the aggregated objective function, and (iii) the variance-based, first-order S_i and total S_{Ti} sensitivity indexes for the model parameters. Details concerning the computation of sensitivity indexes within KarstMod are given in Mazzilli et al. (2019) and Mazzilli et al. (2023).

4.4 Case studies

To illustrate KarstMod application and the use of the above presented functionalities for the assessment of karst groundwater resources, we propose two case studies: (i) the Touvre karst system and (ii) the Lez karst system. Both karst systems consist of strategic freshwater resources for drinking water supply (DWS), for the city of Angoulême (western France) and Montpellier (southern France) respectively.

4.4.1 The Touvre karst system (La Rochefoucauld)

The Touvre karst system is a binary karst system where the infiltration consists of (i) a delayed infiltration of effective rainfall on karstic recharge area and (ii) a direct infiltration of surface water from the Tardoire, Bandiat, and Bonnieure rivers. These lasts are surface stream flow within metamorphic rocks that partly infiltrate to subterranean at the contact with sedimentary formations, mainly composed of Middle to Upper Jurrasic limestones. The springs of the Touvre, located 7 km east of Angoulême (western France), have three main outlets (the Bouillant, the Dormant and the Font de Lussac) and a secondary outlet (the Lèche) (Labat et al., 2022). In the following, the discharge of the four outlets are accumulated and named Touvre spring.

The Touvre karst system constitutes a strategic freshwater resource for drinking water supply (DWS) of Angoulême, with around 110,000 inhabitants, but also contributes to water supply for industry and agriculture. In 2015, there were 84 pumping wells over the karstic impluvium of the Touvre karst system, and around 100 more in the Tardoire, Bandiat, and Bonnieure rivers catchment. Based on the data provided by the Adour-Garonne Water Agency, the annual groundwater abstraction for agriculture represents 4.6 Mm^3 whereas annual groundwater abstraction for DWS represents 1.1 Mm^3 over the karstic impluvium of the Touvre karst system. On the three rivers catchment (out of the karstic impluvium), the annual groundwater abstraction represents 2.5 Mm^3 for agriculture and 3.3 Mm^3 for DWS, mainly through river intakes or alluvial groundwater abstraction. The total annual volume of abstracted groundwater in the area represents around 5 % of the annual volume of transit at the Touvre spring. This is quite low compared with karst aquifers in France exploited for their groundwater resource, such as the Lez spring (Jourde et al., 2014) and the Oeillal's spring karst catchment (Sivelle et al., 2021), where the annual groundwater abstraction volume represents respectively 50 % and 15 % of annual volume of transit at the spring. Therefore, the Touvre catchment seems not to be over exploited at the moment but the impact of groundwater abstraction should be addressed in the actual context of global change to ensure a sustainable management of this strategic fresh water resource.

The area is characterised by an ocean influenced climate with a mean annual precipitation around 800 mm distributed over 255 rainy days. The estimation is performed with Thiessen polygon methods based on eleven meteorological stations over the area (Labat et al., 2022). The mean annual potential evapotranspiration is around 770 mm according to the Penman-Monteith estimation provided by the french meteorological survey (Météo-France). The Touvre spring discharge shows a significant variability ranging from 3 $m^3\ s^{-1}$ to 49 $m^3\ s^{-1}$ with a coefficient of variation around 0.46 (Figure 4.3b). The surface stream flow rates for the Bonnieure, Bandiat and Tardoire rivers are concentrated within the autumn and winter periods. During the summer period, the discharge in the three rivers are very low (Figure 4.3c). The more significant groundwater abstraction is performed during the summer period, while the Touvre spring discharge reaches its lowest values within the late summer and early autumn periods (Figure 4.3c, Figure 4.3d).

The objective of the hydrological modelling is to assess the impact of groundwater abstraction on spring discharge, and more particularly during low flow periods (Labat et al., 2022). So, the calibration is performed according to the KGE_{NP} that improve the simulations during mean and low-flow conditions using the Spearman rank correlation due to it insensitivity to extreme values (Pool et al., 2018). The sampling procedure is set up to find $n_{obj} = 5000$ simulations with KGE_{NP} greater than 0.9. Afterwards the model is evaluated using the various features proposed in KarstMod (Figure 4.4). The diagnostic efficiency plot (Figure 4.4a) testifies of several elements: (i) the model seems to slightly overestimate high flow and underestimate low flow, (ii) the timing error is about 0.9, testifying of suitable flow dynamics in the model, (iii) low flow periods contribute more to the model errors, and (iv) there is no offset in the simulated spring hydrograph. The radar chart (Figure 4.4b) shows a good equilibrium between the various objective functions which values are greater than 0.8, excepted for the NSE criteria (NSE = 0.75). It is the consequence of the design of this criteria that tends to overweight the errors during floods. Here the NSE value still greater than 0.7 and testifies of a "very good" fit according to D. N. Moriasi et al. (2007). Finally, the decomposition of the KGE (Figure 4.4c, Figure 4.4d) shows $r_p = 0.91$, $\alpha = 1.15$ and $\beta = 1.02$ testifying of accurate dynamics

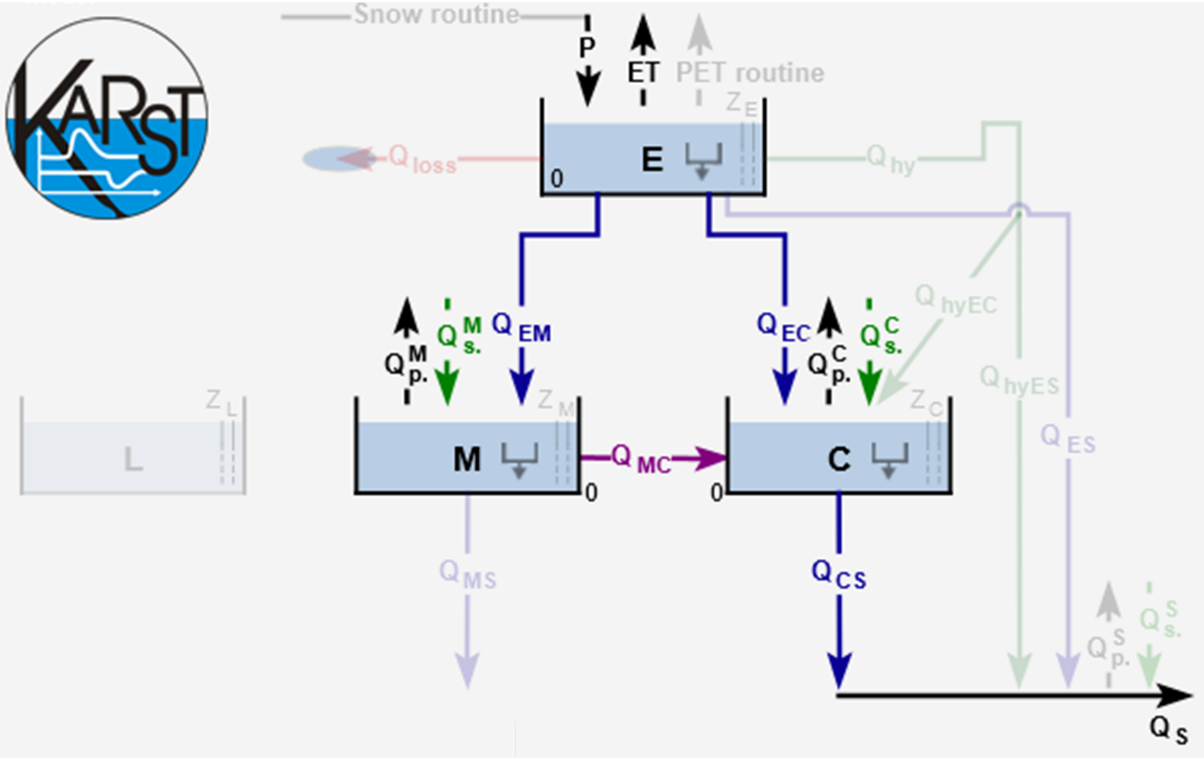


Figure 4.2: Screenshot of KarstMod with a focus on the panel "Model structure" for the Touvre karst system. The solid lines correspond to the activated fluxes whereas the faded color lines are not activated. $Q^M_{p.}$ and $Q^C_{p.}$ stand for groundwater abstraction that affects compartments M and C respectively while $Q^M_{s.}$ and $Q^C_{s.}$ stand for sinking flow that affects compartments M and C respectively.

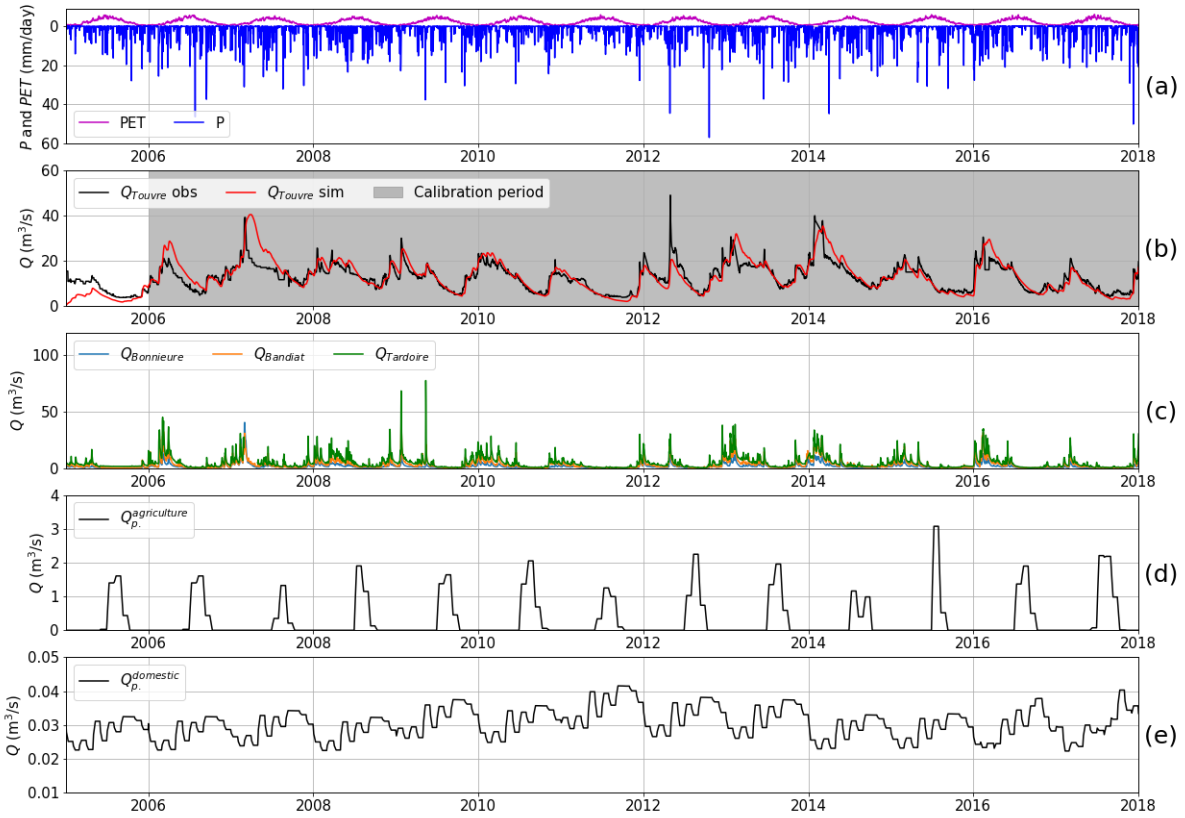


Figure 4.3: Daily time series for the Touvre system: a) precipitation (P) and potential evapotranspiration (PET), b) observed and simulated karst spring discharge (Q_{Touvre} obs and Q_{Touvre} sim), c) river streamflow discharge ($Q_{Bonnieure}$, $Q_{Bandiat}$, $Q_{Tardoire}$), d) and e) groundwater abstraction discharge ($Q_{p.}^{agriculture}$, $Q_{p.}^{domestic}$).

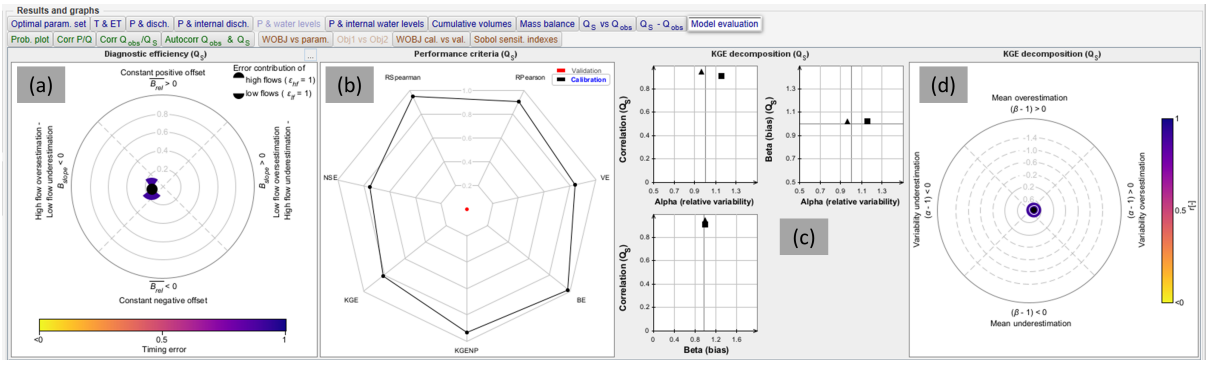


Figure 4.4: Screenshot of KarstMod with a focus on the sub-panel "Model evaluation". Application for the model evaluation on the Touvre system: (a) diagnostic efficiency plot (Schwemmle et al., 2021), (b) radar chart of the objective functions, (c) bi-plot of the KGE's (square) and KGE_{NP}'s (triangle) components, and (d) radar chart of the KGE's components.

and low bias, but slightly too high variability.

4.4.2 The Lez spring

The Lez spring (southern France) consists of the main outlet of a karst system encompassed in the North Montpellieran Garrigue hydrogeological unit delimited to the west by the Herault river, and to the north and east by the Vidourle river. The geology in the area corresponds to the Upper Jurassic layers separated by the Corconne-Matelle fault (oriented N30°), leading to two main compartments in the aquifer (Bérard, 1983; Clauzon et al., 2020). The karst aquifer is unconfined in the western compartment and is locally confined in the eastern compartment. The Lez spring is located about 15 km north of Montpellier. It is of Vauclusian-type with an overflow level at 65 m asl, and a maximum discharge of approximately 15 m^3 s^{-1}. The area is characterised by a typical Mediterranean climate with dry summers and rainy autumns. Over the 2009–2019 period, the mean annual precipitation is around 900 mm distributed over 133 rainy days (estimation with Thiessen's polygon methods based on four meteorological stations over the area: Prades-le-Lez, Saint-Martin-de-Londres, Sauteyrargues and Valflaunès), a mean annual potential evapotranspiration is around 900 mm according to the estimation based on Oudin's formula with the temperature measured at Prades-le-Lez station while the mean annual evapotranspiration is around 450 mm (eddy covariance flux-station of Puéchabon).

Since 1854, the Lez spring supplies the drinking water to Montpellier city and the surroundings. It currently constitutes the main fresh water resource for around 350,000 people in the area. The present water management scheme allows pumping at higher rates than the natural spring discharge during low flow periods, while supplying a minimum discharge rate ($\backsim$ 0.23 m^3 s^{-1}) into the Lez river to ensure ecological flow downstream, and reducing flood hazards via rainfall storage in autumn (Avias, 1995; Jourde et al., 2014). The pumping plant was built in 1982 with four deep wells drilled to intercept the karst conduit feeding the spring, 48 m below the overflow level of the spring. Pumping in these wells allows up to 1.8 m^3 s^{-1} to be withdrawn under low flow periods (with an authorized maximum drawdown of 30 m), while the average annual pumping flow rate is about 1.010 m^3 s^{-1} (over the 2008–2019 period). Due to the pumping management of the aquifer, which supplies about 30 to 35 Mm^3 of water per year to the metropolitan area of Montpellier, the discharge at the Lez spring is often low or nil.

In the present context of global change, Mediterranean karst systems already show significant decrease in spring discharge (Doummar et al., 2018b; Dubois et al., 2020; Fiorillo et al., 2012; Hartmann et al., 2012; Nerantzaki and Nikolaidis, 2020; Smiatek et al., 2013) which could be aggravated with groundwater abstraction (Sivelle et al., 2021). The Lez spring is strongly exposed to global change impact: (i) the Mediterranean area is identified as a climate change hot-spot (Diffenbaugh and Giorgi, 2012) where the projected warming spans 1.83–8.49°C according to CMIP6 and 1.22–6.63°C according to CMIP5 during the summer period (Cos et al., 2022), and (ii) the water management scheme will have to adapt to the future need in drinking water for the growing population in the area as well as changes in the fresh water consumption practice (e.g. water use restriction order). Therefore, a sustainable water management plan for the Lez spring requires a good appreciation of the hydrological functioning as well as operational hydrological model to properly address impacts studies. In this framework, KarstMod allows choosing and calibrating a suitable model structure. This constitutes a first step for global change impact study that requires prediction tools to simulate the aquifer response to various external forcing.

Figure 4.5 shows the model structure for the Lez karst catchment (Mazzilli et al., 2011) that consists of three compartments organised in two levels. The upper level corresponds to compartment E and represents the unsaturated part of the system, including a soil water holding capacity E_{min} and a discharge lost from the compartment Q_{loss}. The compartment E is exposed to precipitation P and evapotranspiration ET and discharge towards the lower level of the model structure starts when the water level exceeds the water holding E_{min}. The lower level consists of two inter-connected compartments M and C allowing to reproduce the lateral exchanges, denoted Q_{MC}, between transmissive function (compartment C) and capacitive function (compartment M) of the karst aquifer. Both M and C compartments are considered bottomless, allowing to reproduce period of non-overflow at the Lez spring when the mean water level in the aquifer stands below 65 m a.s.l., mainly during summer periods due to pumping in the karst conduit. Figure 4.6a and Figure 4.6b show the various daily time series required for the hydrological modelling of the Lez karst system (i.e. P, ET and Q_{pump}).

The available hydrological observations for model calibration consist of spring discharge Q_S, piezometric head measurement Z_C at the Lez spring and surface water discharge from secondary outlets and intermittent springs Q_{loss} (Figure 4.6c, Figure 4.6d and Figure 4.6e). The surface water discharge is estimated as the difference in discharge measured at the Lavalette station (15 km downstream the Lez spring) and the discharge measured at the Lez spring, as performed in Cousquer and Jourde Cousquer and Jourde (2022). Therefore, Q_{loss} includes all the water loss from the epikarst within several seasonal overflowing springs (i.e. Lirou spring, Restinclière spring and Fleurette spring). KarstMod allows to easily handle with the various parameter estimation depending on the considered hydrological observations (i.e. spring discharge, piezometric head measurement, and surface discharge from the epikarst). The sampling procedure is set up to find $n_{obj} = 5000$ simulations with an aggregated objective function Φ greater than 0.6. As suggested by Cousquer and Jourde (2022), using complementary hydrological observations in addition to the spring discharge allows to reduce the parametric uncertainties in the modelling of the Lez spring discharge. Therefore, using a multi-objective calibration procedure implemented in KarstMod, the objective function is build such as:

$$\Phi = \frac{1}{3} \times NSE(Q_s) + \frac{1}{3} \times NSE(Z_C) + \frac{1}{3} \times NSE(Q_{loss}) \tag{4.4}$$

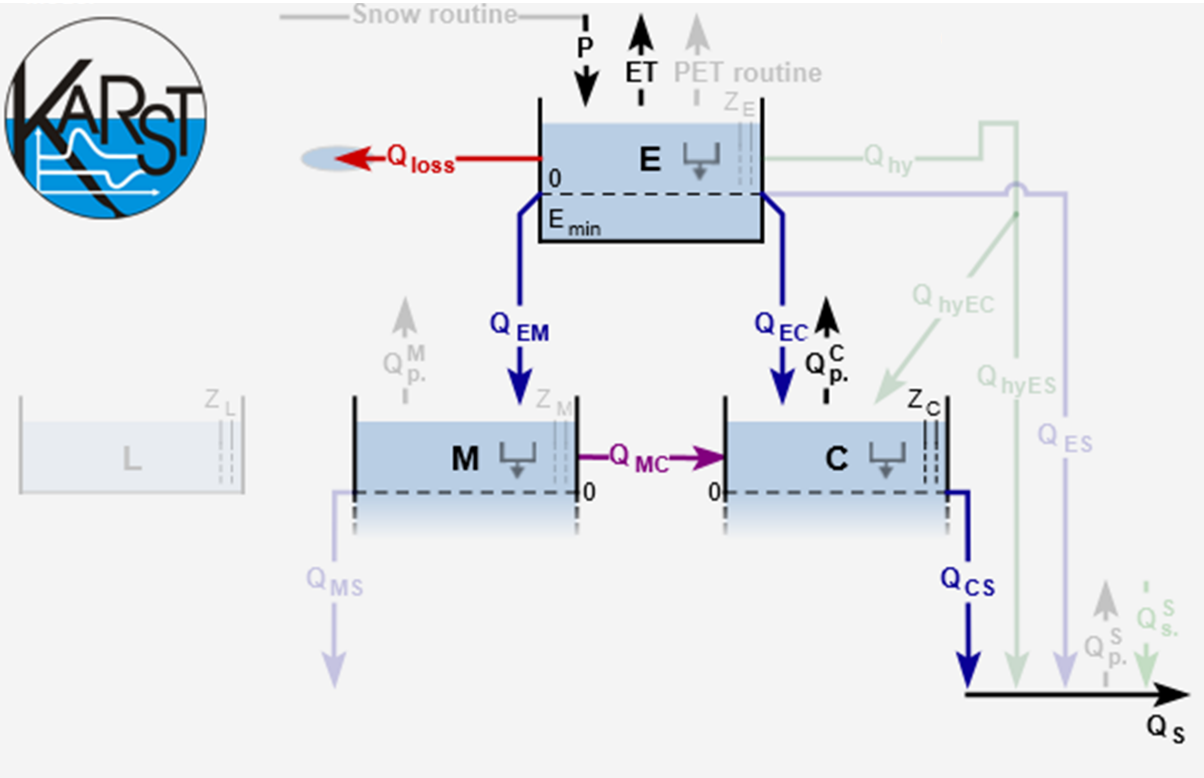


Figure 4.5: Screenshot of KarstMod with a focus on the panel "Model structure" for the Lez karst system. The solid lines correspond to the activated fluxes whereas the faded color lines are not activated. Q_{loss} stands for the surface discharge from the epikarst compartment, $Q^C_{p.}$ stands for groundwater abstraction that affects compartments C while Z_C stands for piezometric head measurements considered as representative of the compartment C.

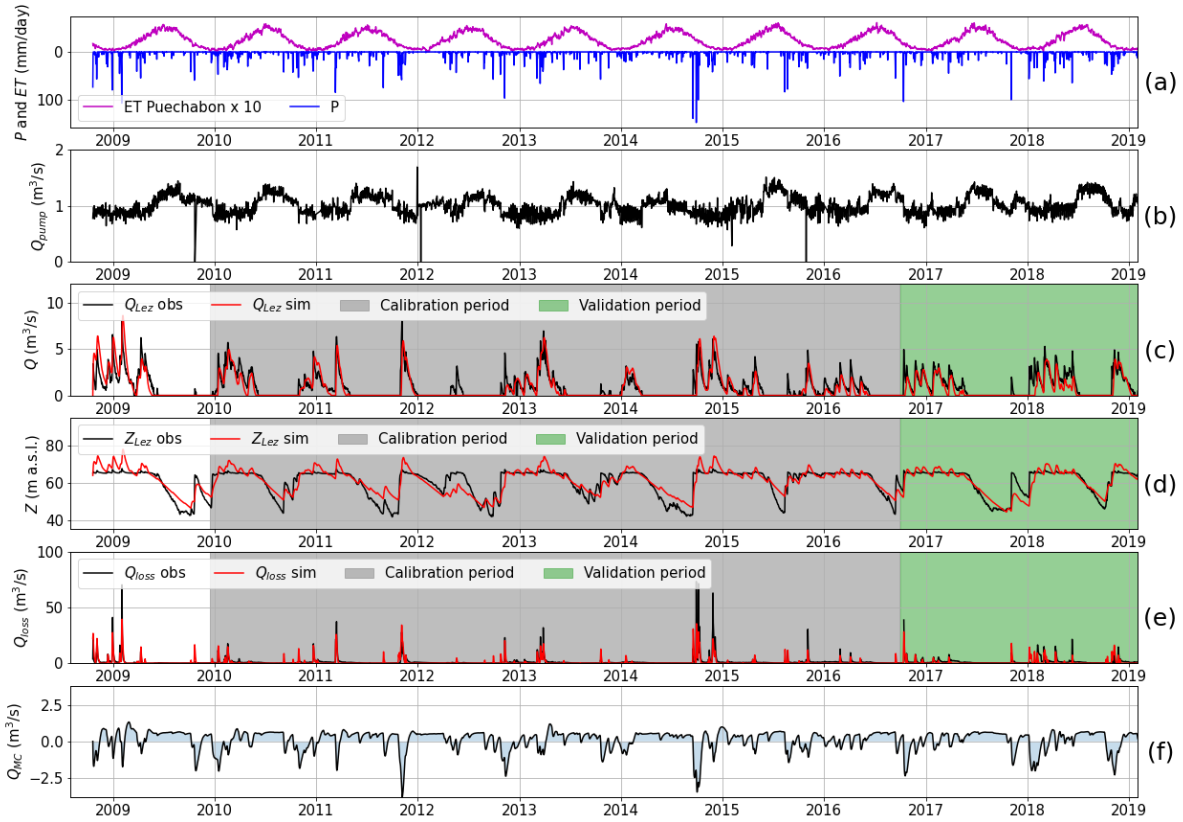


Figure 4.6: Daily time series for the Lez system: a) precipitations (P) and evapotranspiration (ET), b) groundwater abstraction (Q_{pump}), c) observed and simulated karst spring discharge (Q_{Lez} obs and Q_{Lez} sim), d) observed and simulated piezometric head (Z_{Lez} obs and Z_{Lez} sim), e) surface water discharge (Q_{loss}) and f) simulated exchanges fluxes between compartment M and C (Q_{MC}).

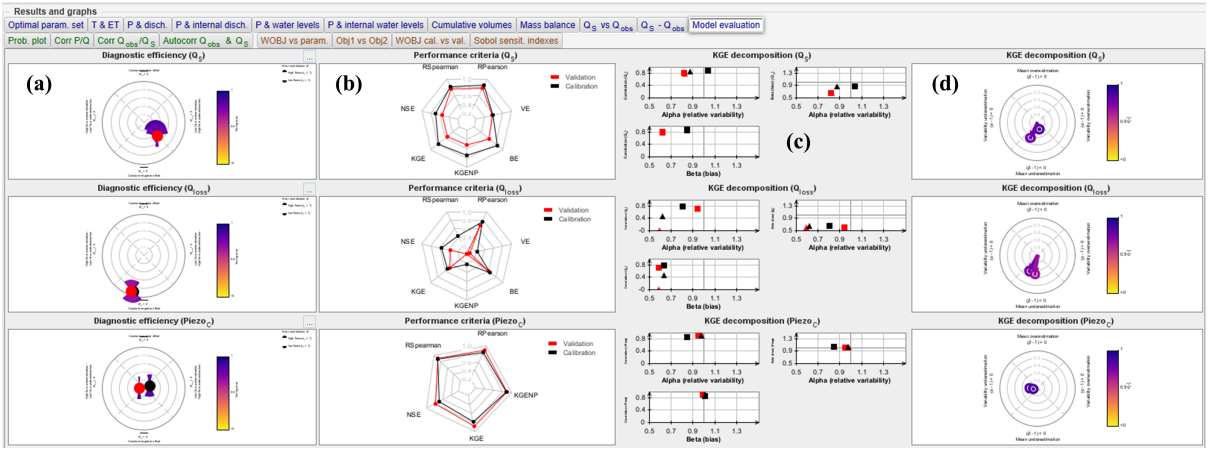


Figure 4.7: Screenshot of KarstMod with a focus on the sub-panel "Model evaluation". Application for the model evaluation on the Lez system. The panel is composed such as (i) each row correspond to the variable for calibration (Q_S, Q_{loss} and Z_C) and (ii) each columns corresponds to (a) diagnostic efficiency plot, (b) radar plots, one should note that VE and BE are not computed according to the piezometric time series, (c) decomposition of KGE (square) and KGE_{NP} (triangle) and (d) radar plot of the KGE decomposition.

The calibration procedure leads to an optimal $\Phi = 0.65$ decomposed such as $\phi\, Q_s = 0.70$, $\phi\, Z_C = 0.57$ and $\phi\, Q_{loss} = 0.70$ within the calibration period. Model performance evaluation on the validation period shows suitable model performance for both spring discharge and piezometric with $\phi\, Q_s = 0.54$ and $\phi\, Z_C = 0.79$, but poor model performance according to the surface water discharge with $\phi\, Q_{loss} = 0.36$. Afterwards the results can be evaluated using the various features proposed in KarstMod (Figure 4.7). The results show higher model performances for Q_S and Z_C than for Q_{loss}. The model performance appears quite satisfactorily concerning the variable of interest to assess the impact of water management scheme on the groundwater resources within the Lez aquifer.

The simulated exchanges fluxes between compartment M and C (Figure 4.6f) show consistent dynamics with the observations. Indeed, during periods of high flow the exchange fluxes are oriented from the compartment C to compartment M (i.e. $Q_{MC} < 0$). Significant precipitation events lead to rapid rises in piezometric head, saturation of the transmissive part of the aquifer and finally the establishment of overflow at the Lez spring (i.e. $Q_S > 0$) as well as the overflowing springs (i.e. $Q_{loss} > 0$). Conversely, during the periods of low piezometric head (i.e. both Q_S and Q_{loss} are nil), the simulated exchange fluxes are oriented from compartment M to compartment C (i.e. $Q_{MC} < 0$). Such flow exchanges between capacitive and transmissive part of karst aquifers has been evidenced using KarstMod on other karst environment (Duran et al., 2020; Frank et al., 2021; Labat et al., 2022; Sivelle et al., 2019).

4.5 Conclusions

KarstMod consists in a useful tool for the assessment of karst groundwater variability and sensitivity to anthropogenic pressures (e.g. groundwater abstraction). This tool is devoted to promote good practices in hydrological modelling for learning and occasional users. KarstMod requires no programming skills and offers a user-friendly interface allowing any user to easily handle hydrological modelling. As a first step, KarstMod can

be used to explore the ability of conceptual representations to explain observations such as discharge or piezometric heads in karst systems. A more advanced use of KarstMod is also possible as it provides a complete framework for (i) primary analysis of the data, (ii) comparison of various model structures, (iii) evaluation of the hydrological model performance as well as (iv) first assessment of parametric uncertainties. The research community increasingly uses KarstMod to address various challenges in karst hydrology, from understanding hydrological processes to practical applications such as evaluation of groundwater management plan, or even assessment of the impact of groundwater abstraction and climate changes on karst groundwater resources.

Future developments of KarstMod might include: (i) the consideration of land cover land use to consider the spatial heterogeneity in recharge processes (Sivelle et al., 2022a), (ii) the simulation of electrical conductivity (Chang et al., 2021), major ions concentration (Hartmann et al., 2013) or natural tracer such as air excess (Sivelle et al., 2022b), and (iii) the assessment of structural uncertainty (Cousquer, 2022). KarstMod should tend toward an open-source research software to avoid duplication of efforts in karst hydrological modelling. Also, a Python version is required for a better connection with additional framework for sensitivity analysis such as SAFE toolbox (Pianosi et al., 2015) and for model calibration procedure such as particle swarm optimization (Kennedy and Eberhart, 1995; Lee, 2014). Finally, the development of the KarstMod modelling platform will benefit better transparency and repeatability with an open-source approach, as observed on other numerical tools (Pianosi et al., 2020).

Acknowledgements

This platform is developed within the framework of the KARST observatory network (SNO KARST) initiative from the INSU/CNRS (France), which aims to strengthen knowledge-sharing and promote crossdisciplinary research on karst systems at the national scale. This work, as well as V. Sivelle post-doctoral position, was supported by the European Commission through the Partnership for Research and Innovation in the Mediterranean Area (PRIMA) program under Horizon 2020 (KARMA project, grant agreement number 01DH19022A).

Author contribution

V. Sivelle: methodology, software, writing — original draft. G. Cinkus: methodology, software, writing — review and editing. N. Mazzilli: methodology, software, project administration, writing — review and editing. H. Jourde: methodology, software, project administration, funding acquisition, writing—review and editing. D. Labat: methodology, software, writing — review and editing. B. Arfib: methodology, software, writing — review and editing. N. Massei: methodology, software, writing — review and editing. Y. Cousquer: writing — review and editing. D. Bertin: methodology, software, writing — review and editing.

Code and data availability

The KarstMod modelling platform is developed and made freely accessible within the framework of the KARST observatory network (SNO KARST) initiative from the INSU/CNRS. The platform can be downloaded on the SNO KARST website: https://sokarst.org/en/softwares-en/karstmod-en/.

4.6 Appendix

4.A Snow routine

P_{sr}^{*} (liquid water leaving the routine) is estimated for each time step t based on the precipitation P and air temperature T time series for each sub-catchment (Figure 4.8). The total snow routine output P_{sr} is calculated as a weighted sum of P_{sr}^{*} time series:

$$P_{sr} = \sum_{i}^{N} P_{sr\,i}^{*} \times p_i \tag{4.5}$$

where p_i is the proportion of the sub-catment i regarding the complete catchment area such as $\sum p_i = 1$, and N is total number of sub-catchments.

The snow routine requires four parameters, whose values are the same for all sub-catchments: the snowmelt temperature threshold T_s [°C], the melt factor MF [L T^{-1} °C^{-1}], the refreezing factor CFR [-], and the water holding capacity of snow CWH [-]. The snow routine allows estimating P_{sr}^{*} according to Algorithm 4.1.

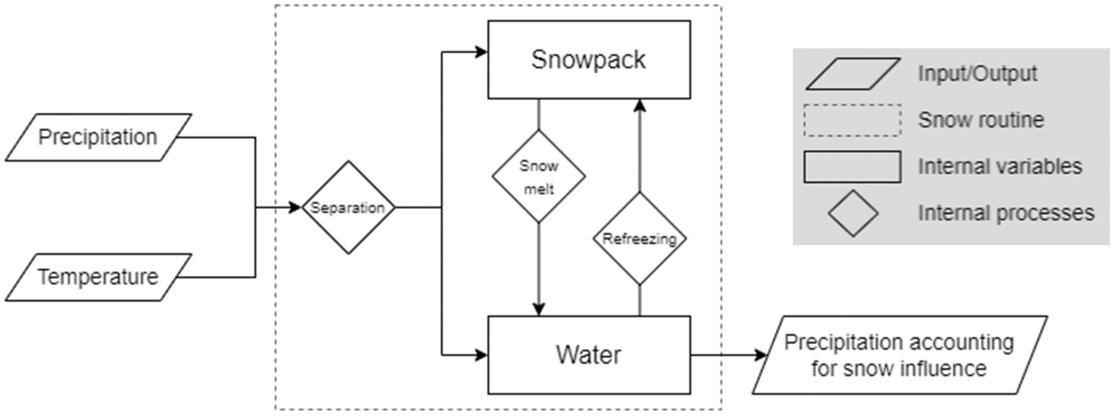


Figure 4.8: Snow routine workflow.

Algorithm 1: Estimating P_{sr}^{*} in sub-catchment

With P_{sr}^{*} = water leaving the routine/recharge to the soil (mm/dt), T_a = active temperature for snowmelt (°C), T_n = active temperature for refreezing (°C), m = snow melt (mm/dt), rfz = refreezing (mm/dt), v = solid component of snowpack depth (mm), vl = liquid component of snowpack depth (mm), and dt = temporal resolution.

for t=*1* **to** t_{max} **do**
- $m[t] = min((MF \times T_a[t]), v[t])$
- $rfz[t] = min(CFR \times MF \times T_n[t], vl[t])$
- $v[t+dt] = v[t] - m[t] + snow[t] + rfz[t]$
- **if** $vl[t+dt] > CWH \times v[t+dt]$ **then**
  - $P_{sr}^{*}[t] = vl[t+dt] - CWH \times v[t+dt]$
  - $vl[t+dt] = CWH \times v[t+dt]$
- **else**
  - $P_{sr}^{*}[t] = 0$
- **end**

end

4.7 References for Chapter 4

Andréassian, V. (2022). On the (im)possible validation of hydrogeological models. *Comptes Rendus. Géoscience*, *355*(S1), 1–9. https://doi.org/10.5802/crgeos.142 (cit. on p. 90).

Avias, J. V. (1995). Gestion active de l'exsurgence karstique de la Source du Lez (Hérault, France) 1957-1994. *Hydrogéologie (Orléans)*, *1*, 113–127 (cit. on p. 96).

Bennett, N. D., Croke, B. F., Guariso, G., Guillaume, J. H., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. *Environ. Model. Softw.*, *40*, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011 (cit. on p. 90).

Bérard, P. (1983). *Alimentation en eau de la ville de Montpellier : Captage de la source du Lez. Études des relations entre la source et son réservoir aquifère. Rapport n°2 : Définition des unités hydrogéologiques.* (BRGM No. 83). (Cit. on p. 96).

Bergström, S. (1992). The HBV model - its structure and applications. *SMHI*, *4*, 32 (cit. on pp. 87, 88).

Beven, K. (2001). On explanatory depth and predictive power. *Hydrol. Process.*, *15*(15), 3069–3072. https://doi.org/10.1002/hyp.500 (cit. on p. 90).

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007 (cit. on p. 92).

Bierkens, M. F. P., & Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: A review. *Environ. Res. Lett.*, *14*(6), 063002. https://doi.org/10.1088/1748-9326/ab1a5f (cit. on p. 85).

Bittner, D., Narany, T. S., Kohl, B., Disse, M., & Chiogna, G. (2018). Modeling the hydrological impact of land use change in a dolomite-dominated karst system. *Journal of Hydrology*, *567*, 267–279. https://doi.org/10.1016/j.jhydrol.2018.10.017 (cit. on p. 86).

Bittner, D., Richieri, B., & Chiogna, G. (2021). Unraveling the time-dependent relevance of input model uncertainties for a lumped hydrologic model of a pre-alpine karst system. *Hydrogeol. J.*, *29*(7), 2363–2379. https://doi.org/10.1007/s10040-021-02377-1 (cit. on p. 87).

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrol. Process.*, *9*(3-4), 251–290. https://doi.org/10.1002/hyp.3360090305 (cit. on p. 87).

Blöschl, G., Bierkens, M. F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., ... Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrol. Sci. J.*, *64*(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507 (cit. on p. 85).

Çallı, S. S., Çallı, K. Ö., Tuğrul Yılmaz, M., & Çelik, M. (2022). Contribution of the satellite-data driven snow routine to a karst hydrological model. *Journal of Hydrology*, *607*, 127511. https://doi.org/10.1016/j.jhydrol.2022.127511 (cit. on p. 87).

Chang, Y., Hartmann, A., Liu, L., Jiang, G., & Wu, J. (2021). Identifying More Realistic Model Structures by Electrical Conductivity Observations of the Karst Spring. *Water Res.*, *57*(4). https://doi.org/10.1029/2020WR028587 (cit. on pp. 86, 101).

Chen, Z., & Goldscheider, N. (2014). Modeling spatially and temporally varied hydraulic behavior of a folded karst system with dominant conduit drainage at catchment scale, Hochifen–Gottesacker, Alps. *J. Hydrol.*, *514*, 41–52 (cit. on p. 86).

Chen, Z., Hartmann, A., Wagener, T., & Goldscheider, N. (2018). Dynamics of water fluxes and storages in an Alpine karst catchment under current and potential future climate conditions. *Hydrol. Earth Syst. Sci.*, *22*(7), 3807–3823. https://doi.org/10.5194/hess-22-3807-2018 (cit. on p. 87).

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., & Goldscheider, N. (2023a). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrol. Earth Syst. Sci.*, *27*(13), 2397–2411. https://doi.org/10.5194/hess-27-2397-2023 (cit. on p. 91).

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J. A., Andreo, B., Goldscheider, N., & Jourde, H. (2023b). Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions. *Hydrol. Earth Syst. Sci.*, *27*(10), 1961–1985. https://doi.org/10.5194/hess-27-1961-2023 (cit. on p. 87).

Clauzon, V., Mayolle, S., Leonardi, V., Brunet, P., Soliva, R., Marchand, P., Massonnat, G., Rolando, J.-P., & Pistre, S. (2020). Fault zones in limestones: Impact on karstogenesis and groundwater flow (Lez aquifer, southern France). *Hydrogeol. J.*, *28*(7), 2387–2408. https://doi.org/10.1007/s10040-020-02189-9 (cit. on p. 96).

Cos, J., Doblas-Reyes, F., Jury, M., Marcos, R., Bretonnière, P.-A., & Samsó, M. (2022). The Mediterranean climate change hotspot in the CMIP5 and CMIP6 projections. *Earth Syst. Dynam.*, *13*(1), 321–340. https://doi.org/10.5194/esd-13-321-2022 (cit. on p. 97).

Cousquer, Y. (2022). *Estimating the Structural Uncertainty of Lumped Parameter Models in Karst Hydrology: A Bayesian Model Averaging (BMA), Tech. Rep. IAHS2022-522* (Oral presentation). Montpellier. (Cit. on p. 101).

Cousquer, Y., & Jourde, H. (2022). Reducing uncertainty of karst aquifer modeling with complementary hydrological observations for the sustainable management of groundwater resources. *Journal of Hydrology*, *612*, 128130. https://doi.org/10.1016/j.jhydrol.2022.128130 (cit. on pp. 86, 97).

Criss, R. E., & Winston, W. E. (2008). Do Nash values have value? Discussion and alternate proposals. *Hydrol. Process.*, *22*(14), 2723–2725. https://doi.org/10.1002/hyp.7072 (cit. on p. 89).

D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, & T. L. Veith. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153 (cit. on p. 93).

Dakhlaoui, H., & Djebbi, K. (2021). Evaluating the impact of rainfall–runoff model structural uncertainty on the hydrological rating of regional climate model simulations. *J. Water Clim. Change*, *12*(8), 3820–3838. https://doi.org/10.2166/wcc.2021.004 (cit. on p. 92).

Darbandsari, P., & Coulibaly, P. (2020). Inter-comparison of lumped hydrological models in data-scarce watersheds using different precipitation forcing data sets: Case study of Northern Ontario, Canada. *J. Hydrol. Reg. Stud.*, *31*, 100730. https://doi.org/10.1016/j.ejrh.2020.100730 (cit. on p. 92).

Diffenbaugh, N. S., & Giorgi, F. (2012). Climate change hotspots in the CMIP5 global climate model ensemble. *Clim. Change*, *114*(3-4), 813–822. https://doi.org/10.1007/s10584-012-0570-x (cit. on p. 97).

Doummar, J., Fahs, M., Aoun, M., Elghawi, R., Othman, J., Alali, M., & Kassem, A. H. (2022). Assessment of Water Quality and Quantity of Springs at a Pilot-Scale: Applications in Semiarid Mediterranean Areas in Lebanon. In M. J. Currell & B. G. Katz (Eds.), *Geophysical Monograph Series* (1st ed., pp. 111–129). Wiley. https://doi.org/10.1002/9781119818625.ch8 (cit. on p. 85).

Doummar, J., Hassan Kassem, A., & Gurdak, J. J. (2018a). Impact of historic and future climate on spring recharge and discharge based on an integrated numerical modelling approach: Application on a snow-governed semi-arid karst catchment area. *J. Hydrol.*, *565*, 636–649. https://doi.org/10.1016/j.jhydrol.2018.08.062 (cit. on p. 87).

Doummar, J., Margane, A., Geyer, T., & Sauter, M. (2018b). Assessment of key transport parameters in a karst system under different dynamic conditions based on tracer experiments: The Jeita karst system, Lebanon. *Hydrogeol. J.*, *26*(7), 2283–2295. https://doi.org/10.1007/s10040-018-1754-x (cit. on p. 97).

Doummar, J., Sauter, M., & Geyer, T. (2012). Simulation of flow processes in a large scale karst system with an integrated catchment model (Mike She) – Identification of relevant parameters influencing spring discharge. *J. Hydrol.*, *426–427*, 112–123. https://doi.org/10.1016/j.jhydrol.2012.01.021 (cit. on p. 86).

Dubois, E., Doummar, J., Pistre, S., & Larocque, M. (2020). Calibration of a lumped karst system model and application to the Qachqouch karst spring (Lebanon) under climate change conditions. *Hydrol. Earth Syst. Sci.*, *24*(9), 4275–4290. https://doi.org/10.5194/hess-24-4275-2020 (cit. on pp. 86, 97).

Duran, L., Massei, N., Lecoq, N., Fournier, M., & Labat, D. (2020). Analyzing multi-scale hydrodynamic processes in karst with a coupled conceptual modeling and signal decomposition approach. *J. Hydrol.*, *583*, 124625. https://doi.org/10.1016/j.jhydrol.2020.124625 (cit. on pp. 86, 100).

Ebel, B. A., & Loague, K. (2006). Physics-based hydrologic-response simulation: Seeing through the fog of equifinality. *Hydrol. Process.*, *20*(13), 2887–2900. https://doi.org/10.1002/hyp.6388 (cit. on p. 92).

Elshall, A. S., Arik, A. D., El-Kadi, A. I., Pierce, S., Ye, M., Burnett, K. M., Wada, C. A., Bremer, L. L., & Chun, G. (2020). Groundwater sustainability: A review of the interactions between science and policy. *Environ. Res. Lett.*, *15*(9), 093004. https://doi.org/10.1088/1748-9326/ab8e8c (cit. on p. 85).

Ferreira, P. M. D. L., Paz, A. R. D., & Bravo, J. M. (2020). Objective functions used as performance metrics for hydrological models: State-of-the-art and critical analysis. *Rev. Bras. Recur. Hidr.*, *25*, e42. https://doi.org/10.1590/2318-0331.252020190155 (cit. on p. 90).

Ficchì, A., Perrin, C., & Andréassian, V. (2016). Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events. *J. Hydrol.*, *538*, 454–470. https://doi.org/10.1016/j.jhydrol.2016.04.016 (cit. on p. 91).

Fiorillo, F., Revellino, P., & Ventafridda, G. (2012). Karst aquifer draining during dry periods. *J. Cave Karst Stud.*, *74*(2), 148–156. https://doi.org/10.4311/2011JCKS0207 (cit. on p. 97).

Ford, D., & Williams, P. (2007). Karst Hydrogeology. In *Karst Hydrogeology and Geomorphology* (pp. 103–144). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118684986.ch5 (cit. on p. 85).

Frank, S., Goeppert, N., & Goldscheider, N. (2021). Improved understanding of dynamic water and mass budgets of high-alpine karst systems obtained from studying a well-defined catchment area. *Hydrol. Process.*, *35*(4), e14033. https://doi.org/10.1002/hyp.14033 (cit. on pp. 86, 100).

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics: Fourth International Student Edition*. W. W. Norton & Company. (Cit. on p. 89).

Gupta, A., & Govindaraju, R. (2019). Propagation of structural uncertainty in watershed hydrologic models. *J. Hydrol.*, *575*, 66–81. https://doi.org/10.1016/j.jhydrol.2019.05.026 (cit. on p. 92).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003 (cit. on pp. 89, 91).

Hartmann, A., Wagener, T., Rimmer, A., Lange, J., Brielmann, H., & Weiler, M. (2013). Testing the realism of model structures to identify karst system processes using water quality and quantity signatures: KARST SYSTEM IDENTIFICATION. *Water Resour. Res.*, *49*(6), 3345–3358. https://doi.org/10.1002/wrcr.20229 (cit. on pp. 86, 101).

Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proc. Natl. Acad. Sci.*, *114*(11), 2842–2847. https://doi.org/10.1073/pnas.1614941114 (cit. on p. 87).

Hartmann, A., Lange, J., Aguado, A. V., Mizyed, N., Smiatek, G., & Kunstmann, H. (2012). A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *J. Hydrol.*, *468–469*, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024 (cit. on pp. 92, 97).

Hauduc, H., Neumann, M., Muschalla, D., Gamerith, V., Gillot, S., & Vanrolleghem, P. (2015). Efficiency criteria for environmental model quality assessment: A review and its application to wastewater treatment. *Environ. Model. Softw.*, *68*, 196–204. https://doi.org/10.1016/j.envsoft.2015.02.004 (cit. on p. 90).

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environ. Model. Softw.*, *119*, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001 (cit. on p. 90).

Jeannin, P.-Y., Artigue, G., Butscher, C., Chang, Y., Charlier, J.-B., Duran, L., Gill, L., Hartmann, A., Johannet, A., Jourde, H., Kavousi, A., Liesch, T., Liu, Y., Lüthi, M., Malard, A., Mazzilli, N., Pardo-Igúzquiza, E., Thiéry, D., Reimann, T., … Wunsch, A. (2021). Karst modelling challenge 1: Results of hydrological modelling. *J. Hydrol.*, *600*, 126508. https://doi.org/10.1016/j.jhydrol.2021.126508 (cit. on p. 86).

Jiang, T., Chen, Y. D., Xu, C.-y., Chen, X., Chen, X., & Singh, V. P. (2007). Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. *J. Hydrol.*, *336*(3-4), 316–333. https://doi.org/10.1016/j.jhydrol.2007.01.010 (cit. on p. 92).

Jones, R. N., Chiew, F. H., Boughton, W. C., & Zhang, L. (2006). Estimating the sensitivity of mean annual runoff to climate change using selected hydrological models. *Adv. Water Resour.*, *29*(10), 1419–1429. https://doi.org/10.1016/j.advwatres.2005.11.001 (cit. on p. 92).

Jourde, H., Lafare, A., Mazzilli, N., Belaud, G., Neppel, L., Dörfliger, N., & Cernesson, F. (2014). Flash flood mitigation as a positive consequence of anthropogenic forcing on the groundwater resource in a karst catchment. *Environ. Earth Sci.*, *71*(2), 573–583. https://doi.org/10.1007/s12665-013-2678-3 (cit. on pp. 93, 96).

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, *4*, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968 (cit. on p. 101).

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrol. Sci. J.*, *31*(1), 13–24. https://doi.org/10.1080/02626668609491024 (cit. on p. 89).

Knoben, W. J. M., Freer, J. E., Peel, M. C., Fowler, K. J. A., & Woods, R. A. (2020). A Brief Analysis of Conceptual Model Structure Uncertainty Using 36 Models and 559 Catchments. *Water Resour. Res.*, *56*(9). https://doi.org/10.1029/2019WR025975 (cit. on p. 92).

Labat, D., Argouze, R., Mazzilli, N., Ollivier, C., & Sivelle, V. (2022). Impact of Withdrawals on Karst Watershed Water Supply. *Water*, *14*(9), 1339. https://doi.org/10.3390/w14091339 (cit. on pp. 86, 92, 93, 100).

Lee, A. (2014). *Pyswarm: Particle swarm optimization (PSO) with constraint support, https://github.com/tisimst/pyswarm.* (Cit. on p. 101).

Liu, Y., Wagener, T., & Hartmann, A. (2021). Assessing Streamflow Sensitivity to Precipitation Variability in Karst-Influenced Catchments With Unclosed Water Balances. *Water Res.*, *57*(1). https://doi.org/10.1029/2020WR028598 (cit. on p. 87).

Lorette, G., Sebilo, M., Buquet, D., Lastennet, R., Denis, A., Peyraube, N., Charriere, V., & Studer, J.-C. (2022). Tracing sources and fate of nitrate in multilayered karstic hydrogeological catchments using natural stable isotopic composition (15N-NO3 and 18O-NO3). Application to the Toulon karst system (Dordogne, France). *J.*

Hydrol., *610*, 127972. https://doi.org/10.1016/j.jhydrol.2022.127972 (cit. on p. 85).

Lukač Reberski, J., Terzić, J., Maurice, L. D., & Lapworth, D. J. (2022). Emerging organic contaminants in karst groundwater: A global level assessment. *J. Hydrol.*, *604*, 127242. https://doi.org/10.1016/j.jhydrol.2021.127242 (cit. on p. 85).

Mazzilli, N., Guinot, V., & Jourde, H. (2012). Sensitivity analysis of conceptual model calibration to initialisation bias. Application to karst spring discharge models. *Adv. Water Resour.*, *42*, 1–16. https://doi.org/10.1016/j.advwatres.2012.03.020 (cit. on p. 89).

Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Soglio, L. D., & Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.*, *122*, 103927. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on pp. 85, 86, 89, 92).

Mazzilli, N., Jourde, H., Guinot, V., Bailly-Comte, D., & Fleure, P. (2011). *Hydrological modelling of a karst aquifer under active groundwater management using a parsimonious conceptual model* (Oral presentation). (Cit. on pp. 86, 97).

Mazzilli, N., Jourde, H., Jacob, T., Guinot, V., Le Moigne, N., Boucher, M., Chalikakis, K., Guyard, H., & Legtchenko, A. (2013). On the inclusion of ground-based gravity measurements to the calibration process of a global rainfall-discharge reservoir model: Case of the Durzon karst system (Larzac, southern France). *Environ. Earth Sci.*, *68*(6), 1631–1646. https://doi.org/10.1007/s12665-012-1856-z (cit. on p. 86).

Mazzilli, N., Sivelle, V., Cinkus, G., Jourde, H., & Bertin, D. (2023). *KarstMod User Guide - version 3.0*. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on pp. 89, 92).

McMillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2011). Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. *J. Hydrol.*, *400*(1-2), 83–94. https://doi.org/10.1016/j.jhydrol.2011.01.026 (cit. on p. 91).

Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2021). Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water*, *13*(1), 28. https://doi.org/10.3390/w13010028 (cit. on p. 91).

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, *10*(3), 282–290 (cit. on pp. 89, 91).

Nerantzaki, S. D., & Nikolaidis, N. P. (2020). The response of three Mediterranean karst springs to drought and the impact of climate change. *J. Hydrol.*, *591*, 125296. https://doi.org/10.1016/j.jhydrol.2020.125296 (cit. on pp. 86, 97).

Ollivier, C., Mazzilli, N., Olioso, A., Chalikakis, K., Carrière, S. D., Danquigny, C., & Emblanch, C. (2020). Karst recharge-discharge semi distributed model to assess spatial variability of flows. *Sci. Total Environ.*, *703*, 134368. https://doi.org/10.1016/j.scitotenv.2019.134368 (cit. on p. 86).

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J. Hydrol.*, *303*(1), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026 (cit. on p. 88).

Oudin, L., Perrin, C., Mathevet, T., Andréassian, V., & Michel, C. (2006). Impact of biased and randomly corrupted inputs on the efficiency and the parameters of

watershed models. *J. Hydrol.*, *320*(1-2), 62–83. https://doi.org/10.1016/j.jhydrol.2005.07.016 (cit. on p. 87).

Palmer, A. N. (1991). Origin and morphology of limestone caves. *Geol. Soc. Am. Bull.*, *103*(1), 1–21. https://doi.org/10.1130/0016-7606(1991)103<0001:OAMOLC>2.3.CO;2 (cit. on p. 85).

Pandi, D., Kothandaraman, S., & Kuppusamy, M. (2021). Hydrological models: A review. *Int. J. Hydrol. Sci. Technol.*, *12*(3), 223. https://doi.org/10.1504/IJHST.2021.117540 (cit. on p. 92).

Pechlivanidis, I., Jackson, B., McMillan, H., & Gupta, H. V. (2013). Using an informational entropy-based metric as a diagnostic of flow duration to drive model parameter identification. *Glob. Nest J.*, *14*(3), 325–334. https://doi.org/10.30955/gnj.000879 (cit. on p. 91).

Perrin, C., Michel, C., & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.*, *242*(3-4), 275–301. https://doi.org/10.1016/S0022-1694(00)00393-0 (cit. on p. 89).

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environ. Model. Softw.*, *70*, 80–85. https://doi.org/10.1016/j.envsoft.2015.04.009 (cit. on p. 101).

Pianosi, F., Sarrazin, F., & Wagener, T. (2020). How successfully is open-source research software adopted? Results and implications of surveying the users of a sensitivity analysis toolbox. *Environmental Modelling & Software*, *124*, 104579. https://doi.org/10.1016/j.envsoft.2019.104579 (cit. on p. 101).

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.*, *63*(13-14), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002 (cit. on pp. 89, 91, 93).

Poulain, A., Watlet, A., Kaufmann, O., Van Camp, M., Jourde, H., Mazzilli, N., Rochez, G., Deleu, R., Quinif, Y., & Hallet, V. (2018). Assessment of groundwater recharge processes through karst vadose zone by cave percolation monitoring. *Hydrol. Process.*, *32*(13), 2069–2083. https://doi.org/10.1002/hyp.13138 (cit. on p. 86).

Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., & Wagener, T. (2018). V2Karst V1.1: A parsimonious large-scale integrated vegetation–recharge model to simulate the impact of climate and land cover change in karst regions. *Geosci. Model Dev.*, *11*(12), 4933–4964. https://doi.org/10.5194/gmd-11-4933-2018 (cit. on p. 86).

Schilling, O. S., Cook, P. G., & Brunner, P. (2019). Beyond Classical Observations in Hydrogeology: The Advantages of Including Exchange Flux, Temperature, Tracer Concentration, Residence Time, and Soil Moisture Observations in Groundwater Model Calibration. *Rev. Geophys.*, *57*(1), 146–182. https://doi.org/10.1029/2018RG000619 (cit. on p. 86).

Schwemmle, R., Demand, D., & Weiler, M. (2021). Technical note: Diagnostic efficiency – specific evaluation of model performance. *Hydrol. Earth Syst. Sci.*, *25*(4), 2187–2198. https://doi.org/10.5194/hess-25-2187-2021 (cit. on pp. 85, 91, 96).

Shmueli, G. (2010). To Explain or to Predict? *Statist. Sci.*, *25*(3). https://doi.org/10.1214/10-STS330 (cit. on p. 90).

Sivelle, V., & Jourde, H. (2021). A methodology for the assessment of groundwater resource variability in karst catchments with sparse temporal measurements. *Hydrogeol. J.*, *29*(1), 137–157. https://doi.org/10.1007/s10040-020-02239-2 (cit. on p. 86).

Sivelle, V., Jourde, H., Bittner, D., Richieri, B., Labat, D., Hartmann, A., & Chiogna, G. (2022a). Considering land cover and land use (LCLU) in lumped parameter modeling in forest dominated karst catchments. *J. Hydrol.*, *612*, 128264. https://doi.org/10.1016/j.jhydrol.2022.128264 (cit. on pp. 86, 101).

Sivelle, V., Cinkus, G., Mazzilli, N., Labat, D., Arfib, B., Massei, N., Cousquer, Y., Bertin, D., & Jourde, H. (2023). Improvement of the KarstMod modeling platform for a better assessment of karst groundwater resources. *Hydrol. Earth Syst. Sci.*, 1–26. https://doi.org/10.5194/hess-2023-17 (cit. on p. 83).

Sivelle, V., Jourde, H., Bittner, D., Mazzilli, N., & Tramblay, Y. (2021). Assessment of the relative impacts of climate changes and anthropogenic forcing on spring discharge of a Mediterranean karst system. *J. Hydrol.*, *598*, 126396. https://doi.org/10.1016/j.jhydrol.2021.126396 (cit. on pp. 86, 92, 93, 97).

Sivelle, V., Labat, D., Mazzilli, N., Massei, N., & Jourde, H. (2019). Dynamics of the Flow Exchanges between Matrix and Conduits in Karstified Watersheds at Multiple Temporal Scales. *Water*, *11*(3), 569. https://doi.org/10.3390/w11030569 (cit. on pp. 86, 87, 100).

Sivelle, V., Pérotin, L., Ladouche, B., De Montety, V., Bailly-Comte, V., Champollion, C., & Jourde, H. (2022b). A lumped parameter model to evaluate the relevance of excess air as a tracer of exchanged flows between transmissive and capacitive compartments of karst systems. *Front. Water*, *4*, 930115. https://doi.org/10.3389/frwa.2022.930115 (cit. on pp. 86, 101).

Smiatek, G., Kaspar, S., & Kunstmann, H. (2013). Hydrological Climate Change Impact Analysis for the Figeh Spring near Damascus, Syria. *J. Hydrometeorol.*, *14*(2), 577–593. https://doi.org/10.1175/JHM-D-12-065.1 (cit. on p. 97).

Sobol, I. (1976). Uniformly distributed sequences with an additional uniform property. *USSR Comput. Math. & Math. Phys.*, *16*(5), 236–242. https://doi.org/10.1016/0041-5553(76)90154-3 (cit. on p. 90).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on p. 85).

Wada, Y., Flörke, M., Hanasaki, N., Eisner, S., Fischer, G., Tramberend, S., Satoh, Y., Van Vliet, M. T. H., Yillia, P., Ringler, C., Burek, P., & Wiberg, D. (2016). Modeling global water use for the 21st century: The Water Futures and Solutions (WFaS) initiative and its approaches. *Geosci. Model Dev.*, *9*(1), 175–222. https://doi.org/10.5194/gmd-9-175-2016 (cit. on p. 85).

Westerberg, I. K., Sikorska-Senoner, A. E., Viviroli, D., Vis, M., & Seibert, J. (2022). Hydrological model calibration with uncertain discharge data. *Hydrol. Sci. J.*, *67*(16), 2441–2456. https://doi.org/10.1080/02626667.2020.1735638 (cit. on p. 91).

Zhou, S., Wang, Y., Li, Z., Chang, J., & Guo, A. (2021). Quantifying the Uncertainty Interaction Between the Model Input and Structure on Hydrological Processes. *Water Resour. Manage.*, *35*(12), 3915–3935. https://doi.org/10.1007/s11269-021-02883-7 (cit. on p. 92).

Chapter 5

Critical evaluation of performance criteria

Performance criteria play a key role in the calibration and evaluation of hydrological models and have been extensively developed and studied, but some of the most used criteria still have unknown pitfalls. This study set out to examine counterbalancing errors, which are inherent to the Kling-Gupta Efficiency (KGE) and its variants. A total of nine performance criteria – including the KGE and its variants, as well as the Nash-Sutcliffe Efficiency (NSE) and the modified index of agreement (d_1) – were analysed using synthetic time series and a real case study. Results showed that, assessing a simulation, the score of the KGE and some of its variants can be increased by concurrent over- and underestimation of discharge. These counterbalancing errors may favour bias and variability parameters, therefore preserving an overall high score of the performance criteria. As bias and variability parameters generally account for 2/3 of the weight in the equation of performance criteria such as the KGE, this can lead to an overall higher criterion score without being associated with an increase in model relevance. We recommend using (i) performance criteria that are not or less prone to counterbalancing errors (d_1, modified KGE, non-parametric KGE, Diagnostic Efficiency), and/or (ii) scaling factors in the equation to reduce the influence of relative parameters.

This work resulted in a publication in Hydrology and Earth System Sciences (Cinkus et al., 2023a) and a presentation during EGU22 (Cinkus et al., 2022).

Article:

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., and Goldscheider, N.: When best is the enemy of good – critical evaluation of performance criteria in hydrological models, Hydrol. Earth Syst. Sci., 27, 2397–2411, https://doi.org/10.5194/hess-27-2397-2023, 2023.

Contents

5.1 Introduction

Hydrological models are fundamental to solve problems related to water resources. They help characterising hydrosystems (Hartmann et al., 2014), predicting floods (Jain et al., 2018; Kauffeldt et al., 2016) and managing water resources (Muleta and Nicklow, 2005). A lot of research efforts are thus dedicated to improve the reliability, the robustness and the relevance of such models. Improvements can be made by working on (i) input data, (ii) model parameters and structure, (iii) uncertainty quantification, (iv) model calibration (Beven, 2019), and also (v) appropriate benchmarks for assessing model performance (Seibert et al., 2018). In this study, we focus on the proper use of performance criteria for calibrating and evaluating hydrological models – an important part that can easily be overlooked (Jackson et al., 2019).

A performance criterion aims to evaluate the goodness-of-fit of a model to an observed data. It is generally expressed as a score, for which the best value corresponds to a perfect fit between predictions and observations. In hydrology, the Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970) is still one of the most commonly used criteria (Kling et al., 2012), although the past decade has seen a gain in popularity of alternatives (Clark et al., 2021), e.g. the Kling-Gupta Efficiency (KGE) (Gupta et al., 2009). Many authors have pointed out the inherent limitations of using performance criteria, especially the fact that a single score metric cannot reflect all relevant hydrological aspects of a model (Gupta et al., 2009). The use of a multi-criteria framework is thus often emphasised to quantify different aspects of a model (Althoff and Rodrigues, 2021; Clark et al., 2021; Gupta et al., 1998; Jackson et al., 2019; Knoben et al., 2019; Krause et al., 2005; Legates and McCabe Jr., 1999; Moriasi et al., 2015; Ritter and Muñoz-Carpena, 2013; van Werkhoven et al., 2009), alongside a scientific evaluation of the results (Biondi et al., 2012). Althoff and Rodrigues (2021), Clark et al. (2021), and Knoben et al. (2019) pointed out that modellers should carefully think about which aspects they consider the most important in their hydrological model and how to evaluate them. In relation to the assessment of model performance, Seibert et al. (2018) argued that the current benchmarks poorly reflect what could and should be expected of a model. They suggested to define lower and upper benchmarks based on the performance of a simple bucket-type model with few parameters, using the same data set.

Performance criteria also have shortcomings at a distinctive level. A number of studies have identified several limitations of the NSE: (i) the contribution of the normalised bias depends on the discharge variability of the basin, (ii) discharge variability is inevitably underestimated because the NSE is maximised when the variability equals the correlation coefficient, which is always smaller than unity, and (iii) mean flow is not a meaningful benchmark for highly variable discharges (Gupta et al., 2009; Willmott et al., 2012). The KGE aims to address these limitations but also has its own issues (Gupta et al., 2009). Santos et al. (2018) identified pitfalls when using the KGE with a prior logarithmic transformation of the discharge. Knoben et al. (2019) warned against directly comparing NSE and KGE scores as the KGE has no inherent benchmark. Clark et al. (2021) and Ritter and Muñoz-Carpena (2013) showed that NSE and KGE scores can be strongly influenced by few data points, resulting in substantial uncertainties on the predictions.

What is not fully addressed yet is the trade-off between individual components (Wöhling et al., 2013) and especially the impact of counterbalancing errors induced by bias and variability parameters, which are integrated in many performance criteria. While accurate bias and variability are desired aspects of hydrological models, sometimes good

evaluations may accidentally result from negative and positive values cancelling each other (Jackson et al., 2019; Massmann et al., 2018). This can be particularly detrimental to model calibration and evaluation, as it generates an increase in criterion score without necessarily being associated with a better model relevance. Some performance criteria naturally address this problem by using absolute or squared error values, but other criteria such as the KGE and its variants do not, as they use relative errors. The aim of this study is to assess the extent to which criteria scores can be trusted for calibrating and evaluating hydrological models when predictions have concurrent over- and underestimated values. The influence of counterbalancing errors is evaluated on nine performance criteria including the NSE and KGE. This selection is far being from exhaustive but includes widely used and recently proposed KGE variants, as well as more traditional criteria such as the NSE or the modified index of agreement (d_1) for comparison purpose. We first use synthetic time series to highlight the counterbalancing errors mechanism. Second, we show how counterbalancing errors can impair the interpretation of hydrological models in a real case study. Finally, we provide some recommendations about the use of scaling factors and the choice of appropriate performance criteria to nullify or reduce the influence of counterbalancing errors.

5.2 Performance criteria

5.2.1 Parameters description

All the performance criteria considered in this study are based on the same or similar statistical indicators, which are first described to avoid repetition.

We use $x_o(t)$ and $x_s(t)$ to refer to observed and simulated values of calibration variable x at a specific time step t. r and r_s correspond to the Pearson and the Spearman rank correlation coefficients (Freedman et al., 2007), respectively.

β is the ratio between the mean of simulated values μ_s and the mean of observed values μ_o:

$$\beta = \frac{\mu_s}{\mu_o} \tag{5.1}$$

β_n corresponds to the bias (mean error) normalised by the standard deviation of observed values σ_o:

$$\beta_n = \frac{\mu_s - \mu_o}{\sigma_o} \tag{5.2}$$

α is the ratio between the standard deviation of simulated values σ_s and the standard deviation of observed values σ_o:

$$\alpha = \frac{\sigma_s}{\sigma_o} \tag{5.3}$$

γ is the ratio between the coefficient of variation of simulated values ($CV_s = \sigma_s/\mu_s$) and the coefficient of variation of observed values ($CV_o = \sigma_o/\mu_o$):

$$\gamma = \frac{CV_s}{CV_o} \tag{5.4}$$

$\overline{B_{rel}}$ and $|B_{area}|$ (Schwemmle et al., 2021) are based on the Flow Duration Curve (FDC). $B_{rel}(i)$ is defined as the relative bias of the simulated and observed flow duration curves at the exceedance probability i:

$$B_{rel} = \frac{x_s(i) - x_o(i)}{x_o(i)} \tag{5.5}$$

where $x_s(i)$ and $x_o(i)$ correspond to the simulated and observed values of calibration variable at exceedance probability i. $\overline{B_{rel}}$ is the mean of $B_{rel}(i)$ when looking at n observations:

$$\overline{B_{rel}} = \frac{1}{n}\sum_{i=0}^{i=1} B_{rel}(i) \tag{5.6}$$

$|B_{area}|$ is calculated as follows:

$$|B_{area}| = \int_0^1 |B_{res}(i)|di \tag{5.7}$$

with B_{res} the residual bias:

$$B_{res} = B_{rel}(i) - \overline{B_{rel}} \tag{5.8}$$

α_{NP} (Pool et al., 2018) is also based on the FDC:

$$\alpha_{NP} = 1 - \frac{1}{2}\sum_{k=1}^{n} \left|\frac{x_s(I(k))}{n\mu_s} - \frac{x_o(J(k))}{n\mu_o}\right| \tag{5.9}$$

where $I(k)$ and $J(k)$ stand for the time steps of the k^{th} largest discharge for the simulated and observed time series, respectively.

As β, β_n and $\overline{B_{rel}}$ all represent the bias, they are therefore designed as "bias parameters" in this study.

5.2.2 Score calculation

A total of nine performance criteria are analysed in this study: the NSE, KGE, 2012-version of the KGE or modified KGE (KGE'), 2021-version of the KGE (KGE"), non-parametric KGE (KGE_{NP}), Diagnostic Efficiency (DE), Liu-Mean Efficiency (LME), Least-squares Combined Efficiency (LCE) and d_1. The value considered as the best score is equal to one for all criteria, except for the DE, for which it is equal to zero.

The NSE (Nash and Sutcliffe, 1970) is a normalised variant of the Mean Squared Error (MSE) and compares a prediction to the observed mean of the target variable:

$$NSE = 1 - \frac{\sum (x_s(t) - x_o(t))^2}{\sum (x_o(t) - \mu_o)^2} \tag{5.10}$$

Gupta et al. (2009) algebraically decomposed the NSE into correlation, variability, and bias components:

$$NSE = 2\alpha r - \alpha^2 - \beta_n^2 \tag{5.11}$$

The Kling-Gupta Efficiency (KGE) was proposed by Gupta et al. (2009) as an alternative to the NSE. The optimal KGE corresponds to the closest point of the three-dimensional Pareto front – of α, β and r – to the ideal value of [1; 1; 1]:

$$KGE = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (r - 1)^2} \tag{5.12}$$

A modified Kling-Gupta Efficiency was proposed by Kling et al. (2012). The coefficient of variation is used instead of the standard deviation to ensure that bias and variability are not cross-correlated:

$$KGE' = 1 - \sqrt{(\gamma - 1)^2 + (\beta - 1)^2 + (r - 1)^2} \tag{5.13}$$

Tang et al. (2021) proposed another variant (KGE") by using the normalised bias instead of β to ensure that the score is not overly sensitive to mean values – μ_o or μ_s – close to zero (Santos et al., 2018; Tang et al., 2021):

$$KGE'' = 1 - \sqrt{(\alpha - 1)^2 + \beta_n^2 + (r - 1)^2} \tag{5.14}$$

Pool et al. (2018) cautioned against the implicit assumptions of the KGE – data linearity, data normality and absence of outliers – and proposed a non-parametric alternative (KGE_{NP}) for limiting their impact. The non-parametric form of the variability is calculated using the FDC and the Spearman rank correlation coefficient is used instead of the Pearson correlation coefficient:

$$KGE_{NP} = 1 - \sqrt{(\alpha_{NP} - 1)^2 + (\beta - 1)^2 + (r_s - 1)^2} \tag{5.15}$$

In a similar way, Schwemmle et al. (2021) used FDC-based parameters to account for variability and bias in another KGE variant: the Diagnostic Efficiency. This criterion is based on constant, dynamic and timing errors and aims to provide a stronger link to hydrological processes (Schwemmle et al., 2021):

$$DE = \sqrt{\overline{B_{rel}}^2 + |B_{area}|^2 + (r - 1)^2} \tag{5.16}$$

In this study, we used a Normalised Diagnostic Efficiency (DE') so that the best error score equals to one for facilitating the comparison with other performance criteria:

$$DE' = 1 - \sqrt{\overline{B_{rel}}^2 + |B_{area}|^2 + (r - 1)^2} \tag{5.17}$$

Liu (2020) proposed another alternative, the Liu-Mean Efficiency, to improve the simulation of extreme events. The LME thus aims to address the underestimation of variability of the KGE, which is still a concern despite being not as severe as with the NSE (Gupta et al., 2009; Mizukami et al., 2019):

$$LME = 1 - \sqrt{(r\alpha - 1)^2 + (\beta - 1)^2} \tag{5.18}$$

Lee and Choi (2022) proposed the Least-squares Combined Efficiency to address the shortcomings of the LME identified by Choi (2022): (i) an infinite number of solutions for the maximum score, and (ii) a inclination to overestimate high flows and underestimate low flows. The LCE is based on the least-squares statistics combined from both-way regression lines $r\alpha$ and r/α:

$$LCE = 1 - \sqrt{(r\alpha - 1)^2 + (r/\alpha - 1)^2 + (\beta - 1)^2} \tag{5.19}$$

Willmott et al. (1985) proposed a modified index of agreement, which aim to address the issues associated with r and the coefficient of determination, as well as the sensitivity of the original index of agreement to outliers (Legates and McCabe Jr., 1999):

$$d_1 = 1 - \frac{\sum |x_s(t) - x_o(t)|}{\sum \left(|x_s(t) - \mu_o| + |x_o(t) - \mu_o|\right)} \tag{5.20}$$

5.3 Synthetic time series

5.3.1 Generating synthetic time series with homothetic transformations

A simulation performance can be assessed in terms of bias, variability and timing errors (Gupta et al., 2009). Bias and variability errors correspond to a difference in volume and amplitude of discharges. Timing errors correspond to a shift in time. We created a synthetic hydrograph corresponding to one flood event as the reference (observed) time series. We also generated synthetic transformations – of the reference time series – with different errors on bias and variability corresponding to time series simulated by a model. We did not consider any timing errors as our aim is to assess counterbalancing errors induced by bias and variability parameters. Synthetic transformations were generated by multiplying the reference time series by a coefficient ω:

$$Q_s(t) = Q_o(t) * \omega \tag{5.21}$$

where $Q_s(t)$ stands for the transformed discharge at the time t, $Q_o(t)$ the reference discharge at the time t and ω a coefficient. ω values were sampled uniformly on the log-transformed interval [-0.36, 0.36] at a defined step of 0.002 to ensure a fair distribution between underestimated and overestimated transformations. The exponentiation in base 10 of the sampled values results in 361 ω values evenly distributed around the $\omega = 1$ homothety, which corresponds to the reference time series (i.e. absence of transformation). We defined ω bounds such that the transformed peak discharge roughly ranges from half ($\omega \approx 0.437 \approx 10^{-0.36}$) to twice ($\omega \approx 2.291 \approx 10^{0.36}$) compared to the reference time series. Note that (i) the data linearity between simulated and observed values is verified, and (ii) ω homotheties still induce small timing errors – which were considered negligible – because the correlation coefficients (r and r_s) also slightly account for the shape of the transformation.

To study counterbalancing errors induced by bias and variability parameters, we generated time series that consist of two successive flood events and considered all possible combinations of the 361 transformations for the simulated time series (Figure 5.1). This results in a total of $361^2 = 130321$ transformations with two flood events, including (i) a "perfect" transformation with $\omega = 1$ for both flood events, (ii) "Bad-Good" (BG) or "Good-Bad" (GB) transformations when $\omega = 1$ for only one out of the two flood events, and (iii) "Bad-Bad" (BB) transformations when $\omega \neq 1$ for both flood events. The performance of the transformations – with regards to the reference time series – were evaluated using the nine performance criteria presented in section 5.2.

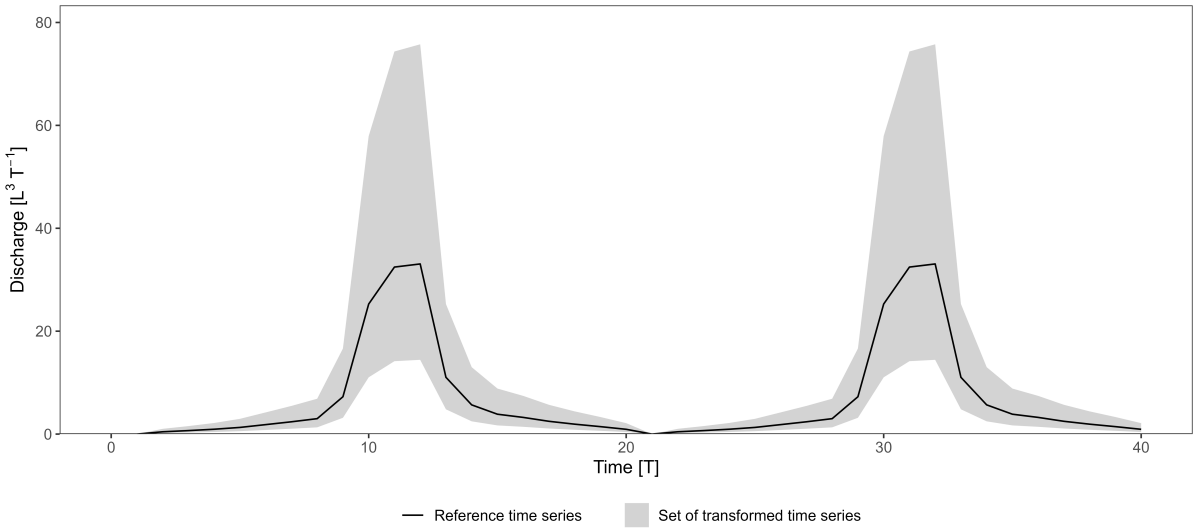


Figure 5.1: Synthetic hydrograph corresponding to two flood events.

5.3.2 Identifying counterbalancing errors on a straightforward example

Figure 5.2 presents two hydrographs extracted from the set of transformations: (i) a BB model with the combination $[\omega_1 = 0.75; \omega_2 = 1.2]$, and (ii) a BG model with the combination $[\omega_1 = 0.75; \omega_2 = 1]$. The BG model stands as a better model because it perfectly reproduces the second flood event and is identical to the BB model on the first flood ($\omega_1 = 0.75$). Nevertheless, the KGE and its variants – KGE', KGE", KGE_{NP}, DE', LME and LCE – all favour the BB model, whereas only the NSE and d_1 evaluate the BG model as better (Figure 5.3a). Further results for common and recently developed performance criteria are presented in Appendix 5.A.

The investigation of the components of the criteria (Figure 5.3b) reveals how a seemingly better model (i.e. the BG model) can have a lower score than expected. Bias parameters are systematically better for the BB model, with 0.98 over 0.88 for β, -0.02 over -0.08 for β_n and -0.04 over -0.12 for $\overline{Brel}$. Timing parameters are systematically better for the BG model, with 0.99 over 0.96 for r and 0.99 over 0.98 for r_s. Variability parameters are mixed: (i) α favours the BB model with 1.01 over 0.89, (ii) γ favours the BG model with 1.01 over 1.04, (iii) α_{NP} slightly favours the BG model with 0.94 over 0.93, and (iv) $|B_{area}|$ is equal for both models. $r\alpha$ and r/α parameters are better for the BB model. $2\alpha r$ is better for the BG model.

β, β_n, $\overline{Brel}$, α, $r\alpha$ and r/α parameters all provide a better evaluation of bias and variability for the BB model. Concurrent over- and underestimation of discharges over the time series result in a good water balance: close to 1 for β and $\overline{B_{rel}}$ and 0 for β_n. Depending on the criterion, the variability parameter can also affect the score in a similar counter-intuitive manner. α is heavily impacted by the counterbalance, whereas it seems mitigated for γ, α_{NP} and $|B_{area}|$. The timing parameters (r and r_s) have an expected score that favour the BG model. However, the score difference on timing errors between BB and BG models is very small (0.03 at best for r). The impact on the overall score is thus minimised compared to the one induced by bias and variability parameters, which can be cumulated (e.g. both β and α counterbalancing errors in the KGE) or have a larger difference – up to 0.12 for α. Counterbalancing errors can thus result in better

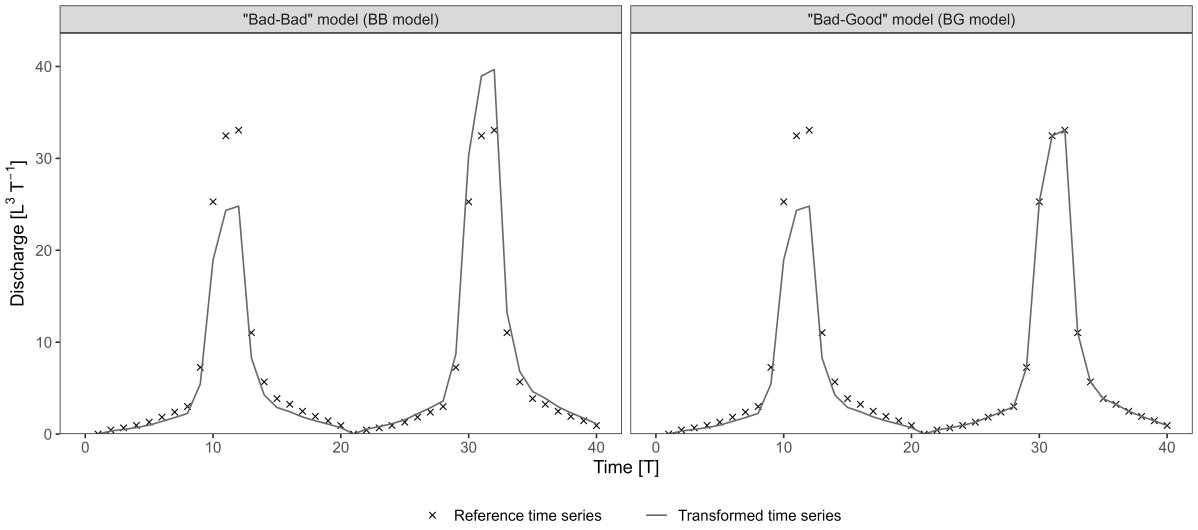


Figure 5.2: Synthetic examples extracted from the set of transformations. The first and second flood events of the "Bad-Bad" and "Bad-Good" transformations were shifted with $[\omega_1 = 0.75; \omega_2 = 1.2]$ and $[\omega_1 = 0.75; \omega_2 = 1]$ combinations, respectively.

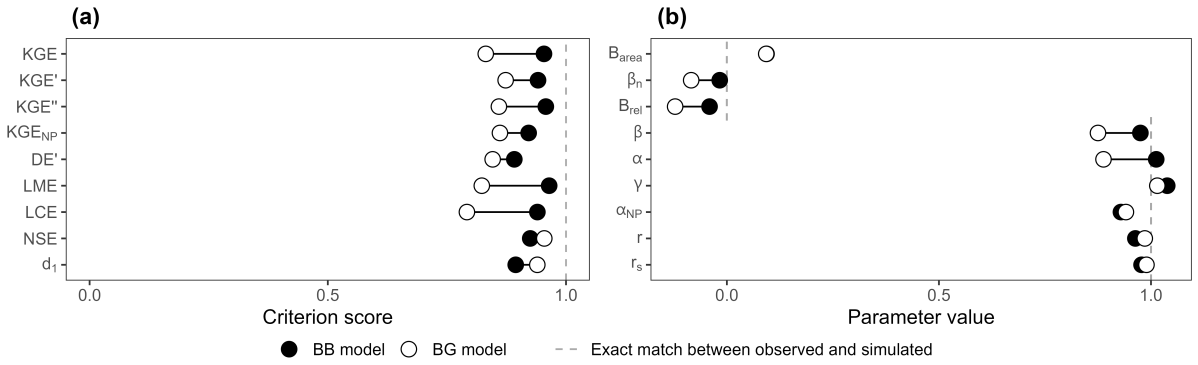


Figure 5.3: (a) Score of the BB and BG transformations according to the different performance criteria. (b) Values of the parameters used in the calculation of the performance criteria.

values for bias and variability, which increase the overall score. In this case, the highest score may not be the most appropriate indicator of model relevance.

The largest differences in score appear for the LME and LCE criteria as all their parameters are affected by counterbalancing errors (β, $r\alpha$ and r/α). The KGE and KGE" also show significant differences as they accumulate the counterbalancing errors of α and β. The KGE' demonstrates a smaller difference than the KGE due to the use of γ. Both FDC-based criteria $\mathrm{KGE_{NP}}$ and DE' show the smallest differences due to α_{NP} and $|B_{area}|$, which have a nearly equal value for both BB and BG models. The NSE has a slightly better score on the BG model, while the difference is more pronounced on $\mathrm{d_1}$.

This example demonstrates how relative error metrics can cancel out each other and affect the design and the evaluation of hydrological models. The counterbalancing errors especially affect bias parameters (β, β_n and $\overline{B_{rel}}$) but also the variability parameter α.

5.3.3 Exploring counterbalancing errors with synthetic transformations

Figure 5.4 shows the score distribution of the synthetic set of hydrographs presented in section 5.3.1. For each value of ω_1, the minimum and maximum criteria scores of the transformations resulting from all combinations with ω_2 provide the dashed envelope of the score distribution, with the maximum transformation score at the top (1 corresponding to a perfect model), and the worst at the bottom. The transformations corresponding to the BG models (with $\omega_2 = 1$) are represented by the black line. All transformations included in the dashed envelope can be identified as "Bad-Bad" models, except when $\omega_1 = 1$ or $\omega_2 = 1$ (black line).

It is obvious that the KGE and its variants – KGE', KGE", $\mathrm{KGE_{NP}}$, DE', LME and LCE – always evaluate one or several BB models as better than the BG model for a same ω_1 value, except for $\omega_1 = 1$. On the other hand, the NSE and $\mathrm{d_1}$ correctly identify the BG model as the best transformation for all combinations of $[\omega_1;\omega_2]$, i.e. the black line is always above the dashed envelope. The envelope of the KGE, KGE' and KGE" criteria are similar, but they do not display the same difference between the best scores and the scores of the BG models. These differences are smaller for the latter two because the KGE' is based on γ instead of α, and the KGE" is based on β_n instead of β, for which it is demonstrated in section 5.3.2 that they both soften counterbalancing errors. The envelope of the LCE criterion looks like that of the KGE. However, the difference between the best scores and the scores of the BG models is much higher. This is likely due to the nature of the equation consisting of 3 parameters affected by counterbalancing errors (β, $r\alpha$ and r/α). The LME criterion has a very distinctive envelope, for which the maximum score of 1 is reached for a lot of BB models, even when both ω_1 and ω_2 are different from 1. This can be explained by the interaction between r and α that leads to an infinite number of solutions (Choi, 2022). The $\mathrm{KGE_{NP}}$ and DE' (FDC-based criteria) both show similar envelopes with a break point near the maximum transformation score in both directions around $\omega_1 = 1$. This is especially pronounced for the DE', for which the BG model is nearly the best model between $\omega_1 = 0.83$ and $\omega_1 = 1.17$. These results show that counterbalancing errors can happen on a large range of parameters, and when using the KGE or its variants, there is a possibility for the more meaningful model (i.e. BG model) to have a lower score than a "compensated" or "Bad-Bad" model.

Figure 5.5 shows the value of ω_2 corresponding to the best evaluation for a given ω_1, by performance criteria. As identified above, the NSE and $\mathrm{d_1}$ both evaluate the BG

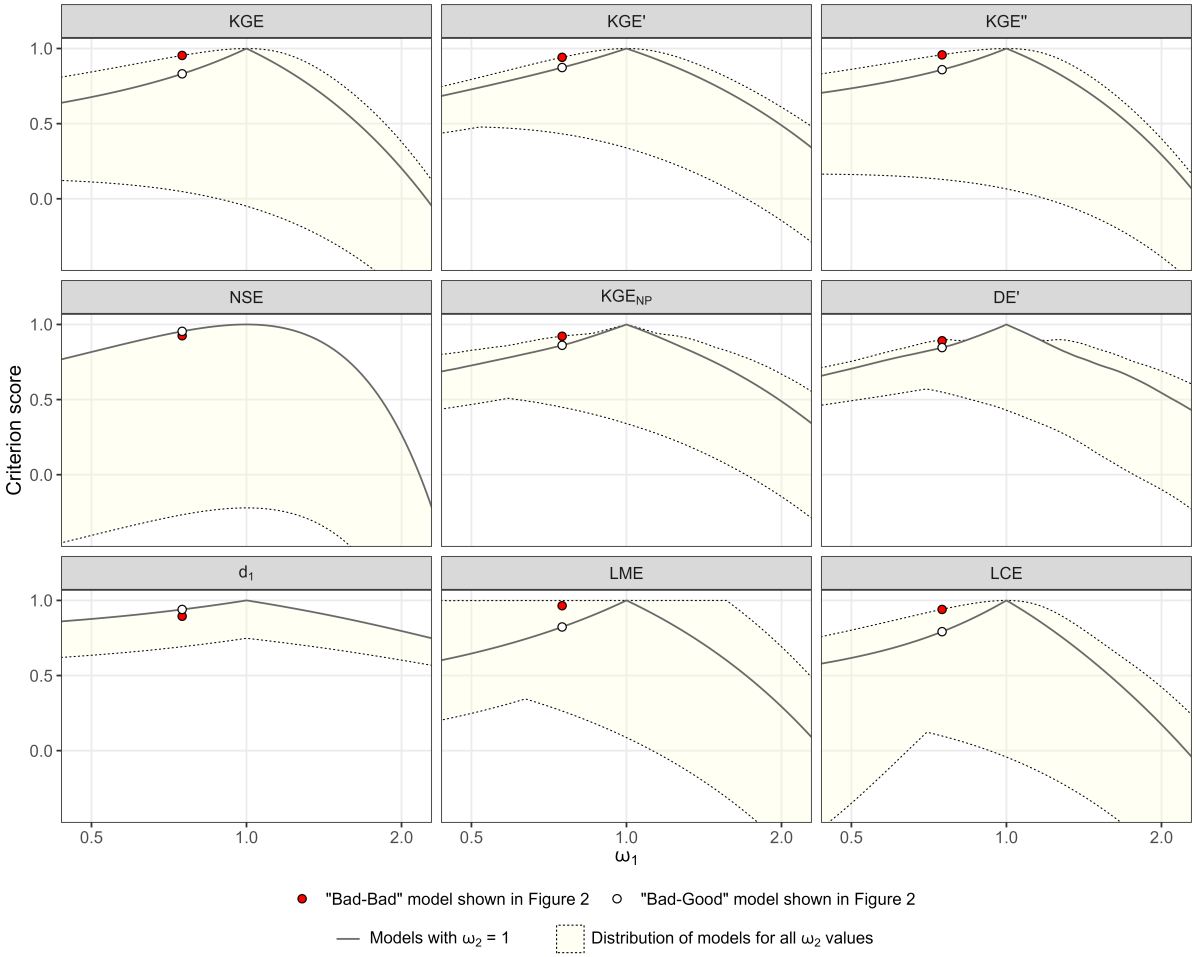


Figure 5.4: Score of each transformation for all $[\omega_1; \omega_2]$ combinations by performance criteria.

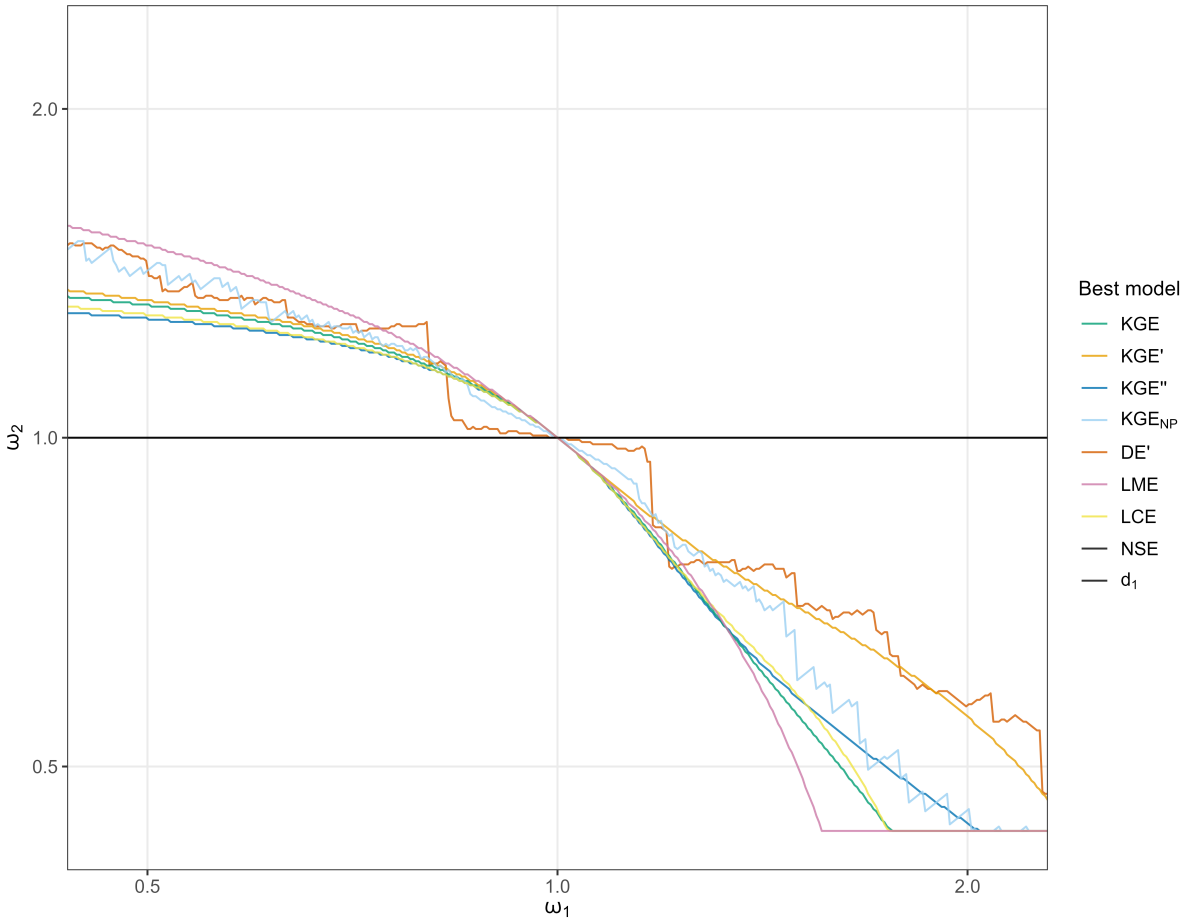


Figure 5.5: Graph of each $[\omega_1;\omega_2]$ combination identified as the best transformation by each performance criteria. The NSE and the d_1 black lines coincide at $\omega_2 = 1$.

models as the best transformations (NSE and d_1 black lines coincide at $\omega_2 = 1$, Figure 5.5). Counterbalancing errors are apparent for the KGE and its variants. For $\omega_1 \neq 1$, best transformations are always BB models and follow two conditions: (i) if $\omega_1 < 1$ then $\omega_2 > 1$, and (ii) if $\omega_1 > 1$ then $\omega_2 < 1$. This means that, in this case, such performance criteria will always be flawed towards concurrent under- and overestimation of discharges in a transformation.

5.4 Real case study

To highlight how counterbalancing errors can affect the assessment of hydrological models on a real case study, we used two different modelling approaches: artificial neural networks (ANN) and bucket-type models. The simulations of karst spring discharges of both models were evaluated on the same 1-year validation period. To clearly highlight the problem, we deliberately chose a bucket-type simulation that is noticeably affected by counterbalancing errors – yet still realistic. Further information on the modelling approaches, the input data, the calibration strategy and the simulation procedure can be found in Cinkus et al. (2023b).

5.4.1 Study site

The Unica springs are the outlet of a complex karstic system influenced by a network of poljes. The recharge area is about 820 km^2 and is located in a moderate continental climate with a strong snow influence. Recharge comes from both (i) allogenic infiltration from two sub-basins drained by sinking rivers, and (ii) autogenic infiltration through a highly karstified limestone plateau (Gabrovšek et al., 2010; Kovačič, 2010; Petric, 2010). The network of connected poljes constitutes a common hydrological entity that induces a high hydrological variability in the system, and long and delayed high discharges at the Unica springs (Mayaud et al., 2019). The limestone massif can reach a height of 1800 m above sea level and has significant groundwater resources (Ravbar et al., 2012). A polje downstream of the springs can flood when the Unica discharge exceeds 60 $m^3 s^{-1}$ for several days. If the flow reaches 80 $m^3 s^{-1}$, the flooding can reach the gauging station and influence its measurement. The flow data are from the gauging station in Unica-Hasberg (ARSO, 2021b). Precipitation, height of snow cover, and height of new snow data are from the meteorological stations in Postojna and Cerknica (ARSO, 2021a). Temperature and relative humidity data are from the Postojna station. Potential evapotranspiration is calculated from the Postojna station data with the Penman-Monteith formula (Allen et al., 1998).

5.4.2 Modelling approaches

The first modelling approach is based on Convolutional Neural Networks (CNN) (LeCun et al., 2015), which is a specific type of ANN that is powerful in processing image-like data but also very useful for processing sequential data. The model consists of a single 1D Convolutional layer with a fixed kernel size of three and an optimised number of filters. This layer was complemented by a Max-Pooling layer a Monte-Carlo dropout layer with 10 % dropout rate and two dense layers. The first dense layer has an optimised number of neurons and the second a single output neuron. We programmed our models in Python 3.8 (van Rossum, 1995), using the following frameworks and libraries: Bayesian

Optimization (Nogueira, 2014), Matplotlib (Hunter, 2007), Numpy (van der Walt et al., 2011), Pandas (McKinney, 2010; Reback et al., 2021), Scikit-Learn (Pedregosa et al., 2018), TensorFlow 2 .7 (Abadi et al., 2016) and its Keras API (Chollet, 2015).

The second modelling approach is a bucket-type model, which is a conceptual representation of a hydrosystem consisting of several buckets that are supposed to be representative of the main processes involved. We used the adjustable modelling platform KarstMod (Mazzilli et al., 2019). The model structure consists of one upper bucket for simulating soil and epikarst processes (including a soil available water capacity), and two lower buckets corresponding to matrix and conduits compartments. A very reactive transfer function from the upper bucket to the spring is used to reproduce very fast flows occurring in the system.

5.4.3 Impact of counterbalancing errors on model evaluation

Figure 5.6a shows the results of the two hydrological models on Unica springs. The models have overall good dynamics and successfully reproduce the observed discharges. Regarding high flow periods, both models show a small timing error, inducing a delay in the simulated peak flood. The first flood event (February 2017) is slightly underestimated by the ANN model and highly overestimated by the bucket-type model. The second flood event (March 2017) is similarly underestimated by both models but the bucket-type model demonstrates a slightly better performance. The third flood event (May 2017) is poorly simulated by the models – both underestimate the flood peak – but the ANN model is more accurate in terms of timing and volume estimate, while the bucket-type model has a better recession coefficient and flow variability. The last flood event (September 2017) comprise a small peak followed by a very high and long-lasting flood. Both models fail to account for the small peak. The following important flood event is highly overestimated by the bucket-type model, while being nicely simulated by the ANN model – despite the small underestimation and timing error. The small flood events are better simulated by the ANN model than the bucket-type model: (i) the ANN model simulates them satisfactorily, except for the second one (mid-April), where the simulated discharges are overestimated; (ii) the bucket-type model does not simulate the first two events at all (mid-January and mid-April) and largely overestimates the last two (early and late June), in addition to timing errors. Both models can be improved during recession and low flow periods. The ANN model is rather close to the observed discharges but seems to be too sensitive to precipitation (continuous oscillations). On the other hand, the bucket-type model shows no oscillations but either overestimates or underestimates the observed discharges. Some events are not well simulated by both models (e.g. the May 2017 flood), which may be due to uncertainties in the input data. Also, the data linearity between simulated and observed values is slightly skewed for both models, which can affect the relevance of r (Barber et al., 2020).

In general, the ANN model can be described as better because it is closer to the observed values in the high and low flow periods. While this statement cannot be supported by performance metrics, we believe that an expert assessment based on intuition and experience is still valuable despite being intrinsically subjective. In this particular case, one can assess the main, distinctive flaws of each model: (i) the ANN model has continuous oscillations – especially on recession and low flow periods – and lacks of accuracy during recession periods; (ii) the bucket-type model highly overestimates several flood events and is inaccurate during a lot of recession and low flow periods. Figure 5.6b also shows

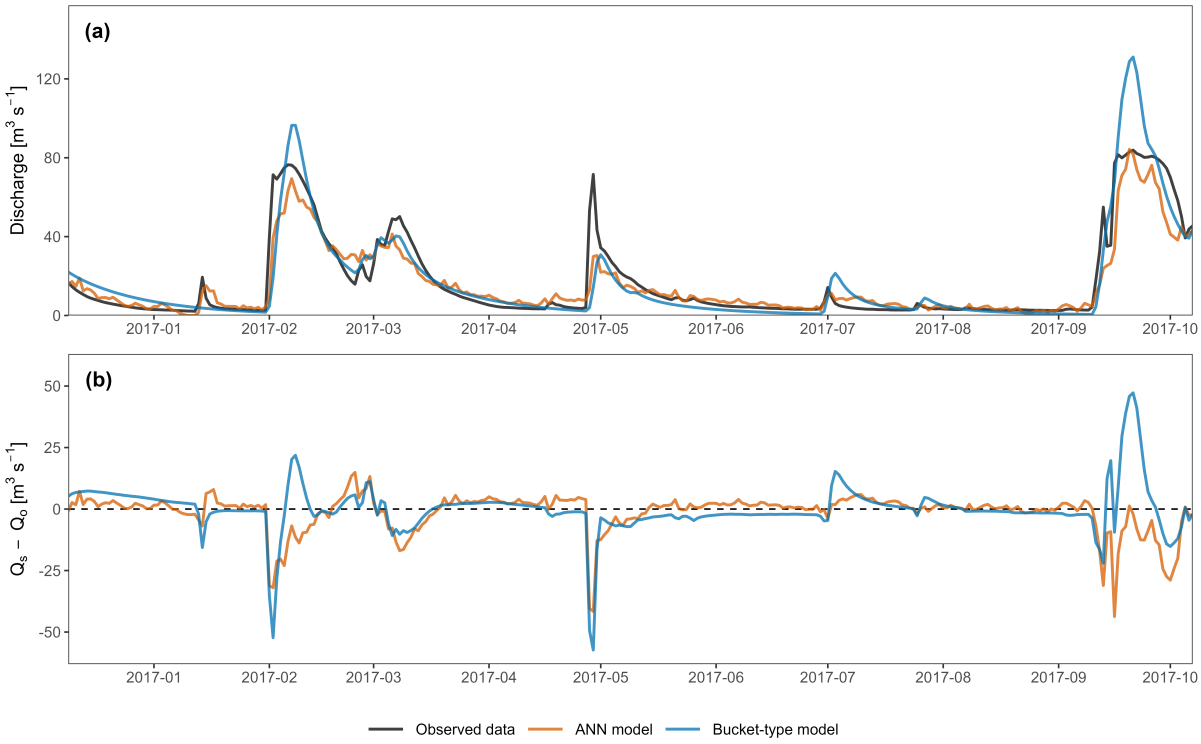


Figure 5.6: (a) Observed and simulated spring discharge time series on the validation period. (b) Relative difference between simulated and observed discharge on the validation period.

that the bucket-type model has an overall higher bias than the ANN model. Hydrological models are generally used for (i) the prediction/forecast of water flood/inrush, (ii) the management of water resources, (iii) the characterisation of hydrosystems, and more recently (iv) the study of the impact of climate change on water resources. Most studies thus put the emphasis on volumes, and also extremes events (i.e. dry and flood periods), which in this case are more satisfactorily reproduced by the ANN model – in terms of volume estimate, timing and variability.

This visual assessment is confirmed only by few performance criteria: the NSE, d_1 and KGE_{NP} (Figure 5.7a). These criteria evaluate the ANN model as better, although the performances of both models are quite close for the d_1. However, the KGE and most of its variants (except the KGE_{NP}) all favour the bucket-type model over the ANN model – sometimes by a large margin. Further results for common and recently developed performance criteria are presented in Appendix 5.A. It is interesting to note how similar these results are to those of the synthetic example (Figure 5.3a, Appendix 5.A). Looking at the values of the equations' parameters (Figure 5.7b), we find that bias parameters are systematically better for the bucket-type model, with 1 over 0.92 for β, 0 over -0.06 for β_n and -0.07 over 0.18 for $\overline{B_{rel}}$. Timing errors are systematically better for the ANN model, with 0.95 over 0.92 for r and 0.94 over 0.83 for r_s. Variability parameters favour the bucket-type model with 1.1 over 0.78 for α, 1.1 over 0.85 for γ, 0.22 over 0.3 for $|B_{area}|$, and a very close better value by 0.005 on the α_{NP} parameter. In summary, all bias and variability parameters have better values for the bucket-type model, while timing and shape parameters are better for the ANN model.

As the KGE and its variants are generally composed of equally-weighted bias, variability and timing, their overall score is heavily affected by compensation effects – except

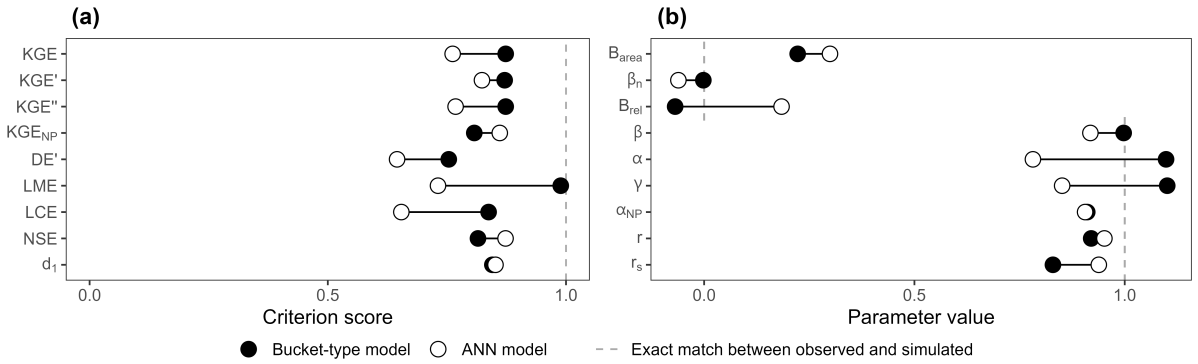


Figure 5.7: (a) Score of the ANN and bucket-type models according to the different performance criteria. (b) Values of the parameters used in the calculation of the performance criteria.

in the case of a large error on one parameter. In our case, all parameters have similar errors, which results in a better KGE for the bucket-type model compared to the ANN model. This applies to all the KGE variants except the KGE_{NP} where the error on r_s is significant, resulting in a better score for the ANN model. The LME score is extremely high (0.99) for the bucket-type model, which is probably due to the compensation of r and α identified by Choi (2022). Also, using γ instead of α for assessing the variability seems to lower counterbalancing errors.

Interestingly, the cumulative sum of the absolute bias error between simulated and observed values (Figure 5.6b) is smaller for the ANN model (1394 m^3) than the bucket-type model (1611 m^3), but still the relative bias and variability parameters are better for the bucket-type model. This observation highlights how counterbalancing errors can impair the evaluation of hydrological models: seemingly better parameters values (bias and variability) that increase criteria scores are not necessarily associated with an increase in model relevance.

5.5 Recommendations

The aim of this paper is primarily to raise awareness among modellers. Performance criteria generally comprise several aspects of the characteristics of a model into a single value, which can lead to an inaccurate assessment of said aspects. Ultimately, all criteria have their flaws and should be carefully selected with regards to the aim of the model.

5.5.1 Use of relevant performance criteria

Table 5.1 summarises the presence and impact of counterbalancing errors, as well as the advantages and drawbacks (as reported in other studies) of the different performance criteria. The recommendations on counterbalancing errors are based on the results of this research – i.e. synthetic and real case studies. The KGE and all its variants are affected by counterbalancing errors with varying degrees of intensity: (i) mildly impacted (+) for the KGE', KGE_{NP} and DE, (ii) moderately impacted (++) for the KGE, KGE" and LCE, and (iii) strongly impacted (+++) for the LME. In this study, the NSE and d_1 stand out as clearly better since they have no counterbalancing errors. However, they have other drawbacks that are not associated with counterbalancing errors, especially the NSE with its limitations related to variability (Gupta et al., 2009). We thus recommend

using performance criteria that are not or less prone to counterbalancing errors (d_1, KGE', KGE_{NP}, DE).

5.5.2 Use of scaling factors

The assessment of the hydrological models in the real case study shows how concurrent over- and underestimation can generate counterbalancing errors on bias and variability parameters. For the case study considered in this paper, the ANN model, although offering a better simulation, is evaluated as – sometimes considerably – worse than the bucket-type model, because it slightly underestimates the total volume. This has a great impact on the overall score, as the KGE and its variant are calculated with both bias and variability parameters accounting for 2/3 of the overall criterion score.

While the overall balance (bias) may be a desired feature in a model, we showed that a good value may be accidental and result from counterbalancing errors. The common use of the KGE neglects one of the original proposals which is to weight the parameters β, α and r in the equation. Gupta et al. (2009) proposed an alternative equation for adjusting the emphasis on the different aspects of a model:

$$KGE_s = 1 - \sqrt{[s_\alpha(\alpha - 1)]^2 + [s_\beta(\beta - 1)]^2 + [s_r(r - 1)]^2} \tag{5.22}$$

with s_r, s_β and s_α the scaling factors of r, β and α, respectively. By default, these factors are equal to 1, which induces a weight of 1/3 on the parameter in absolute value (r) and 2/3 on the parameters in relative values (β, α). To the best of our knowledge, only Mizukami et al. (2019) ever considered changing the scaling factors when using the KGE. We suggest to carefully consider such scaling factors for the calibration and the evaluation of hydrological models using the KGE and its variants. Depending on the purpose of the model, they can help to emphasise particular aspects of a model or reduce the influence of relative parameters and counterbalancing errors.

Figure 5.8 shows how emphasising absolute parameters with scaling factors helps to reduce the influence of counterbalancing errors for the KGE (Figure 5.8a) and its most used variant KGE' (Figure 5.8b). The default value (1-1-1) – corresponding to scaling factors of 1 for α (KGE) or γ (KGE'), 1 for β and 1 for r, respectively – is compared to other factor combinations with different ratios between absolute and relative parameters. The 2:1 ratio (2-1-1) increases counterbalancing errors as the emphasis is on the relative parameters, while the 1:2, 1:3, 1:4 and 1:5 ratios decrease counterbalancing errors. The ANN model is evaluated as better with the 1:4 ratio for the KGE and the 1:3 ratio for the KGE', highlighting that the KGE' is less sensitive to counterbalancing errors. This also shows how the score of a performance criterion and by extension its interpretation can be radically different depending on the parameters used in the equation. This is why a multi-criteria framework can strengthen the evaluation of models and reduce the uncertainties of performance criteria scores.

Criterion	Year	Affected by CE	Impact of CE	Advantages	Drawbacks
KGE	2009	Yes	++	Variability is not underestimated (Gupta et al., 2009)	Still slight underestimation of high discharges (Gupta et al., 2009) Bias and variability are cross correlated (Kling et al., 2012) Implicit assumptions of data linearity, data normality and absence of outliers (Pool et al., 2018) No inherent benchmark (Knoben et al., 2019) Not suited to logarithmic transformation of discharge (Santos et al., 2018)
KGE'	2012	Yes	+	Bias and variability are not cross correlated (Kling et al., 2012)	
KGE"	2021	Yes	++	The score is not overly sensitive to mean values close to zero (Santos et al., 2018; Tang et al., 2021)	
KGE_{NP}	2018	Yes	+	Reduce the impact of implicit assumptions of data linearity, data normality and absence of outliers by using non-parametric parameters (Pool et al., 2018)	
DE	2021	Yes	+	Aims to provide a stronger link to hydrological processes (Schwemmle et al., 2021)	
LME	2020	Yes	+++	Improve the simulation of extreme events (Liu, 2020)	Infinite number of solutions for the maximum score (Lee and Choi, 2022) Inclination to overestimate high flows and underestimates low flows (Lee and Choi, 2022)
LCE	2022	Yes	++	Improve the simulation of extreme events (Lee and Choi, 2022)	
NSE	1970	No	/		The contribution of β_n depends on the variability (Gupta et al., 2009) Variability is underestimated (Gupta et al., 2009) The benchmark is inappropriate for highly variable discharges (Gupta et al., 2009)
d_1	1985	No	/	Address the shortcomings of r and the coefficient of determination (Willmott, 1981) The score is less sensitive to errors concentrated in outliers in comparison to the original index of agreement (Willmott et al., 1985)	

Table 5.1: Presence and impact of counterbalancing errors (CE) on the assessment of model performance of different performance criteria. The impact of CE is denoted as null (/), mild (+), moderate (++), or strong (+++).

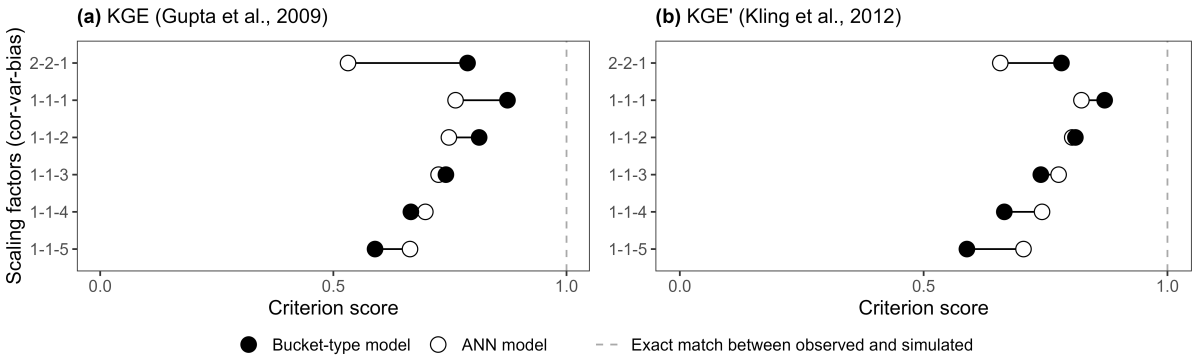


Figure 5.8: (a) KGE and (b) KGE' scores of the ANN and bucket-type models (Figure 5.6a) according to different scaling factors. The y-axis numbers correspond to the scaling factors of the variability, bias and timing parameters, with the default being 1-1-1.

5.6 Conclusion

This study sets out to explore the influence of counterbalancing errors and raise awareness among modellers about the use of performance criteria for calibrating and evaluating hydrological models. A total of nine performance criteria (NSE, KGE, KGE', KGE", KGE_{NP}, DE, LME, LCE and d_1) are analysed. The investigation of synthetic time series and real hydrological models shows that concurrent over- and underestimation of multiple parts of a discharge time series may favour bias and variability parameters. This especially concerns the bias parameters (β, β_n and B_{rel}) as their values are all influenced by counterbalancing errors in both synthetic time series and the real case study. On the other hand, the impact of counterbalancing errors on the variability parameters seems to depend on the time series: only the value of α is influenced in the synthetic time series, while the values of all variability parameters (α, γ, $|B_{area}|$ and α_{NP}) are influenced in the real hydrological models. As bias and variability parameters generally account for 2/3 of the weight in the equation of certain performance criteria, this can lead to an overall higher criterion score without being associated with an increase in model relevance. This is especially concerning for the KGE and its variants, as they generally use relative parameters for evaluating bias and variability in hydrological models. These findings highlight the importance of carefully choosing a performance criterion adapted to the purpose of the model. Recommendations also include the use of scaling factors to emphasise different aspects of a hydrological model and reduce the influence of relative parameters on the overall score of the performance criterion. Further research could explore the appropriate values of scaling factors to use, depending on the modelling approach and the purpose of the study.

Acknowledgements

We thank the French Ministry of Higher Education and Research for the thesis scholarship of G. Cinkus as well as the European Commission for its support through the Partnership for Research and Innovation in the Mediterranean Area (PRIMA) program under Horizon 2020 (KARMA project, grant agreement number 01DH19022A). We further thank the Slovenian Research Agency for financial support within the project Infiltration processes in forested karst aquifers under changing environment (No. J2-1743). For the data provided, we also acknowledge the Slovenian Environment Agency (ARSO, 2021a; ARSO,

2021b).

The analyses were performed using R (R Core Team, 2021) and the following packages: readxl, readr, dplyr, tidyr, ggplot2, lubridate (Wickham et al., 2019), cowplot (Wilke, 2020), diag-eff (Schwemmle et al., 2021), flextable (Gohel, 2021), hydroGOF (Zambrano-Bigiarini, 2020), HydroErr (Roberts et al., 2018) and padr (Thoen, 2021). The manuscript was written with the Rmarkdown framework (Allaire et al., 2021; Xie et al., 2018; Xie et al., 2020).

Author contribution

GC, NM and HJ conceptualised the study and designed the methodology. GC and AW developed the software code. GC performed the experiments and investigated and visualised the results. AW provided the ANN results for the case study. GC wrote the original paper draft with contributions from AW and NR. All the authors contributed to the interpretation of the results and review and editing of the paper draft. NM and HJ supervised the work.

Code and data availability

We provide complete scripts for reproducing the results on the synthetic time series (section 3), as well as ANN model code and KarstMod `.properties` file (bucket-type model) on GitHub (Cinkus and Wunsch, 2022). Unica spring discharge time series and meteorological data are available from the Slovenian Environment Agency (ARSO, 2021a; ARSO, 2021b).

Competing Interest

The authors declare that they have no conflict of interest.

5.7 Appendix

5.A Common and recently developed performance criteria applied to the synthetic time series and the real case study

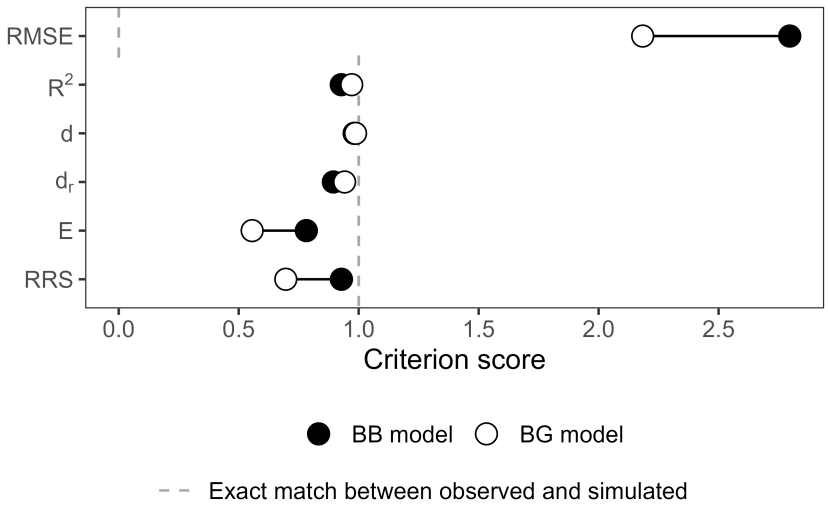


Figure 5.9: Score of the BB and BG transformations according to other common and recently developed performance criteria: the Root Mean Square Error (RMSE), the coefficient of determination R2, the index of agreement d (Willmott, 1981), the refined index of agreement d_r (Willmott et al., 2012), the Onyutha efficiency E and the revised R-squared RRS (Onyutha, 2022).

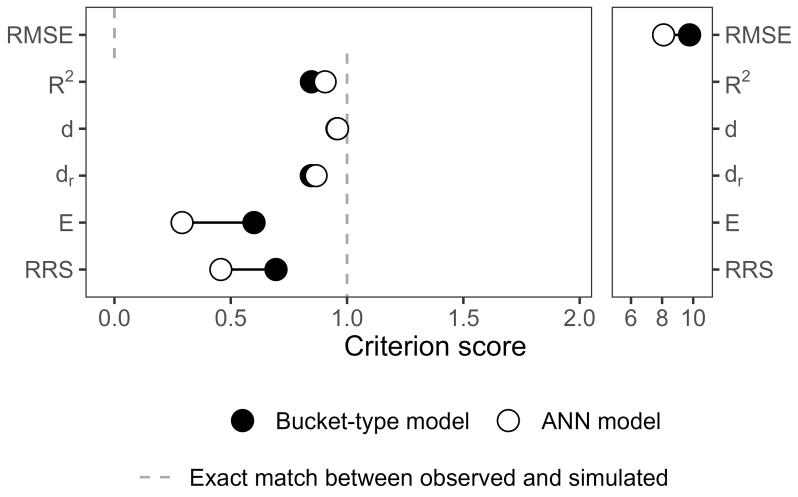


Figure 5.10: Score of the ANN and bucket-type models according to other common and recently developed performance criteria: the Root Mean Square Error (RMSE), the coefficient of determination R2, the index of agreement d (Willmott, 1981), the refined index of agreement d_r (Willmott et al., 2012), the Onyutha efficiency E and the revised R-squared RRS (Onyutha, 2022).

5.8 References for Chapter 5

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.* (Cit. on p. 124).

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2021). *Rmarkdown: Dynamic documents for r.* (Version 2.22, R package). (Cit. on p. 130).

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., & FAO (Eds.). (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements.* Food and Agriculture Organization of the United Nations. (Cit. on p. 123).

Althoff, D., & Rodrigues, L. N. (2021). Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *J. Hydrol.*, *600*, 126674. https://doi.org/10.1016/j.jhydrol.2021.126674 (cit. on p. 113).

ARSO. (2021a). *Archive of meteorological data, Ministry of the Environment and Spatial Planning, Slovenian Environment Agency, Archive of hydrological data.* Retrieved March 6, 2023, from http://vode.arso.gov.si/hidarhiv/ (cit. on pp. 123, 129, 130).

ARSO. (2021b). *Archive of meteorological data, Ministry of the Environment and Spatial Planning, Slovenian Environment Agency, Archive of meteorological data.* Retrieved June 6, 2023, from http://www.meteo.si/ (cit. on pp. 123, 129, 130).

Barber, C., Lamontagne, J. R., & Vogel, R. M. (2020). Improved estimators of correlation and R2 for skewed hydrologic data. *Hydrological Sciences Journal*, *65*(1), 87–101. https://doi.org/10.1080/02626667.2019.1686639 (cit. on p. 124).

Beven, K. (2019). How to make advances in hydrological modelling. *Hydrol. Res.*, *50*(6), 1481–1494. https://doi.org/10.2166/nh.2019.134 (cit. on p. 113).

Biondi, D., Freni, G., Iacobellis, V., Mascaro, G., & Montanari, A. (2012). Validation of hydrological models: Conceptual basis, methodological approaches and a proposal for a code of practice. *Physics and Chemistry of the Earth, Parts A/B/C*, *42–44*, 70–76. https://doi.org/10.1016/j.pce.2011.07.037 (cit. on p. 113).

Choi, H. I. (2022). Comment on Liu (2020): A rational performance criterion for hydrological model. *J. Hydrol.*, *606*, 126927. https://doi.org/10.1016/j.jhydrol.2021.126927 (cit. on pp. 116, 120, 126).

Chollet, F. (2015). *Keras.* Retrieved June 6, 2023, from https://github.com/keras-team/keras (cit. on p. 124).

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Goldscheider, N., Ravbar, N., Fernández-Ortega, J., Barberá, J. A., Andreo, B., & Chen, Z. (2022). *Possible bias in the assessment of karst hydrological model performance. Example of alpha and beta parameters compensation when using the KGE as performance criterion.* (Oral presentation). EGU22, the 24th EGU General Assembly. Vienna, Austria. https://doi.org/10.5194/egusphere-egu22-5149 (cit. on p. 111).

Cinkus, G., Mazzilli, N., Jourde, H., Wunsch, A., Liesch, T., Ravbar, N., Chen, Z., & Goldscheider, N. (2023a). When best is the enemy of good – critical evaluation of performance criteria in hydrological models. *Hydrol. Earth Syst. Sci.*, *27*(13), 2397–2411. https://doi.org/10.5194/hess-27-2397-2023 (cit. on p. 111).

Cinkus, G., & Wunsch, A. (2022). *Busemorose/KGE_critical_evaluation: Model code release.* https://doi.org/10.5281/zenodo.7274031 (cit. on p. 130).

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J. A., Andreo, B., Goldscheider, N., & Jourde, H. (2023b). Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions. *Hydrol. Earth Syst. Sci.*, *27*(10), 1961–1985. https://doi.org/10.5194/hess-27-1961-2023 (cit. on p. 123).

Clark, M. P., Vogel, R. M., Lamontagne, J. R., Mizukami, N., Knoben, W. J. M., Tang, G., Gharari, S., Freer, J. E., Whitfield, P. H., Shook, K. R., & Papalexiou, S. M. (2021). The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resour. Res.*, *57*(9), e2020WR029001. https://doi.org/10.1029/2020WR029001 (cit. on p. 113).

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics: Fourth International Student Edition*. W. W. Norton & Company. (Cit. on p. 114).

Gabrovšek, F., Kogovšek, J., Kovačič, G., Petrič, M., Ravbar, N., & Turk, J. (2010). Recent Results of Tracer Tests in the Catchment of the Unica River (SW Slovenia). *Acta Carsologica*, *39*(1), 27–37. https://doi.org/10.3986/ac.v39i1.110 (cit. on p. 123).

Gohel, D. (2021). *Flextable: Functions for tabular reporting* (Version 0.9.1, R package). (Cit. on p. 130).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003 (cit. on pp. 113, 115–117, 126–128).

Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour. Res.*, *34*(4), 751–763. https://doi.org/10.1029/97WR03495 (cit. on p. 113).

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.*, *52*(3), 218–242. https://doi.org/10.1002/2013RG000443 (cit. on p. 113).

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55 (cit. on p. 124).

Jackson, E. K., Roberts, W., Nelsen, B., Williams, G. P., Nelson, E. J., & Ames, D. P. (2019). Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. *Environ. Model. Softw.*, *119*, 32–48. https://doi.org/10.1016/j.envsoft.2019.05.001 (cit. on pp. 113, 114).

Jain, S. K., Mani, P., Jain, S. K., Prakash, P., Singh, V. P., Tullos, D., Kumar, S., Agarwal, S. P., & Dimri, A. P. (2018). A Brief review of flood forecasting techniques and their applications. *Int. J. River Basin Manag.*, *16*(3), 329–344. https://doi.org/10.1080/15715124.2017.1411920 (cit. on p. 113).

Kauffeldt, A., Wetterhall, F., Pappenberger, F., Salamon, P., & Thielen, J. (2016). Technical review of large-scale hydrological models for implementation in operational flood forecasting schemes on continental level. *Environ. Model. Softw.*, *75*, 68–76. https://doi.org/10.1016/j.envsoft.2015.09.009 (cit. on p. 113).

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011 (cit. on pp. 113, 116, 128).

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019 (cit. on pp. 113, 128).

Kovačič, G. (2010). Hydrogeological study of the Malenščica karst spring (SW Slovenia) by means of a time series analysis. *Acta Carsologica*, *39*(2), 201–215. https://doi.org/10.3986/ac.v39i2.93 (cit. on p. 123).

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005 (cit. on p. 113).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539 (cit. on p. 123)
Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Reviews Subject_term: Computer science;Mathematics and computing Subject_term_id: computer-science;mathematics-and-computing.

Lee, J. S., & Choi, H. I. (2022). A rebalanced performance criterion for hydrological model calibration. *J. Hydrol.*, *606*, 127372. https://doi.org/10.1016/j.jhydrol.2021.127372 (cit. on pp. 116, 128).

Legates, D. R., & McCabe Jr., G. J. (1999). Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998WR900018 (cit. on pp. 113, 117).

Liu, D. (2020). A rational performance criterion for hydrological model. *J. Hydrol.*, *590*, 125488. https://doi.org/10.1016/j.jhydrol.2020.125488 (cit. on pp. 116, 128).

Massmann, C., Woods, R., & Wagener, T. (2018). Reducing equifinality by carrying out a multi-objective evaluation based on the bias, correlation and standard deviation errors, 11457 (cit. on p. 114)
ADS Bibcode: 2018EGUGA..2011457M.

Mayaud, C., Gabrovšek, F., Blatnik, M., Kogovšek, B., Petrič, M., & Ravbar, N. (2019). Understanding flooding in poljes: A modelling perspective. *J. Hydrol.*, *575*, 874–889. https://doi.org/10.1016/j.jhydrol.2019.04.092 (cit. on p. 123).

Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Soglio, L. D., & Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.*, *122*, 103927. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on p. 124).

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a (cit. on p. 124).

Mizukami, N., Rakovec, O., Newman, A. J., Clark, M. P., Wood, A. W., Gupta, H. V., & Kumar, R. (2019). On the choice of calibration metrics for "high-flow" estimation using hydrologic models. *Hydrol. Earth Syst. Sci.*, *23*(6), 2601–2614. https://doi.org/10.5194/hess-23-2601-2019 (cit. on pp. 116, 127).

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Trans. ASABE*, *58*(6), 1763–1785. https://doi.org/10.13031/trans.58.10715 (cit. on p. 113).

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.*, *306*(1), 127–145. https://doi.org/10.1016/j.jhydrol.2004.09.005 (cit. on p. 113).

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, *10*(3), 282–290 (cit. on pp. 113, 115).

Nogueira, F. (2014). *Bayesian Optimization: Open source constrained global optimization tool for Python*. Retrieved June 6, 2023, from https://github.com/fmfn/BayesianOptimization (cit. on p. 124).

Onyutha, C. (2022). A hydrological model skill score and revised R-squared. *Hydrology Research*, *53*(1), 51–64. https://doi.org/10.2166/nh.2021.071 (cit. on p. 131).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python*. (Cit. on p. 124).

Petric, M. (2010). Chapter 10.3 - Case Study: Characterization, exploitation, and protection of the Malenščica karst spring, Slovenia. In N. Kresic & Z. Stevanovic (Eds.), *Groundwater Hydrology of Springs* (pp. 428–441). Butterworth-Heinemann. https://doi.org/10.1016/B978-1-85617-502-9.00021-9 (cit. on p. 123).

Pool, S., Vis, M., & Seibert, J. (2018). Evaluating model performance: Towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.*, *63*(13-14), 1941–1953. https://doi.org/10.1080/02626667.2018.1552002 (cit. on pp. 115, 116, 128).

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. Retrieved June 6, 2023, from https://www.R-project.org/ (cit. on p. 130).

Ravbar, N., Barberá, J. A., Petrič, M., Kogovšek, J., & Andreo, B. (2012). The study of hydrodynamic behaviour of a complex karst system under low-flow conditions using natural and artificial tracers (the catchment of the Unica River, SW Slovenia). *Environ. Earth Sci.*, *65*(8), 2259–2272. https://doi.org/10.1007/s12665-012-1523-4 (cit. on p. 123).

Reback, J., jbrockmendel, McKinney, W., Bossche, J. V. den, Augspurger, T., Cloud, P., Hawkins, S., Roeschke, M., gfyoung, Sinhrks, Klein, A., Petersen, T., Hoefler, P., Tratner, J., She, C., Ayd, W., Naveh, S., Garcia, M., Darbyshire, J. H. M., … Seabold, S. (2021). *Pandas-dev/pandas: Pandas 1.3.5*. https://doi.org/10.5281/zenodo.5774815 (cit. on p. 124).

Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, *480*, 33–45. https://doi.org/10.1016/j.jhydrol.2012.12.004 (cit. on p. 113).

Roberts, W., Williams, G. P., Jackson, E., Nelson, E. J., & Ames, D. P. (2018). Hydrostats: A Python Package for Characterizing Errors between Observed and Predicted Time Series. *Hydrology*, *5*(4), 66. https://doi.org/10.3390/hydrology5040066 (cit. on p. 130).

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrol. Earth Syst. Sci.*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018 (cit. on pp. 113, 116, 128).

Schwemmle, R., Demand, D., & Weiler, M. (2021). Technical note: Diagnostic efficiency – specific evaluation of model performance. *Hydrol. Earth Syst. Sci.*, *25*(4), 2187–2198. https://doi.org/10.5194/hess-25-2187-2021 (cit. on pp. 115, 116, 128, 130).

Seibert, J., Vis, M. J. P., Lewis, E., & van Meerveld, H. (2018). Upper and lower benchmarks in hydrological modelling. *Hydrological Processes*, *32*(8), 1120–1125. https://doi.org/10.1002/hyp.11476 (cit. on p. 113).

Tang, G., Clark, M. P., & Papalexiou, S. M. (2021). SC-Earth: A Station-Based Serially Complete Earth Dataset from 1950 to 2019. *J. Clim.*, *34*(16), 6493–6511. https://doi.org/10.1175/JCLI-D-21-0067.1 (cit. on pp. 116, 128).

Thoen, E. (2021). *Padr: Quickly get datetime data ready for analysis.* (Version 0.6.2, R package). (Cit. on p. 130).

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37 (cit. on p. 124).

van Rossum, G. (1995). *Python Tutorial.* CWI. (Cit. on p. 123).

van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2009). Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Adv. Water Resour.*, *32*(8), 1154–1169. https://doi.org/10.1016/j.advwatres.2009.03.002 (cit. on p. 113).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *J. Open Source Softw.*, *4*(43), 1686. https://doi.org/10.21105/joss.01686 (cit. on p. 130).

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* (Version 1.1.1, R package). (Cit. on p. 130).

Willmott, C. J. (1981). On the validations of models. *Phys. Geogr.*, *2*(2), 184–194. https://doi.org/10.1080/02723646.1981.10642213 (cit. on pp. 128, 131).

Willmott, C. J., Ackleson, S. G., Davis, R. E., Feddema, J. J., Klink, K. M., Legates, D. R., O'Donnell, J., & Rowe, C. M. (1985). Statistics for the evaluation and comparison of models. *J. Geophys. Res.*, *90*(C5), 8995. https://doi.org/10.1029/JC090iC05p08995 (cit. on pp. 117, 128).

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *Intern. J. Climatol.*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419 (cit. on pp. 113, 131).

Wöhling, T., Samaniego, L., & Kumar, R. (2013). Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. *Environ Earth Sci*, *69*(2), 453–468. https://doi.org/10.1007/s12665-013-2306-2 (cit. on p. 113).

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide.* Chapman and Hall/CRC. (Cit. on p. 130).

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook.* Chapman and Hall/CRC. (Cit. on p. 130).

Zambrano-Bigiarini, M. (2020). *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series.* https://doi.org/10.5281/zenodo.839854 (cit. on p. 130).

Chapter 6

Comparison of artificial neural networks and reservoir models

Hydrological models are widely used to characterise, understand and manage hydrosystems. Lumped parameter models are of particular interest in karst environments given the complexity and heterogeneity of these systems. There is a multitude of lumped parameter modelling approaches, which can make it difficult for a manager or researcher to choose. We therefore conducted a comparison of two lumped parameter modelling approaches: artificial neural networks (ANN) and reservoir models. We investigate five karst systems in the Mediterranean and Alpine regions with different characteristics in terms of climatic conditions, hydrogeological properties and data availability. We compare the results of ANN and reservoir modelling approaches using several performance criteria over different hydrological periods. The results show that both ANN and reservoir models can accurately simulate karst spring discharge, but also that they have different advantages and drawbacks: (i) ANN models are very flexible regarding the format and amount of input data, (ii) reservoir models can provide good results even with few years of relevant discharge in the calibration period, and (iii) ANN models seem robust for reproducing high-flow conditions while reservoir models are superior for reproducing low-flow conditions. However, both modelling approaches struggle to reproduce extreme events (droughts, floods), which is a known problem in hydrological modelling. For research purposes, ANN models have been shown to be useful for identifying recharge areas and delineating catchments, based on insights into the input data. Reservoir models are adapted to understand the hydrological functioning of a system, by studying model structure and parameters.

A part of this work has contributed to the KARMA project in the form of a deliverable and has been presented during a progress meeting. This work resulted in a publication in Hydrology and Earth System Sciences (Cinkus et al., 2023).

Article:

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J. A., Andreo, B., Goldscheider, N., and Jourde, H.: Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean regions, Hydrol. Earth Syst. Sci., 27, 1961–1985, https://doi.org/10.5194/hess-27-1961-2023, 2023.

Contents

6.1 Introduction

Karst systems are complex and heterogeneous media. High contrasts in porosity and permeability induce a high variability in infiltration and internal flow processes (Bakalowicz, 2005; Ford and Williams, 2007) which can be difficult to assess. Considering the increasing demand for water and that around 9 % of the world's population (up to 90 % in some parts of the Mediterranean area) depends on karst water resources for drinking water supply (Stevanović, 2019), the characterisation of karst systems functioning and water availability become a major challenge for water resource management. Among the numerous methods to study karst systems (Goldscheider, 2015), hydrological models are useful to characterise karst functioning, and specially to predict the impact of climate and land use changes (Hartmann et al., 2014). Hydrological models can be grouped into lumped parameter and distributed approaches (Kovács and Sauter, 2007). While distributed models divide a karst system into a two- or three-dimensional grid, for which each cell is assigned appropriate hydraulic parameters and system states, lumped parameter models are based on the mathematical analysis of input data (e.g. precipitation, temperature) for simulating spring discharge time series. They include (i) "black-box" models such as neural-networks-based approaches, which use no a priori information about the functioning of a system; and (ii) "conceptual" models, which are based on a conceptual representation of a karst system – e.g., for the reservoir models, a succession of one or several reservoirs using simplified physical transfer functions.

The choice of a modelling approach depends mainly on the objective of the study, but also on the current knowledge of the system, the available data, and regional/institutional preferences (Addor and Melsen, 2019). For karst systems, the available data are often scarce and poorly reflect the heterogeneity of the meteorological and karst processes. Distributed models require a lot of diverse data with high spatial and temporal resolution for defining physical parameters and thus can be tough to use in a scarce data context (Hartmann et al., 2014). On the other hand, lumped parameter models permit the study of complex and heterogeneous karst systems without requiring extensive meteorological and system-related data with high spatial resolution. Artificial neural networks (ANN) have been successfully used to simulate karst spring discharge (Hu et al., 2008; Kurtulus and Razack, 2007; Meng et al., 2015; Wunsch et al., 2022), predict and forecast water flood/inrush (Kong-A-Siou et al., 2011; Wu et al., 2008) and manage the exploitation of karst aquifers (Kong-A-Siou et al., 2015; Yin et al., 2011). Reservoir models have also been successfully used to simulate karst spring discharge (Dubois et al., 2020; Fleury et al., 2007), manage the exploitation of karst aquifers (Fleury et al., 2009; Zhou et al., 2021), as well as characterise specific functioning in karst systems (Bittner et al., 2020; Jukić and Denić-Jukić, 2009; Perrin et al., 2003; Tritz et al., 2011). This approach is well suited to karst systems due to the high heterogeneity and low level of knowledge of their structure (Fleury et al., 2009; Hartmann et al., 2012). Although several authors compared the performance of different ANN models (Cheng et al., 2020; Kovačević et al., 2018; Kurtulus and Razack, 2010) and studied structure and parameter equifinality in reservoir models (Gondwe et al., 2011; Hartmann et al., 2012; Makropoulos et al., 2008; Mazzilli et al., 2012), only a few studies have been conducted on the comparison of both approaches in karst environments (Jeannin et al., 2021; Kong-A-Siou et al., 2014; Sezen et al., 2019). Kong-A-Siou et al. (2014) observed that ANN models are more effective at accounting for the non-linearity of karst systems during extreme events (dry and flood periods), while reservoir models were better at representing the hydrological functioning

of the system during intermediate water periods. Sezen et al. (2019) observed that ANN models were better for simulating low-flow periods and reservoir models for simulating spring discharges on predominantly non-karst catchments. Jeannin et al. (2021) emphasised the great potential of ANN models but highlighted two main limitations: (i) they require long time series to accurately learn the functioning of a karst system, and (ii) usually no information about specific functioning of a system can be deduced from the results.

The performance of ANN and reservoir models can therefore be influenced by the characteristics of the catchment, as well as the format and length of the input data. The aim of the present study is to help researchers and stakeholders to choose between ANN and reservoir modelling approaches for simulating karst spring discharge, depending on their purpose and the available data. This research provides the first extensive comparison of ANN and reservoir models in karst hydrology by investigating results on five study sites with different context and input data. We use ANN as they have proven to be fast and reliable for modelling hydrological time series (Jeannin et al., 2021; Van et al., 2020; Wunsch et al., 2021). We specifically apply one-dimensional convolutional neural networks (CNNs) because in an earlier study (Wunsch et al., 2022) we were able to demonstrate their high ability to perform karst spring discharge modelling. Furthermore, they have some favourable properties compared to popular recurrent neural networks (e.g. the LSTMs), such as a batch-wise training procedure which makes them considerably faster and computationally less expensive. Reservoir modelling is carried out using the KarstMod platform, as it provides a powerful modular interface for varying the structure, parameters and transfer functions of the conceptual model (Mazzilli et al., 2019). This research seeks to address the following research questions:

i. What are the advantages and drawbacks of ANN and reservoir models in karst hydrogeology?

ii. To which extent can ANN and reservoir models be used to get a better understanding of system functioning?

iii. What are the implications from a stakeholder's perspective?

6.2 Data and study sites

We compare ANN and reservoir modelling approaches using data from five different well-studied karst systems (Table 6.1, Figure 6.1). All the systems have different characteristics in terms of hydrogeological properties (e.g. catchment area, karstification), data availability (e.g. length of the time series, number of meteorological stations, time step), and environmental conditions (e.g. climate, anthropogenic influence). Each study site is detailed in the following sub-sections and further details about the meteorological data can be found in Appendix 6.A.

6.2.1 Aubach spring, Austria

Aubach spring (1080 m asl) is a large non-permanent spring located in the Hochifen-Gottesacker area, on the border between Germany and Austria (Northern Alps). The Hochifen-Gottesacker system covers an area of about 35 km^2 and its altitude varies between 1000 and 2230 m asl (Chen et al., 2018). The area is under a cool temperate,

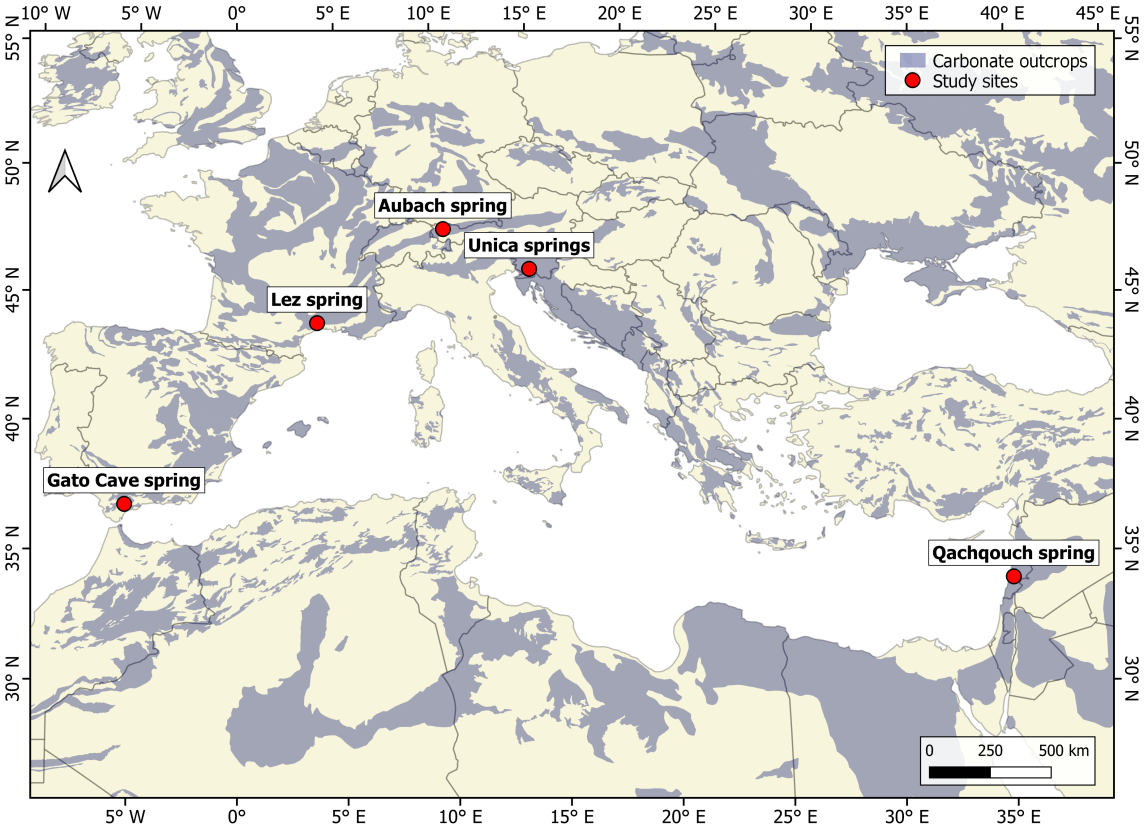


Figure 6.1: Locations of the study sites (carbonate outcrops from Goldscheider et al., 2020).

Spring	Country	Climate	Catchment area [km^2]	Q_{mean} [m^3 s^{-1}]	P_{an} [mm]	Period
Aubach	Austria	Cool temperate and humid	9	0.91	2113	2012-11-20 – 2020-10-31
Gato Cave	Spain	Mediterranean	69–79	1.50	1872	1963-10-02 – 2015-04-29
Lez	France	Mediterranean	130	0.84	904	2008-10-21 – 2020-12-03
Qachqouch	Lebanon	Mediterranean	56	2.01	1293	2015-09-06 – 2020-02-05
Unica	Slovenia	Mediterranean	820	21.97	1605	1961-01-02 – 2018-12-31

Table 6.1: Summary of the studied springs and areas. Q_{mean} corresponds to the mean observed discharge and P_{an} to the annual mean precipitation over the considered period.

humid climate and is strongly affected by snow accumulation and melting, which typically occur between November and May (Chen et al., 2018). The spring is located in the Schwarzwasser valley, which follows the geological contact between highly karstified limestone (Schrattenkalk formation) in the northern and western parts and impermeable sedimentary rocks of the Flysch zone in the southern part (Goldscheider, 2005). The main catchment of Aubach spring is estimated to be approximately 9 km^2 (Chen and Goldscheider, 2014; Goldscheider, 2005). The spring also receives inflow from several upstream karst catchments and the Flysch zone, where surface runoff can sink into an estavelle and pass through an underground karst conduit during low-flow periods, as demonstrated by multiple tracer tests (Goldscheider, 2005).

Precipitation and temperature data were obtained from three meteorological stations located outside the catchment. The potential evapotranspiration is calculated using data from one station with the modified Turk-Ivanov approach after (Wendling and Müller, 1984), described in (Conradt et al., 2013).

6.2.2 Gato Cave spring, Spain

Gato cave spring (462 m asl) is one of the main outlets of the karst system of Sierra de Lìbar. It is located in the north-western part of the province of Málaga, within the boundaries of the Grazalema Natural Park, about 75 km west of Málaga. The altitude of the Sierra de Lìbar varies between 400 and 1400 m asl according to the main north-east/south-west mountain alignments. The area is under a Mediterranean climate, with an average annual precipitation of about 1500 mm and is defined by a strong seasonal pattern (Andreo et al., 2006). The site is located within the External Zone of the Betic Cordillera and presents mainly Jurassic limestones and dolomites, Cretaceous-Paleogene marly-limestones and Tertiary clays and sandstones (Flysch) that cover the whole Mesozoic rock sequence. The Jurassic rocks outcrop as anticlinal cores, while the synclines and tectonic grabens are composed of Cretaceous rocks (Martín-Algarra, 1987). The Hundidero-Gato system constitutes a binary karst system where a wide range of well-developed karst landforms are found, such as karrenfields, swallow holes and caves. These features strongly condition recharge, which is primarily produced in two ways: (i) autochthonous, by direct infiltration of rainfall through carbonate outcrops (20–40 km^2) as well as rainwater that infiltrates through swallow holes in poljes; and (ii) allochthonous, as a contribution from runoff produced in the Gaduares River basins (43.5 km^2). This runoff is stored in the Montejaque dam, which was built on karstified limestone, resulting in water losses in the reservoir and, consequently, the artificial recharge of the aquifer through the Hundidero cave (Andreo et al., 2004).

Precipitation and temperature data are from the meteorological station of Grazalema, which is the closest to the catchment, and therefore the most representative. Potential evapotranspiration is calculated with the Hargreaves-Samani approach (Hargreaves and Samani, 1985).

6.2.3 Lez spring, France

The Lez spring (64 m asl) is located 15 km north of Montpellier, and the altitude of its catchment varies between 64 and 655 m asl. The Lez catchment is exposed to a Mediterranean climate, which is characterised by hot, dry summers, mild winters and wet autumns. As a large part of the hydrogeological basin is relatively impermeable due

to the presence of marl and marly-limestone formation, the effective recharge area of the Lez spring covers about 130 km^2 (Fleury et al., 2009) and corresponds to Jurassic limestone outcrops. Localised infiltration occurs through fractures and sinkholes along the basin and through the major geologic fault of Corconne-Les Matelles. The Lez aquifer is subject to anthropic pressure (i.e. exploitation for water supply) with pumping directly into the karstic conduit. The discharge is measured at the spring pool and is regularly zero during low water periods, when the pumping rate exceeds the natural discharge of the spring.

Precipitation data are from four meteorological stations. Three are located in the catchment and one is located about 5 km west of the catchment. Potential evapotranspiration is calculated with the Oudin approach (Oudin et al., 2005). Temperature data are from the Prades-le-Lez meteorological station.

6.2.4 Qachqouch spring, Lebanon

Qachqouch spring (64 m asl) is located in the Nahr el-Kalb catchment and originates from a Jurassic karst aquifer. The recharge area is estimated to be about 56 km^2 with altitudes ranging from 60 to over 1500 m asl (Doummar et al., 2018; Dubois et al., 2020). The catchment is primarily exposed to a Mediterranean climate, with snow influence at higher altitudes (Dubois et al., 2020)). The lithology mainly consists in Jurassic karstified limestone and dolomitic limestone (on the higher plateaus) changing to more massive micritic limestone in the lower part of the catchment. The Qachqouch system is characterised by a duality of flow in a low permeability matrix and a high permeability conduit system (Dubois, 2017). Potential runoff inflows from higher altitudes and infiltrates downstream into the Jurassic karst aquifer.

Precipitation and temperature data are from two meteorological stations. One is located in the catchment at 950 m asl. The other, with a heated rain gauge, is located 22 km north-east of the catchment at 1700 m asl (Doummar et al., 2018). Potential evapotranspiration is calculated using data from the 950 m station with the modified Penman-Monteith approach (Allen et al., 1998).

6.2.5 Unica springs, Slovenia

Unica springs (450 m asl) are the outlets of a complex karst system with an estimated recharge area of about 820 km^2. The area is under a moderate continental climate and is strongly influenced by snow accumulation and melting. It is subdivided into three subcatchments, with a predominance of (i) allogenic infiltration from two subcatchments drained by sinking rivers flowing through a chain of karst poljes and a river valley, and (ii) autogenous infiltration through a karst plateau with highly karstified limestone (Gabrovšek et al., 2010; Kovačič, 2010; Petric, 2010). The poljes follow each other in a descending series at altitudes between 450 and 750 m asl and are connected in a common hydrological system. Characterised by a network of surface rivers and frequent flooding, this induces a very particular response at the Unica springs with very high hydrological variability (by several orders of magnitude), as well as delayed and prolonged high-flow values (Mayaud et al., 2019). Low-flow periods are sustained by flows from the karstified limestone aquifer, which reaches heights up to 1800 m asl and has significant groundwater storage (Ravbar et al., 2012). Part of the discharge is lost due to an underground bifurcation (Kogovšek et al., 1999). When the discharge exceeds about 60 $m^3\ s^{-1}$

and remains high for a few days, a polje downstream of the springs becomes flooded. When the discharge reaches about 80 m^3 s^{-1}, the flooding reaches the monitoring station, influencing the measurement. The water from the lake is drained by several ponors downstream of the monitoring station, but their absorption capacity is much lower than the discharges of the springs.

Precipitation, snow cover height, and height of new snow data were obtained from two meteorological stations located on the catchment. Temperature and relative humidity data are from Postojna meteorological station only. Potential evapotranspiration is calculated using data from the Postojna station with the modified Penman-Monteith approach (Allen et al., 1998).

6.3 Methodology

6.3.1 Artificial neural networks

ANN are a branch of machine learning, i.e. a technique to learn complex relations from existing data. They imitate the basic functioning of biological nervous systems and similarly consist of mathematical representations of neurons structured and interconnected in layers. Given sufficient data from which to learn, ANN can establish complex input-output relations with only limited domain knowledge.

In this study, CNNs (LeCun et al., 2015) – a specific model type from the ANN-subfield of deep learning (DL) – were used. CNNs are predominantly successful in processing image-alike data, but are also useful in signal processing for sequential data. They usually consist of sequences or blocks of convolutional layers for feature recognition and pooling layers for information consolidation. In the former, filters of a specific size (defining their receptive field) are used to produce feature maps. These feature maps are subsequently downsampled (often by maximum selection) into pooling layers to consolidate the contained information. Several of these blocks with varying properties can be stacked on top of each other, also in combination with other layer types such as batch normalization layers (Ioffe and Szegedy, 2015) to prevent exploding gradients or dropout layers (Srivastava et al., 2014). Lastly, one (or multiple) fully connected dense layers follow to produce the model output. For the models in this study, we used a single one-dimensional convolutional layer with a fixed kernel size (three) and an optimised number of filters. This layer was succeeded by (i) a Max-Pooling layer, (ii) a Monte Carlo dropout layer (10 % dropout rate), and (iii) two dense layers: the first with an optimised number of neurons and the second with a single output neuron (Figure 6.2). Besides number of filters and number of neurons in the first dense layers, we optimised the training batch size and the length of the input sequence for each simulation step using the Bayesian Optimization library (Nogueira, 2014). The number of minimum and maximum optimisation steps was individually selected for each site and can be found in the provided modelling scripts (Cinkus and Wunsch, 2022). To ensure proper learning, the models are regularised with several measures. Hence, early stopping with a patience of 20 steps is applied to prevent the model from overfitting. Except for Qachqouch, where few data are available, the size of the according stopset ranges between one and four annual cycles (see the provided scripts for details). This stopset is considered a part of the calibration period mentioned in section 6.3.4. Further, dropout ensures robust training and serves as another measure against overfitting. We applied the Adam optimizer for a maximum of 150 to 300 training epochs with an initial learning rate of 0.001 and applied gradient

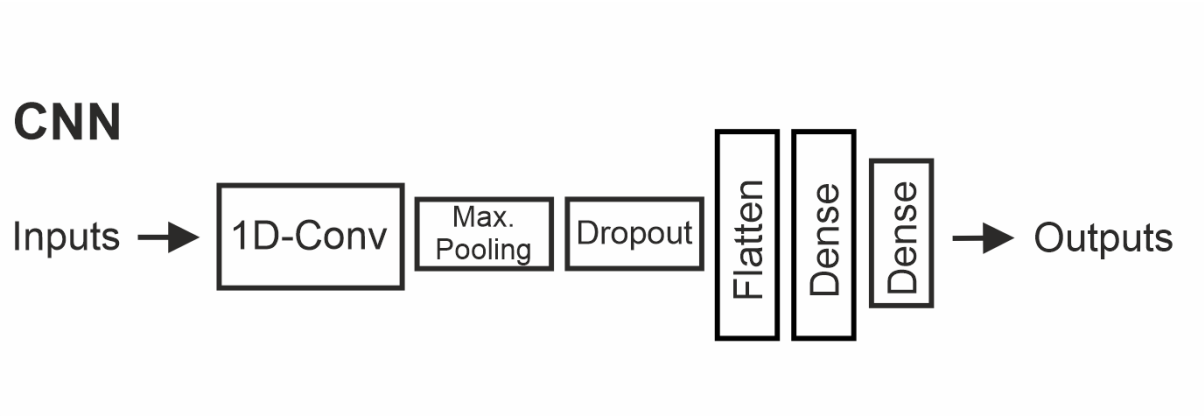


Figure 6.2: Selected structure of the CNN model.

clipping to prevent exploding gradients.

6.3.2 Reservoir models

Reservoir models are a conceptual representation of a hydrosystem, which involves the association of several reservoirs that are thought to be representative of the main processes at stake. Each reservoir is characterised by its water height and a flow equation that translates the variations of water height into discharges. The flow equation is function of a specific discharge coefficient and a positive exponent (different from 1 for non-linear flows), which are defined by calibration against observed data.

Many reservoir models have been developed to study the relation between precipitation and discharge in karst systems (Hartmann et al., 2014). They all differ in complexity with respect to the number of reservoirs and parameters, which need to be well thought out in order to preserve physical realism and limit equifinality on model parameters. Careful sensitivity analyses and uncertainty assessment should be considered along with model results to avoid over-interpretation (Refsgaard et al., 2007). Reservoir models can be seen as a compromise between simulation performance and insight into the functioning of a system.

We used the adjustable modelling platform KarstMod to perform reservoir modelling. KarstMod provides a modular, user-friendly interface for simulating spring discharge at karst outlets. The structure of models built using the KarstMod platform is based on the conceptual model of a karst aquifer with infiltration and saturated zones (Mazzilli et al., 2019). The infiltration zone (soil and epikarst) drains water from the surface through a vertical network of fissures and conduits. Water storage can occur in the unsaturated zone, as well as local saturation. The saturated zone comprises a dual porosity functioning, with a network of high-permeability fractures and conduits, and a low-permeability matrix with a high storage capacity.

In KarstMod, the model structure can include up to four reservoirs. One at the upper level reflects the processes (infiltration, storage and drainage) occurring in the soil and epikarst zone. Three at the lower level can be connected with the first one and correspond to the infiltration and/or saturated zones. The discharge can be simulated with (i) several linear and non-linear water level-discharge laws, (ii) a hysteretic water level-discharge function to reproduce the hysteretic functioning observed on the wet-dry cycles in the unsaturated zone (Lehmann et al., 1998; Tritz et al., 2011), and (iii) an exchange function

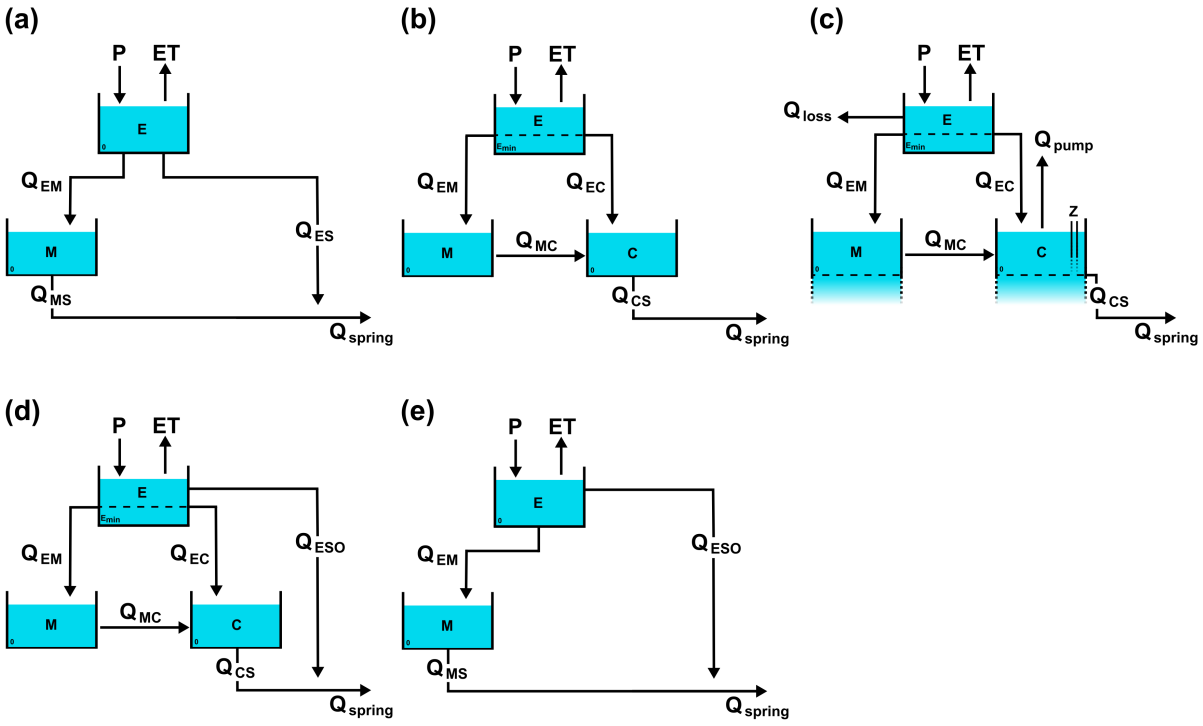


Figure 6.3: Selected model structures for (a) Aubach, (b) Gato Cave, (c) Lez, (d) Qachqouch and (e) Unica springs. Flux names correspond to the terminology of the KarstMod platform (Mazzilli and Bertin, 2019).

that aims to reproduce the interactions between the matrix and conduits. More details on the balance equations, the parameters involved and the KarstMod platform in general can be found in (Mazzilli et al., 2019) or in the KarstMod User Guide (Mazzilli and Bertin, 2019).

In this study, we first addressed the structure of the models taking into account our expert knowledge and previous studies. For each site, we examined the major characteristics that determine the functioning of the system and associated the corresponding conceptual modelling. We then modified this base structure according to the performance of the model while trying to maintain physical realism. The most efficient model structures that we obtained after performing the modelling are shown in Figure 6.3.

Aubach spring selected model (Figure 6.3a) is close to the conceptual model with a very reactive transfer function (Q_{ES}), corresponding to the well-developed conduit network, and a matrix reservoir (M), which in this case mostly reflects the storage properties in the unsaturated limestone. We tested different configurations (lost discharge from upper level reservoir and/or pumping in lower reservoirs) to simulate the lost discharges through overflow springs and underground flows, but there were no significant increases in model performance. Gato Cave spring selected model (Figure 6.3b) is different from the conceptual model as the platform could not account for the allochthonous recharge on the catchment. The model structure includes a soil available water capacity (E_{min}), matrix and conduits compartments (M and C), as well as matrix-conduits exchanges (Q_{MC}), which may translate the processes occurring through the dam. Lez spring selected model (Figure 6.3c) is accurate with the conceptual model and includes an overflow transfer function (Q_{loss}), matrix and conduit compartments (M and C), matrix-conduit exchanges (Q_{MC}), and pumping into the main conduit (Q_{pump}). We considered a low soil available

water capacity (E_{min}) as it greatly increased the performance of the model. Qachqouch spring selected model (Figure 6.3d) is consistent with previous conceptual models that considered many different response times. The model structure features a very reactive transfer function (Q_{ESO}), matrix and conduits compartments (M and C), matrix-conduits exchanges (Q_{MC}) as well as a soil available water capacity. The multiple different transfer functions help to reproduce the reactive and dampened responses of the Qachqouch karst aquifer. Unica springs selected model (Figure 6.3e) is significantly simpler than the conceptual model, which includes polje flooding, allochthonous recharge, overflow springs and matrix-conduits exchanges. We only retained a very simple structure as it was the best trade-off between physical realism and model performance. The very reactive transfer function Q_{ESO} allows reproduction of fast flows through conduits, while the matrix reservoir (M) likely translates processes occurring in the matrix and surface flooding.

6.3.3 Input data

Input data are the time series that are used for simulating karst spring discharge. They can be derived from either a direct observation (e.g. observed discharge, temperature, sinking stream discharge or pumping) or a calculation from raw input data (e.g. potential evapotranspiration derived from temperature). The nature of input data usually differs between ANN and reservoir modelling approaches, as ANN models tend to make good use of direct observations, whereas reservoir models often require one to preprocess the raw input data. We decided to work with raw input data to ensure equitable performance between ANN and reservoir models. The raw input data was either used directly or preprocessed, depending on the modelling approach.

The data used for each modelling approach and site is summarised in Table 6.2. Observed discharge time series were used directly (without further preprocessing) in ANN and reservoir models. In the case of the Lez spring, the models were simultaneously calibrated on the spring discharge (Q) as well as on the water level in the aquifer (Z). Furthermore, the pumped discharge time series in reservoir C (Q_{pump}, Figure 6.3c) was used as an input. Precipitation time series were used differently as there are often several meteorological station per study site. For ANN models, precipitation time series were used as raw input P_{raw}, except for Lez spring where the individual raw precipitation data had too many missing values, so we used the same input as the reservoir model (P_{in}). In the case of Aubach, Qachqouch and Unica, P_{raw} includes all the precipitation time series from the different meteorological stations (Appendix 6.A). For reservoir models, the precipitation time series were either (i) used directly if there was no snow dynamics in the catchment and only one meteorological station was available (Gato Cave), (ii) preprocessed with Thiessen's polygon interpolation (Appendix 6.B) if there were several meteorological stations (Lez), (iii) preprocessed with a snow routine (Appendix 6.C) to simulate snow accumulation and melting over the catchment (Aubach) if snow dynamics could not be neglected, or (iv) preprocessed with both Thiessen's polygon interpolation and a snow routine (Qachqouch, Unica). For reservoir models, evapotranspiration processes were considered using time series of potential evapotranspiration, which were calculated for each site using different methods depending on the available meteorological data, the climate of the area and local expert knowledge. For ANN models, we used temperature time series instead of evapotranspiration because calculating potential evapotranspiration is generally not necessarily beforehand. Additionally, we used a sinusoidal temperature signal time series (T_{sin}, derived from the observed temperature) to better

Spring	Time step	Date range	Data used		Maximum gap [days]			
			ANN	Reservoir	P	T	Q	ET
Aubach	Hourly	2012–2020	Q_{obs}, P_{raw}, T, T_{sin}[a]	Q_{obs}, P_{sr}, PET[c]	0	0	0	0
Gato Cave	Daily	1963–2015	Q_{obs}, P_{raw}, T_{max}, T_{min}, T_{med}	Q_{obs}, P_{raw}, PET	0	0	0	0
Lez	Daily	2008–2020	Q_{obs}, Q_{pump}, Z_{obs}, P_{in}, T_{sin}	Q_{obs}, Q_{pump}, Z_{obs}, P_{in}, PET	0	2	7	0
Qachqouch	Daily	2015–2020	Q_{obs}, P_{raw}, T_{max}[b]	Q_{obs}, P_{in}, PET	0	0	11	0
Unica	Daily	1961–2018	Q_{obs}, P_{raw}, T, T_{sin}, NS	Q_{obs}, $P_{in\text{-}sr}$, PET	0	1	0	29

[a] P_{raw}, T and T_{sin} data are from Diedamskopf, Oberstdorf and Walmendinger Horn meteorological stations
[b] T_{max} data are from the 1700 m meteorological station
[c] P_{sr} data are calculated with the data from Diedamskopf station

Table 6.2: Summary of input data. (i) P_{raw}, (ii) P_{in} and (iii) P_{sr} refer to (i) raw precipitation data, (ii) precipitation data interpolated with Thiessen's polygon method, and (iii) precipitation data redistributed by applying the snow routine. Q_{obs}, Z_{obs} and T refer to observed discharge, observed water level and temperature, respectively. ET (Evapotranspiration) refers to either PET (Potential Evapotranspiration) or AET (Actual Evapotranspiration) time series.

account for seasonality in Aubach, Lez and Unica ANN models.

We handled missing values in the different time series as follows: (i) temperature gaps were linearly interpolated, (ii) precipitation and evapotranspiration gaps were considered to be equal to 0, and (iii) discharge gaps were interpolated with a Lagrange polynomial function. Maximum observed gaps for precipitation, temperature, discharge and evapotranspiration time series are detailed in Table 6.2. Note that (i) for Lez spring, we observed maximum gaps of 17 and 16 days for pumped discharge and piezometric level, respectively; and (ii) for Unica springs, there are no missing values in the Cerknica new snow height (NS) time series.

6.3.4 Model calibration and simulation

The calibration period is the period used for selecting the parameters that provide the best results according to the performance measure. The validation period is intended to assess the relevance of the parameters over a time interval that is not used for calibration. In the domain of the ANN modelling, the validation is usually denoted as testing period. However, we unify the terminology at this point. The calibration period is again split into three different parts in the case of ANN modelling, (i) to train the model, (ii) to prevent the model from overfitting, and (iii) to optimize its hyperparameters. We defined the same calibration and validation periods for both modelling approaches (Table 6.3), which ensures fair initial conditions and a meaningful comparison of the results. We have chosen the periods in a way to maximise the length of the calibration periods, which allows for relevant model results (especially in ANN models). In reservoir model, we considered a short warm-up interval at the beginning of the calibration period for the model to adjust and reach an optimal state.

We calibrated the models by applying the mean squared error (MSE) to simulated and observed discharge time series. For Lez spring, we used a composite function of discharge and water level in order to consider both variables in the same modelling process.

Multiple simulations were carried out for each modelling approach at each site. The obtained simulated discharge (or water level) time series corresponds to the mean of the distribution of simulated values at each time step. We also considered the uncertainties in the model prediction by calculating the 90 % confidence interval, whose limits correspond to the 5^{th} and 95^{th} percentiles of the distribution at each time step.

Spring	Calibration period	Validation period	Objective function
Aubach	2014-04-18 – 2019-12-31	2020-01-01 – 2020-10-31	MSE(Q)
Gato Cave	1963-10-02 – 2011-09-03	2011-09-04 – 2015-04-29	MSE(Q)
Lez	2008-10-21 – 2017-12-31	2018-01-01 – 2020-12-30	MSE(Q,Z)
Qachqouch	2015-09-06 – 2019-09-30	2019-10-01 – 2020-01-22	MSE(Q)
Unica	1961-01-01 – 2016-09-28	2016-09-29 – 2018-12-31	MSE(Q)

Table 6.3: Calibration and validation periods.

In KarstMod (reservoir models), the retained simulations correspond to all the results satisfying a maximum MSE threshold on the calibration period for a 6-hour model run. The confidence interval reflects the uncertainty in the parameters used in the model, which are not fixed but are defined as a range (e.g. catchment area = 150 to 200 km^2). In the case of ANN models, we used a model ensemble of 10 models based on different random number generator seeds for model initialization. Using the Monte Carlo dropout layer, for each of the ensemble members a total of 100 simulation results were generated.

6.3.5 Model evaluation

We evaluated the performance of the models using the MSE and several performance criteria that assess different aspects of the discharge: modified Kling-Gupta efficiency (KGE'), KGE' components (r, γ, β) (Kling et al., 2012), and Diagnostic efficiency (DE) (Schwemmle et al., 2021). These criteria were all applied to the whole discharge regime, but also to sub-regimes of high- and low-flow conditions (with the exception of DE, which already takes sub-regimes into account). For Lez spring, we also applied the MSE and KGE' criteria on water level. Model performance is usually evaluated on both calibration and validation periods for reservoir models. However, this approach is not suited to ANN models, for which the calibration period corresponds to the learning period of the model. Thus, we chose to only evaluate and compare the reservoir and ANN models on their validation periods.

The Kling-Gupta efficiency (KGE) has gained in popularity as it aims to address some limitations of the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970), i.e. (i) the discharge variability is underestimated, (ii) the mean of observed values is not a meaningful benchmark for variables with high variability, and (iii) the normalised bias is dependent on the variability (Gupta et al., 2009; Willmott et al., 2012). The KGE' is based on the KGE and aims to ensure that bias and variability are not cross-correlated by using the coefficient of variation ratio (γ) instead of the standard deviation ratio (α):

$$KGE' = 1 - \sqrt{(\gamma - 1)^2 + (\beta - 1)^2 + (r - 1)^2} \tag{6.1}$$

with r the Pearson correlation coefficient between the simulated and observed discharge, β the ratio between mean simulated and mean observed discharge, and γ the ratio between simulated and observed coefficients of variation of discharge. The three components of KGE' help to evaluate different aspects of a model: (i) r is related to shape and timing (Santos et al., 2018), (ii) β is used to assess the overall volume of water discharged at the spring (further referred to as "volume"), and (iii) γ gives insight into the flow variability. The KGE' and r criteria can range from $-\infty$ to 1, whereas γ and β can range from $-\infty$ to $+\infty$. A KGE' score equal to 1 means a perfect match between

simulated and observed discharge, while a score lower than -0.41 indicates that the model is worse than using the observed mean as a predictor (Knoben et al., 2019).

The DE criterion is intended to help define the weaknesses of a model. It is based on constant, dynamic and timing errors. DE is proposed as a numerical measure (ranging from 0 to $+\infty$, with 0 indicating a perfect model), but can also be visualised on a polar plot that effectively differentiates error contributions. The overall error is calculated with the following equation:

$$DE = \sqrt{\overline{B_{rel}}^2 + |B_{area}|^2 + (r-1)^2} \tag{6.2}$$

with $\overline{B_{rel}}$ and $|B_{area}|$ the measures for constant and dynamic errors, respectively. As these measures are based on the flow duration curve, they give information in terms of exceedance probability. Details of their calculation can be found in (Schwemmle et al., 2021).

The performance criteria applied to high- and low-flow conditions are denoted by the lower script indices "L" and "H", respectively. These criteria allow the performance of the models to be evaluated over different flow regimes (i.e. dry/intermediate, wet). Discharge thresholds were set manually based on our knowledge of the system and a careful assessment of the distribution of discharge values. They are equal to 1, 2, 0.8, 5, and 20 m^3 s^{-1} for Aubach, Gato Cave, Lez, Qachqouch and Unica springs, respectively.

6.4 Results and discussion

The obtained models and their confidence intervals for the two approaches and each test site are presented in Figure 6.4 for discharge and Figure 6.5 for water level (Lez spring). Their performance – assessed with multiple criteria – are shown in Figure 6.6, Table 6.4, and Appendix 6.D. The DE polar plots for each site are presented in Figure 6.7.

6.4.1 Modelling results

6.4.1.a Aubach spring

ANN model is very good with a KGE' of 0.88 (Table 6.4). The snow-influenced period from April to mid-June is accurately modelled, as are the peaks in summer and early autumn (Figure 6.4a). The highest peaks of the whole time series occurring in February, July and November are only slightly underestimated. The model is balanced and accurate on volume ($\beta = 0.93$), variability ($\gamma = 1.01$) and shape and timing ($r = 0.91$). The model is very good for simulating high flows and is decent on low flows, but could be improved, especially on shape and timing ($r_H = 0.84$, $r_L = 0.66$). The slightly higher value of γ_L (1.26) may be related to the tendency of the model to "oscillate" during low/medium flows (e.g. in September, Appendix 6.Ea). This wave-like behaviour may be related to a high sensitivity to precipitation events or to inappropriate learning from other data. DE is very good (0.14, Figure 6.7a). The model shows negative dynamic and constant errors with a higher share of high flows, which points a small underestimation of the occurrence of high flows.

Reservoir model is decent with a KGE' of 0.69 (Table 6.4), but the model fails to accurately reproduce the discharges in all seasons. There is a deficit in water during winter/early spring and an excess during spring (Figure 6.4a). The model is balanced

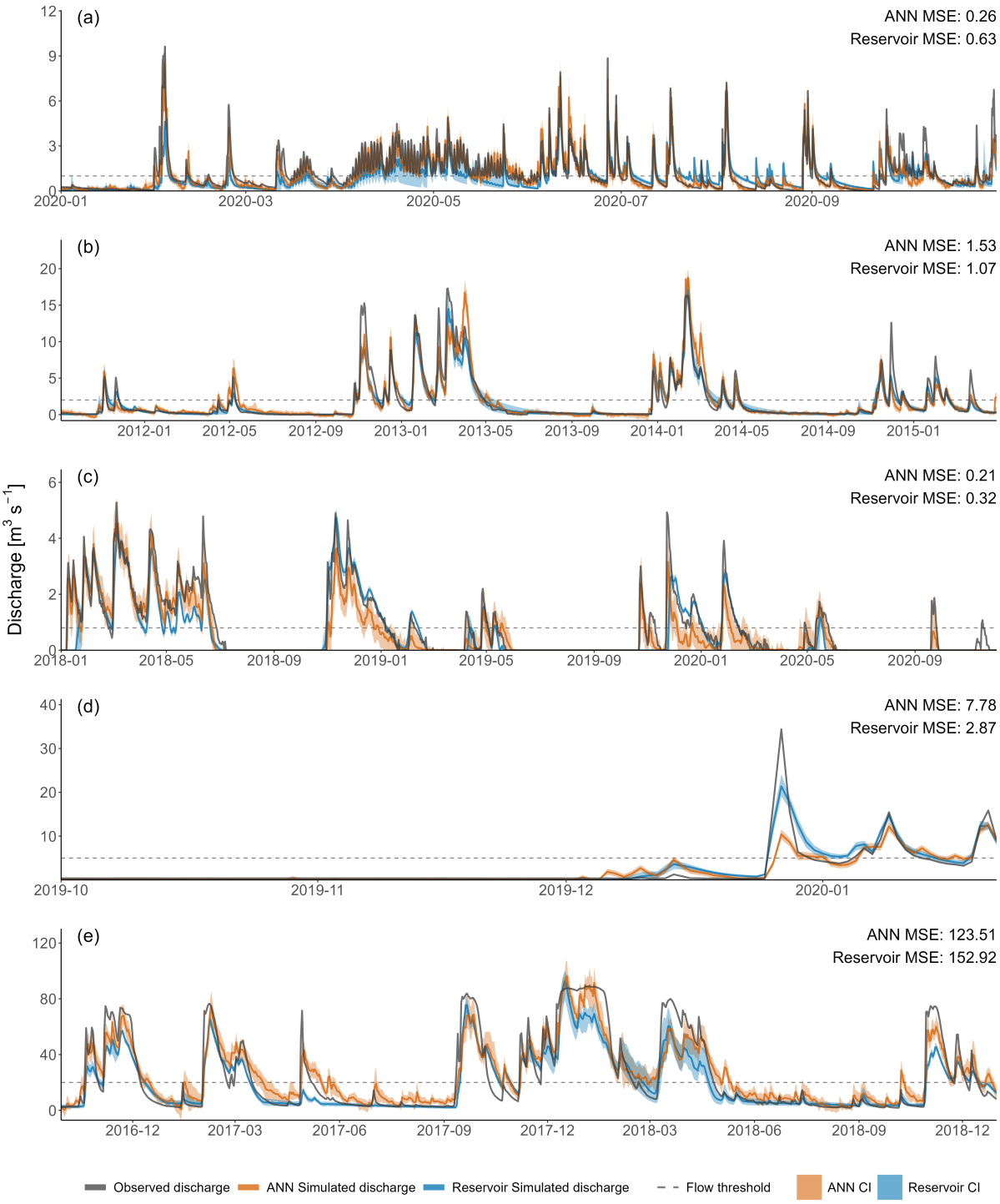


Figure 6.4: Observed and simulated spring discharge time series with (i) 90 % confidence intervals (CI) on the validation period and (ii) the threshold for high and low flows used for the calculation of the performance criteria. (a) Aubach, (b) Gato Cave, (c) Lez, (d) Qachqouch and (e) Unica springs.

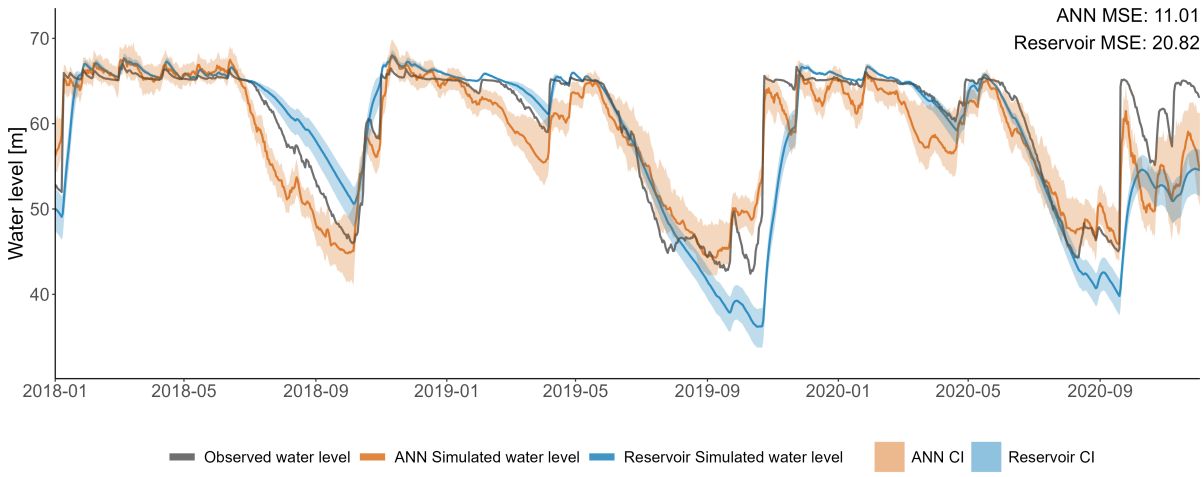


Figure 6.5: Observed and simulated spring water level time series with 90 % confidence intervals (CI) on the validation period (Lez spring).

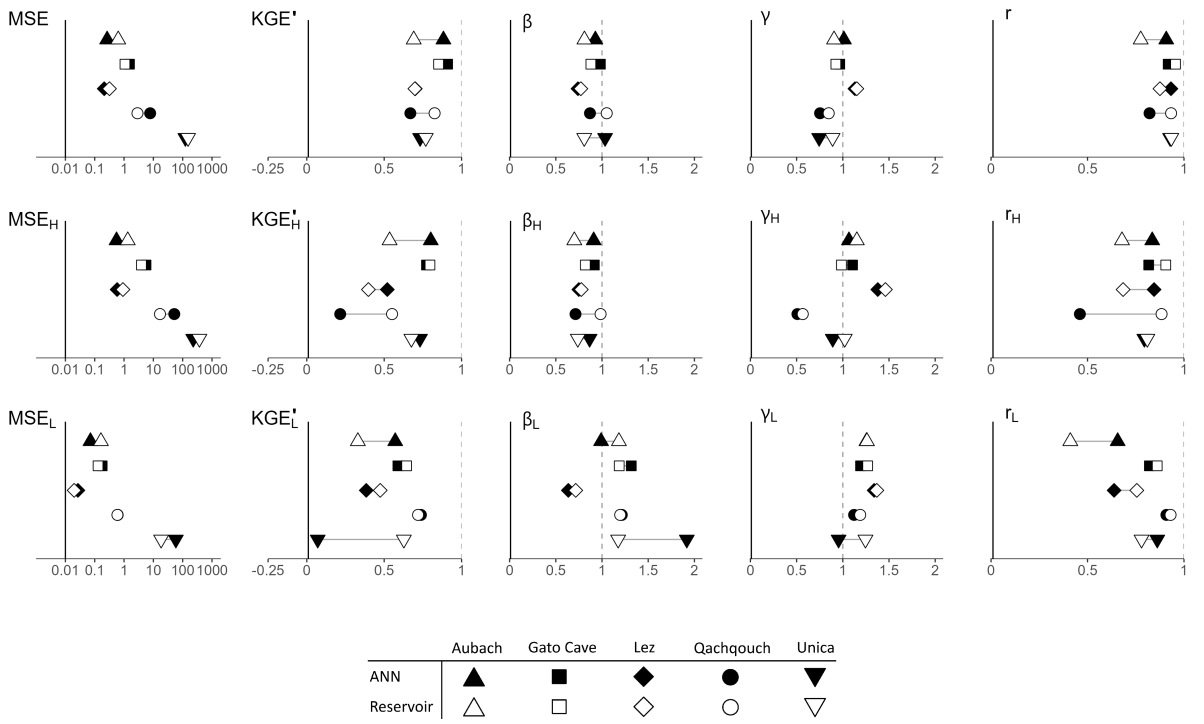


Figure 6.6: Performance of the ANN and reservoir models on the validation period, according to different performance criteria. Exact values can be found in Table 6.4.

Spring	Flow regime	MSE		KGE'		β		r		γ	
		ANN	RSV	ANN	RSV	ANN	RSV	ANN	RSV	ANN	RSV
Aubach	Total	**0.26**	0.63	**0.88**	0.69	**0.93**	0.81	**0.91**	0.78	**1.01**	0.91
Aubach	High flow	**0.55**	1.34	**0.80**	0.53	**0.91**	0.70	**0.84**	0.68	**1.07**	1.15
Aubach	Low flow	**0.07**	0.16	**0.57**	0.33	**0.99**	1.18	**0.66**	0.41	1.26	1.26
Gato Cave	Total	1.53	**1.07**	**0.91**	0.85	**0.98**	0.88	0.92	**0.96**	**0.97**	0.92
Gato Cave	High flow	5.62	**3.93**	0.77	**0.80**	**0.92**	0.82	0.82	**0.91**	1.11	**0.99**
Gato Cave	Low flow	0.18	**0.13**	0.59	**0.64**	1.32	**1.19**	0.82	**0.86**	**1.19**	1.27
Lez	Total	**0.21**	0.32	0.70	0.70	0.74	**0.77**	**0.93**	0.88	**1.13**	1.15
Lez	High flow	**0.58**	0.92	**0.52**	0.40	0.75	**0.78**	**0.84**	0.68	**1.38**	1.46
Lez	Low flow	0.03	**0.02**	0.38	**0.47**	0.64	**0.72**	0.64	**0.76**	**1.34**	1.39
Qachqouch	Total	7.78	**2.87**	0.67	**0.83**	0.87	**1.05**	0.82	**0.93**	0.75	**0.85**
Qachqouch	High flow	51.74	**16.84**	0.22	**0.55**	0.71	**0.98**	0.46	**0.88**	0.51	**0.57**
Qachqouch	Low flow	0.61	**0.60**	**0.74**	0.72	1.21	**1.20**	0.91	**0.93**	**1.12**	1.19
Unica	Total	**123.51**	152.92	0.73	**0.77**	**1.03**	0.81	0.93	0.93	0.74	**0.89**
Unica	High flow	**229.00**	371.37	**0.73**	0.68	**0.87**	0.74	0.79	**0.81**	0.89	**1.02**
Unica	Low flow	58.55	**18.43**	0.07	**0.63**	1.92	**1.17**	**0.86**	0.78	**0.95**	1.25

Table 6.4: Details of indicator values for the reservoir (RSV) and ANN models in the validation period. For each site, the simulations are evaluated with different performance criteria on total, high- and low-flow conditions. Values in bold correspond to the better score between ANN and reservoir models.

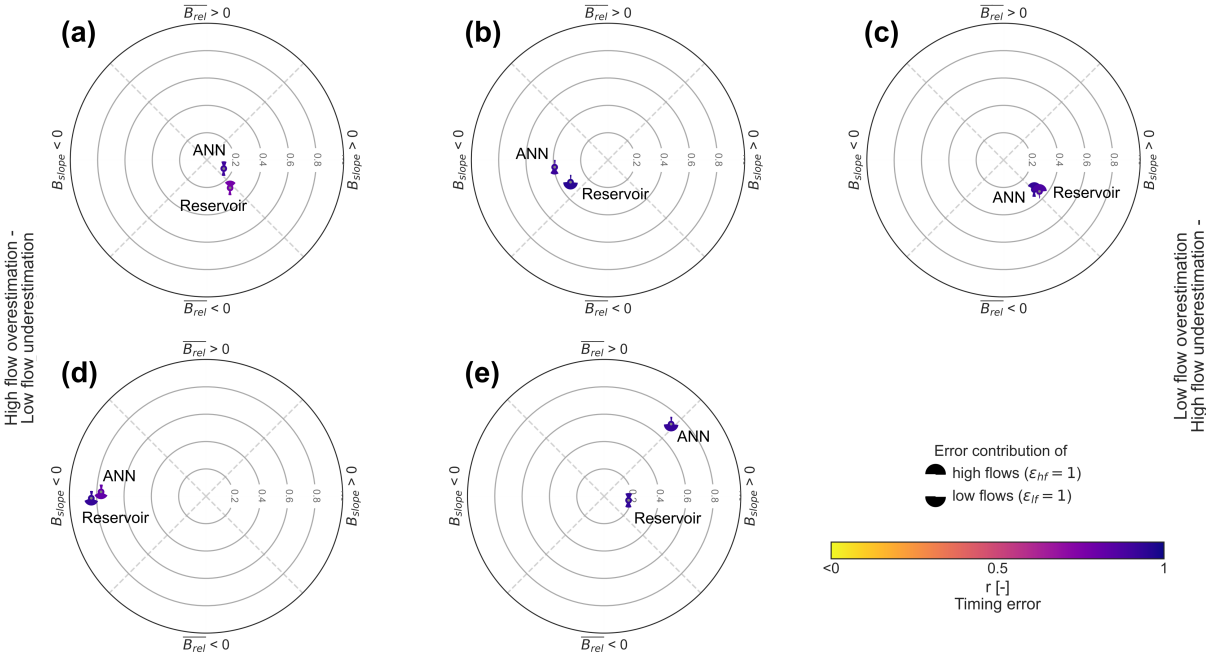


Figure 6.7: Diagnostic efficiency polar plots on the validation period. (a) Aubach, (b) Gato Cave, (c) Lez, (d) Qachqouch and (e) Unica springs.

and accurate on variability ($\gamma = 0.91$), but has middling shape and timing ($r = 0.78$) and underestimates volume ($\beta = 0.81$). The model is particularly bad for simulating low flows, with high errors on volume ($\beta_L = 1.18$), variability ($\gamma_L = 1.26$) and shape and timing ($r_L = 0.41$). The simulated high flows are decent, although they can be improved on shape and timing ($r_H = 0.68$) and volume ($\beta_H = 0.70$). DE is good (0.26, Figure 6.7a). The model has negative dynamic and constant errors, with a higher share of high flows. These errors can be either due to (i) a miscalibration of the snow routine, retaining too much water as snow in winter and thus releasing too much in warmer periods, (ii) the uncertainties related to the meteorological data in mountainous catchments, or (iii) the snow dynamics which cannot be taken into account within the KarstMod platform, e.g. by adding a snow storage above the epikarst (Chen et al., 2018).

In October, a series of peaks is not well captured by the outputs of both models (Figure 6.4a). A plausible explanation is that the inputs do not capture the respective local precipitation events due to the location of the climate stations outside the catchment.

The modelling of discharges from Aubach spring is challenging due to the large elevation differences and the heterogeneity of precipitation on the catchment. This makes it difficult to provide accurate data to the model, especially with regard to snow dynamics. The reservoir model is particularly confronted with these aspects because (i) it can only handle a single precipitation input (from one weather station or interpolated from several stations) and (ii) the snow dynamics must be simulated by a snow module. As these preprocessings cannot really catch the spatial heterogeneity of complex snow processes, they strongly limit the model performance (Çallı et al., 2022). Leaving aside the mismatches related to inadequate meteorological inputs, the structure of the reservoir model seems appropriate to simulate the hydrological response of the spring. In contrast, the ANN model is able to consider snow dynamics without any preprocessing, using only the precipitation and temperature time series during calibration. It shows great versatility with respect to the input data, similar to that of a two-dimensional approach.

6.4.1.b Gato Cave spring

ANN model is very good with a KGE' of 0.91 (Table 6.4), but the model struggles to reproduce the discharges during flood events (Figure 6.4b). Very high peaks are either overestimated (e.g. May 2012, April 2013, March 2014) or underestimated (e.g. December 2011, November 2012, March 2013). The model is balanced and accurate on volume ($\beta = 0.98$), variability ($\gamma = 0.97$) and shape and timing ($r = 0.92$). The model is good for simulating high flows and is somewhat decent on low flows. It shows a slight lack on shape and timing on both high and low flows ($r_H = 0.82$, $r_L = 0.82$), and also seems to overestimate low flows ($\beta_L = 1.32$). In the same way as Aubach ANN model, the slightly high variability ($\gamma_L = 1.19$) may be related to the "oscillations" that can be observed, especially on medium and low flows (e.g. January 2012, May 2013, Appendix 6.Eb).

Reservoir model is very good with a KGE' of 0.85 (Table 6.4), although the model tends to slightly underestimate the discharges during high-flow events ($\beta_H = 0.82$, Figure 6.4b). This seems to happen when precipitation occur during several days without reaching really high values, which may indicate either (i) some kind of hysteresis functioning with flow occurring after a connection has been made in the system, or (ii) inflows into the system that are not taken into account in the model. The model is balanced and accurate on variability ($\gamma = 0.92$) and shape and timing ($r = 0.96$), but generally un-

derestimates volume ($\beta = 0.88$). The model has good performance on high flows and is decent on low flows. After flood periods, the model seems to simulate a slower draining than observed – higher volume ($\beta_L = 1.19$) and variability ($\gamma_L = 1.27$) of low/medium flows – resulting in inaccurate recession periods for which the discharge is overestimated (e.g. November 2011, January 2013, April 2014).

Some periods like November 2012 or February 2015 are not simulated very well by both models (Figure 6.4b), which may be related to uncertainties in the meteorological data input. DE is decent (0.39) for ANN model and good (0.32) for reservoir model (Figure 6.7b). Both models have a positive dynamic error with a higher share of low flows, which highlight a small underestimation of the occurrence of low flows.

The modelling of discharges from Gato Cave spring shows that both approaches can have great performance given few modelling constraints. Raw precipitation input was used in both models, thereby avoiding additional uncertainties from interpolation or snow preprocessings.

6.4.1.c Lez spring

ANN model is good with a KGE' of 0.7 for discharge (Table 6.4) and 0.89 for water level (Figure 6.5). The high piezometric levels (above 55 m asl) seem a bit too sensitive to precipitation events, especially at the end of 2019 (Figure 6.5). On discharges, the model is accurate on variability ($\gamma = 1.13$) and shape and timing ($r = 0.93$), but underestimates volume ($\beta = 0.74$). The overall underestimation of volume mainly comes from high flows ($\beta_H = 0.75$) as they are the most represented on the time series. The model is decent on high flows, although having too much variability ($\gamma_H = 1.38$). On low flows, the model performs poorly mainly due to high underestimation of volume ($\beta_L = 0.64$) and insufficient shape and timing ($r_L = 0.64$).

Reservoir model is good with a KGE' of 0.7 for discharge (Table 6.4) and 0.75 for water level (Figure 6.5). However, the model fails to reproduce the observed discharge for several months for the period between September 2020 and March 2020 (Figure 6.4c). During dry periods, there is a too high deficit in the lower reservoirs, leading to a strong delay in the spring response when fresh precipitation occur – the C reservoir having to be replenished beforehand. The model is balanced and accurate on shape and timing ($r = 0.88$), but overestimates variability ($\gamma = 1.15$) and underestimates volume ($\beta = 0.77$). The model is decent on high flows, but has poor variability ($\gamma_H = 1.46$) and shape and timing ($r_H = 0.68$), and also underestimates volume ($\beta_H = 0.78$). On low flows, the model has too much variability ($\gamma_L = 1.37$) and middling shape and timing ($r_L = 0.76$). The piezometric levels are satisfactory when the spring is flowing (greater than 65 m asl), but lose accuracy during dry periods. The model could not reproduce the changes in flow dynamics at 46 m asl (August 2019, August 2020, Figure 6.5). Also, the draining of the aquifer seems to be simulated more slowly than observed (July 2018, July 2019), which can be a result of the model trying to fit the aforementioned periods during calibration.

On both models, the poor overall KGE' value on low/medium flows is probably due to the small occurrences of low discharges (except 0), thus inducing a high error on volume. DE is good for both ANN (0.31) and reservoir models (0.35) (Figure 6.7c). Both models have negative dynamic and constant errors with a higher share of high flows, which highlight an underestimation of the occurrence of high flows.

The time series is generally characterised by distinct dry periods without any recharge due to the anthropogenic pumping of water into the saturated zone of the aquifer. These

periods are simulated fairly accurately by both models but ANN model is better at simulating first floods after or during dry periods. Several boreholes at the north of the spring showed flow-bearing structures at 50 m asl (Dausse et al., 2019). These fast water transfers could explain the rapid increases in observed piezometric level and the reactive spring responses. We also suspect an evolution of the carbonate's facies with depth, which could affect the effective porosity of the medium and induce different flow dynamics. These aspects are not considered in the reservoir model which results in poor simulations when the water level is below 60 m asl. However, this failure provides information on the structure of the aquifer, which is valuable for research. On the other hand, ANN model was successful in learning these particular behaviours, especially as it only had a medium learning time of about 8 years.

6.4.1.d Qachqouch spring

ANN model is decent with a KGE' of 0.67 (Table 6.4), but strongly overestimates low flows at the beginning of December and then underestimates the flood peak at the end of the month (Figure 6.4d). The model slightly underestimates volume ($\beta = 0.87$), and is lacking in variability ($\gamma = 0.75$) and shape and timing ($r = 0.82$). The high flows are poorly simulated but the low flows are well fitted, although volume is slightly overestimated ($\beta_L = 1.21$). The oscillations of the simulated discharges (Appendix 6.Ec) may appear because the model does not account the time needed for the aquifer to replenish and generate an increase of discharge at the spring.

Reservoir model is very good with a KGE' of 0.83 (Table 6.4). At the end of the dry period, the low flows are overestimated and the first flood is underestimated (Figure 6.4d). This may be due to heterogeneous precipitation occurring on highly transmissive parts of the catchment. In this case, the soil available water capacity (E_{min}) – which is necessary for a good simulation of low-flow periods – may not be representative of the whole catchment, thus inducing a more dampened response than observed. The model is balanced and accurate on volume ($\beta = 1.05$) and shape and timing ($r = 0.93$), but slightly lacks in variability ($\gamma = 0.85$). The model is decent on high flows but has middling variability ($\gamma_H = 0.57$) which can be due to the underestimation of the late December flood peak. The low flows and recession periods are overestimated ($\beta_L = 1.20$ and $\gamma_L = 1.19$).

DE is bad for ANN and reservoir models (0.77 and 0.84, respectively, Figure 6.7d). Both models have a positive dynamic error with a higher share of low flows, which highlight an underestimation of the occurrence of low flows. Here, the positive dynamic error is influenced by the constant underestimation of the observed discharge during the dry period (October–December 2019), accounting for more than 50 % of the observations.

The very short data length is particularly detrimental to the ANN model as the learning period is only about 4 years. Furthermore, even when data are available, there is a significant amount of time without (relevant) discharge, for which no input-output relation can be learned. Due to the characteristics of the discharge time series, it can be assumed that a much longer time series of daily values would be needed to successfully simulate the discharges of Qachqouch spring. On the other hand, the reservoir model seems more appropriate to simulate Qachqouch spring discharges even with the limited data available. This highlights the strength of conceptual modelling to take into account recharge processes and reservoir replenishment, even on a short dataset with long dry periods.

6.4.1.e Unica springs

ANN model is good with a KGE' of 0.73 (Table 6.4). The model is accurate on volume ($\beta = 1.03$) and shape and timing ($r = 0.93$), but insufficient on variability ($\gamma = 0.74$). The model is good at simulating high flows, although slightly lacking in volume ($\beta_H = 0.87$), variability ($\gamma_H = 0.89$) and shape and timing ($r_H = 0.79$). The model is poor for simulating low flows, which are often significantly overestimated ($\beta_L = 1.92$), especially the recession periods which systematically have a slower draining (Figure 6.4e). The overestimation of low flows could be the result of the model trying to better fit the high-flow periods during training, which may shift the whole discharge curve slightly towards the upper limits. The model also seems to be too sensitive regarding precipitation events, hence the wave-like behaviour of the simulated time series (Appendix 6.Ed). DE is bad (0.72, Figure 6.7e). The model has a negative dynamic error and a positive constant error with a higher share of low flows, which highlights an overestimation of the occurrence of low flows.

Reservoir model is good with a KGE' of 0.77 (Table 6.4). The model is balanced and accurate on variability ($\gamma = 0.89$) and shape and timing ($r = 0.93$), but has shortcomings on volume ($\beta = 0.77$). In some winter months (December 2017, March 2018), the model has a delayed response of the rising limb (Figure 6.4e), which may be due to a slightly wrong parametrisation of the snow routine. The model is good on high flows, but shape and timing ($r_H = 0.81$) and volume ($\beta_H = 0.74$) could be improved. The model accurately simulates low flows volume, but has too much variability ($\gamma_L = 1.25$) and is middling on shape and timing ($r_L = 0.78$). The difficulty of the model to reproduce the depletion of the capacitive function may be due to the size and complexity of the catchment and the very specific influence of poljes draining over the catchment, which cannot be simulated within the KarstMod platform. DE is very good (0.18, Figure 6.7e). The model has negative dynamic and constant errors with a higher share of high flows, which highlight a small underestimation of the occurrence of low flows.

Both models were unable to reproduce the plateau-like behaviour observed at very high discharges (Figure 6.4e), which is due to the flooding of a polje at Unica springs that influences the monitoring station (Mayaud et al., 2022). They are simulated as separate peaks, which is false in terms of model accuracy but may also have some underlying conceptual truth. Only two meteorological stations were considered, which is very few for such a large catchment (820 km^2). Moreover, the major recharge area (Javorniki plateau) does not have any direct climate data available. Both models have difficulties in consistently reproducing the very particular hydrological functioning of the system (influenced by polje and surface water). ANN model is more reactive, which helps for reproducing the dynamics of high flood peaks but hinders the simulation of low flows. Reservoir model has better dynamics for medium and low flows but does not always manage to reproduce high flood peaks, which may be a consequence of the simple structure of the model.

6.4.2 Sources of uncertainties

Both ANN and reservoir models have similar trends on water volume and hydrological variability (Figure 6.6). Overall volumes are great with β ranging from 0.74 to 1.05. High-flow volumes are systematically underestimated with β_H ranging from 0.70 to 0.98. Low-flow volumes are mainly overestimated – β_L ranging from 0.99 to 1.92 – except at Lez spring with β_L of 0.64 and 0.72. Overall hydrological variability is mainly underestimated, with only Lez and Aubach springs having γ values slightly above 1. High-flow

hydrological variability does not show a distinct trend, being either overestimated or underestimated depending on the studied system. Low-flow hydrological variability is mainly overestimated with γ_L ranging from 0.95 to 1.37. These overestimations may be due to (i) improper – and generally softer – simulation of recession periods or (ii) too high sensitivity to precipitation events, especially in ANN models, inducing discharge oscillations during recession and low-flow periods. The performance on shape and timing (r) are mixed between the two approaches. They depend mainly on the system studied and the quality of the model, but also on the hydrological period considered.

These similar results between the two approaches highlight a common struggle to simulate extreme water conditions. As ANN and reservoir modelling approaches are very different, explanation must be sought in common factors to both approaches such as input data, observed data, internal/external system dynamics or the consideration of extreme events during calibration:

i. Input data: Generally, in one-dimensional modelling approaches, input data only come from at most few meteorological stations and do not accurately reflect the heterogeneity of meteorological processes on a catchment. Spatial variability of precipitation can be very high and not fully captured by meteorological stations, (i) resulting in different travel times and generating a different response at the spring (Ollivier et al., 2020), and (ii) hindering the simulation of very high flows (Hohmann et al., 2021; Pereira et al., 2014) – especially in areas where strong convective storms are frequent (Lobligeois et al., 2014). McMillan et al. (2018) suggested that uncertainties in precipitation data are about 0–10 % at point scale but can go up to 40 % when considering interpolation uncertainties. Temperature data are generally less heterogeneous than precipitation, although they can be affected by complex topography (Aalto et al., 2017). In the case of snow-covered areas, this can result in strong uncertainties on the timing of snow accumulation and melting (Zhang et al., 2016), and therefore the recharge of the aquifer. The uncertainties related to precipitation and temperature input in one-dimensional hydrological models can thus – partly – explain the difficulties to reproduce extreme events (Bittner et al., 2021; Huang et al., 2019; Lobligeois et al., 2014; Ollivier et al., 2020), especially high flows.

ii. Observed data: Discharge time series are generally derived from water height measured at the spring, using water level–discharge calibration curves. Numerous uncertainties are related to this determination method (Pelletier, 1988), including extrapolation errors for extreme values (Di Baldassarre and Montanari, 2009; Moges et al., 2021). Extreme events occur more rarely and are harder to measure, especially high flows. This can result in inaccurately observed discharge time series that are difficult to reproduce with simulations (e.g. Unica springs at very high flows). The uncertainties related to discharge measurements are highly dependent on the quality of the gauging station and usually range between 10 and 40 % (McMillan et al., 2018). Although they are expected to be higher in a karst context (Westerberg et al., 2016), some authors reported uncertainties of about 20 % (Jeannin et al., 2021) or 10–15 % (Katz et al., 2009).

iii. Internal/external system dynamics: Karst systems are inherently complex media. Internal dynamics are not necessarily captured in hydrological models (Hartmann et al., 2017; Sidle, 2006; Sidle, 2021) and can be related to numerous processes in karst media, e.g. the saturation state of the system, surface water exchanges,

temporary storage of water, incoming or outgoing flows from/to another aquifer, change in physical properties beyond a certain level, or karst features such as poljes or sinkholes. These complex processes do not occur systematically and can change from year to year (Ollivier et al., 2020). This can lead to difficulties in training ANN models or in adapting the structure of reservoir models.

iv. Extreme events during calibration: ANN and reservoir models are both trained on a calibration period. By definition, extreme events are rare. Therefore, models may have less opportunities to successfully fit model parameters to such events (Seibert, 2003), preferring more balanced parameters that are appropriate to the rest – and most – of the time series (Onyutha, 2019). In addition, models are generally calibrated over the whole time series using one performance criterion against observed data. In this case, extreme events are not explicitly emphasised in the objective function. A solution could be to give more weight to the reproduction of certain parts of the time series, such as flood and dry periods (Singh and Bárdossy, 2012), or to use different model optimisation techniques, such as cross-validation (Wilks, 2011).

Both approaches can also benefit from a careful assessment of the calibration period. For example, the ANN model is thought to overestimate low flows in Unica springs by trying to fit the plateaus at very high discharges. In Lez spring, the reservoir model simulates a slower draining in the aquifer (piezometric level) because it does not account for a potential change in underground dynamics. These limitations emphasise the need for a meticulous investigation of the results in regard to the characteristics of the system and the input data. Such errors can be avoided or lessened by excluding abnormal periods during the calibration, which can be justified by inaccurate input data or limitation in the conceptual model.

6.4.3 Comparison of general model properties

The main findings of this study are presented in Table 6.5. The extensive analysis of high and low flows did not show a clear trend, but did reveal slight differences between the two approaches for this study. For high-flow periods, results slightly favour the ANN approach (except for Qachqouch spring), with consistently accurate volumes and shape and timing (Figure 6.6). ANN models also tend to achieve higher flows than reservoir models (Figure 6.4); due to the noticeable/strong karstification of the studied systems, the high occurrence of high discharge data may benefit the learning of the ANN models. On the other hand, reservoir models are more dependent on the relevance and the quality of the input data preprocessing, thus can be more affected by the uncertainties presented in section 6.4.2, especially regarding high flows. For low-flow periods, results slightly favour the reservoir approach (except for Aubach spring), with good estimation of volumes and only a slight overestimation of the hydrological variability (Figure 6.6). The conceptual representation of the aquifer with reservoirs and transfer functions may help to simulate the recharge process (especially for inertial systems): a precipitation event will not directly result in a discharge increase at the spring if the reservoir is not fully saturated. On the other hand, ANN models seem to not always account for the time needed for the aquifer to replenish, inducing wave-like behaviours during medium- and low-flow periods (Appendix 6.E), which can hinder the simulation of low flows. The water level (Lez spring) was correctly simulated by both approaches, with only some imprecision during dry periods (Figure 6.5).

	ANN models	Reservoir models
Advantages	Fast and reliable	The simulation is supported by a conceptual model
	Flexible regarding input data	Slightly better on low flows
	Slightly better on high flows	Can be used to gain knowledge about system functioning
	Can be used to gain knowledge about input data, catchment delineation, recharge processes	Can work with short observed time series
Drawbacks	Struggle to reproduce extreme events	Struggle to reproduce extreme events
	Need medium/long observed time series for a proper learning	Input data generally need preprocessing
	Essentially a black-box approach	Can be time consuming
		Potential platform/coding limitations

Table 6.5: Advantages and drawbacks of ANN and reservoir models, based on the results of this research.

ANN models are flexible and provide numerous advantages over reservoir models with respect to input data. They can easily integrate meteorological processes (e.g. snow dynamics, evapotranspiration) without any preprocessing of the raw data, whereas this is generally calculated beforehand in reservoir models. It is also possible to add a large amount of raw data in ANN models and let the model select those relevant for a good simulation, which makes the modelling easier and can also give insight into the input data or catchment characteristics (Wunsch et al., 2022). This helps to avoid additional uncertainties related to (i) arbitrary decisions over the raw data (e.g. choosing precipitation from one station rather than another), (ii) interpolation (when data from several meteorological stations over a catchment are available) or (iii) preprocessing (e.g. snow routine, potential evapotranspiration). This great flexibility regarding input data makes ANN models close to a 2D or semi-distributed approach. If necessary, the transition between 1D and 2D input data are comparably easy, whereas in reservoir models this usually involves changing or adapting the tool.

Reservoir models do not need a long calibration period to provide accurate and relevant simulation results. In contrast, a very short time series or a short time series with long dry periods can be detrimental for the learning of ANN model, which seems to benefit from long periods of relevant discharge (at least 5 years). We have seen that the ANN model has difficulties in simulating the flows of the Qachqouch spring, mainly because of (i) the short calibration period, and (ii) the long low water periods which are not relevant for training the model. On the other hand, the reservoir model has been able to integrate key elements (e.g. double porosity, matrix-conduit exchanges, fast conduit transfer in wet periods) by relying on the conceptual model.

The ANN approach does not require any prior knowledge of the system and inherently considers model structure and parameters. This makes the modelling process easier and faster thus saving the operator a great amount of time. On the other hand, reservoir models require a significant investment in reading the literature, analysing expert knowl-

edge, and doing trial and error during model design. Moreover, the cost of a change of structure is not trivial. Depending on the modelling platform (e.g. software, raw code), it may take more or less time – or even be impossible – to take certain elements into account. For example, in this study, the KarstMod platform does not allow one to (i) consider different porosities in the same reservoir, leading to difficulties in modelling the piezometric levels during dry periods for the Lez system; (ii) use different E_{min} values, which may benefit the Qachqouch model; (iii) consider polje and surface water influence in the Unica model; or (iv) consider snow dynamics in the structure of the Aubach model.

Both ANN and reservoir models can be used for research purposes. Model structure, transfer functions and parameters are explicitly expressed in reservoir models, which can provide valuable insights into the hydrogeological structure of the reservoir and the internal processes of the karst system, e.g. (i) the relative contributions of fast and slow flows; (ii) the draining of each compartment; (iii) the activation thresholds of the overflow transfer functions (either to the spring or out of the system); (iv) the changes in flow dynamics with respect to dry and wet periods; and (v) the exchanges between the matrix and conduit compartments. In comparison, ANN models act rather as a "black-box", whose parameters are more difficult to exploit and associate with the hydrological functioning of a system. However, ANN model can help to explore input data, thus indirectly providing insights into catchment delineation or external recharge processes (Wunsch et al., 2022).

6.5 Conclusion

Our objective was to provide researchers and stakeholders with guidelines for choosing either artificial neural networks or reservoir models to simulate karst spring discharges, depending on their purpose, data availability, data length and time budget. Five test sites were considered, allowing different hydrological conditions and input data to be studied. The results of ANN and reservoir models were compared on the basis of several performance criteria, distinguishing between high- and low-flow conditions. Both models succeeded in simulating spring discharges satisfactorily, but struggled to reproduce extreme events (drought, flood), generally overestimating low flows and underestimating high flows. This can be related to common problems in hydrological modelling regarding uncertainties in the input data or observed data, internal/external system dynamics or the consideration of extreme events during calibration.

ANN models seem robust for reproducing high-flow conditions and reservoir models for reproducing low-flow conditions. The input data are also a critical factor in choice. Reservoir models can work with relatively short time series while ANN models need a minimum number of relevant years to learn the functioning of a karst system. On the other hand, ANN models are very flexible on the format and amount of input data. They can learn many meteorological processes with no prior need for preprocessing the raw data, as well as use several time series for a single variable. This avoids arbitrary interpolation decisions (e.g. precipitation between several stations), parameter definitions (e.g. snow routine) or meteorological calculation (e.g. potential evapotranspiration), and allows these aspects to be integrated into the model calibration.

Both ANN and reservoir models can be used for karst aquifer management, flood forecasting and system characterisation. ANN models may be more appropriate for simulating high flows, delineating catchments, or assessing external recharge processes. Reservoir

models seem more robust for simulating low flows and gaining insights into the internal functioning of a system. ANN models can also be interesting time-wise as (i) they do not require any prior knowledge of the system and (ii) model design is more flexible regarding input data and system functioning.

One of the difficulties this paper faced was to distinguish the general limitations of the reservoir modelling approach from those specific to the chosen modelling platform. In comparison to user-defined models, the modelling platform constrains the structure and the transfer functions of the conceptual model. Remaining within the KarstMod platform provided the time advantages of a turnkey toolbox (which are widely used in research and by stakeholders), but limited the possibilities of the conceptual models.

Acknowledgements

We thank the French Ministry of Higher Education and Research for the thesis scholarship of G. Cinkus as well as the European Commission for its support through the Partnership for Research and Innovation in the Mediterranean Area (PRIMA) program under Horizon 2020 (KARMA project, grant agreement number 01DH19022A). The data collection and instrumentation on the Qachqouch catchment were funded by USAID and National Academy of Science (Peer Science; project award: 102881; Cycle 3) and KARMA project (L-CNRS in the framework of the PRIMA program; Award 103895; Project 25713). Moreover, Beirut and Mount Lebanon Water Establishment are thanked to facilitate the installation of instruments and access to field sites. We thank the Slovenian Research Agency for financial support within the project "Infiltration processes in forested karst aquifers under changing environment" (No. J2-1743) and the Karst Research Programme (No. P6-0119). We further thank the National Agency of Research of the Spanish Ministry of Science, Innovation and Universities for the funding of the KARMA project (PCI2019-103675) and the Autonomous Government of Andalusia (Spain) for the funding of the Research Group RNM-308. For the data provided, we also acknowledge the French Karst National Observatory Service (SNO KARST, 2019), the French national meteorological service Météo-France, the office of the federal state of Vorarlberg – division of water management, the German Meteorological Service, the Slovenian Environment Agency (ARSO, 2021a; ARSO, 2021b) and the Meteorological National Agency (AEMET).

The manuscript was written with the Rmarkdown framework (Allaire et al., 2021; Xie et al., 2018; Xie et al., 2020), using R (R Core Team, 2021) and the following packages: readxl, readr, dplyr, tidyr, ggplot2, lubridate (Wickham et al., 2019), cowplot (Wilke, 2020), flextable (Gohel, 2021), hydroGOF (Zambrano-Bigiarini, 2020) and padr (Thoen, 2021). We programmed our ANN models in Python 3.8 (van Rossum, 1995), using the following frameworks and libraries: Numpy (van der Walt et al., 2011), Pandas (McKinney, 2010; Reback et al., 2021), Scikit-Learn (Pedregosa et al., 2018), Matplotlib (Hunter, 2007), Bayesian Optimization (Nogueira, 2014), TensorFlow 2 .7 (Abadi et al., 2016) and its Keras API (Chollet, 2015).

Author contribution

GC, AW, NM, TL, HJ and NG conceptualised the study and designed the methodology. GC and AW developed the software code. GC and AW performed the experiments and investigated and visualised the results. GC and ZC performed formal analysis. GC wrote the original paper draft with contributions from AW, NR, JD, JFO. All authors

contributed to the interpretation of the results and review and editing of the paper draft. NM and HJ supervised the work.

Code and data availability

We provide complete codes for ANN models and .properties files for reservoir models on Github (Cinkus and Wunsch, 2022). Due to redistribution restrictions from several parties, a dataset cannot be provided. However, the data are available from the local authorities upon request. Aubach spring discharge time series and meteorological data from Diedamskopf and Walmendinger Horn stations are available from the office of the federal state of Vorarlberg – division of water management. Meteorological data from Oberstdorf station are available on the DWD Open Data Server (DWD, 2022). Gato Cave spring discharge time series is available from the Confederación Hidrográfica de las Cuencas Mediterráneas Andaluzas (Cuenca Mediterránea Andaluza, 2022) and meteorological data is available in "Datos a la carta" section in Consejería de Agricultura, Pesa, Agua y Desarrollo rural (Consejería de Agricultura, Pesa, Agua y Desarrollo rural, 2022). Lez spring discharge time series is available on the OSU OREME website (SNO KARST, 2019), water level time series can be requested from Montpellier Méditerranée Métropole, and meteorological data are available on request from Météo-France. Qachqouch discharge time series and meteorological data are available on request from the Department of Geology at the American University of Beirut. Unica spring discharge time series and meteorological data are available from ARSO (Slovenian Environment Agency) (ARSO, 2021a; ARSO, 2021b).

Competing Interest

The authors declare that they have no conflict of interest.

6.6 Appendix

6.A Origin of the meteorological data

6.B Calculation details for the Thiessen's polygon interpolation method

The Thiessen's polygon interpolation method consists of calculating a weighted average of precipitation data from several meteorological stations. The contribution percentages of the stations are proportional to their influence area on the catchment. An influence area corresponds to a polygon where the precipitation is considered to be identical to that measured at the associated meteorological station. The polygons are defined in two steps: (i) drawing the straight-line segments between all adjacent stations and (ii) adding the perpendicular bisectors of each segment, which correspond to the edges of the polygons. The weighted average of the precipitation P_{TH} is calculated with the following equation:

$$P_{TH} = \frac{\sum_{i=1}^{n} A_i P_i}{A} \tag{6.3}$$

with A the area of the catchment, n the number of meteorological stations, A_i the area of the polygon associated with the i^{th} station and P_i the precipitation measured at the i^{th} station.

Spring	Station	Altitude [m asl]	Latitude [°]	Longitude [°]	Data measured
Aubach	Diedamskopf	1790	47.3389	10.0256	P, T, R_{SO}
Aubach	Oberstdorf	806	47.3984	10.2759	P, T, R_{SO}
Aubach	Walmendinger Horn	1650	47.3216	10.1225	P, T, R_{SO}
Gato Cave	Grazamela	901	36.7678	-5.3658	P, T
Lez	Prades-le-Lez	69	43.7176	3.8573	P, T, RH, U
Lez	Puéchabon	250	43.7414	3.5958	AET
Lez	Saint-Martin-de-Londres	214	43.7903	3.7326	P
Lez	Sauteyrargues	150	43.8345	3.9207	P
Lez	Valflaunès	155	43.8001	3.8707	P
Qachqouch	*950 m station*	950	33.9180	35.6763	P, T, RH, U, R_S
Qachqouch	*1700 m station*	1700	34.0253	35.8360	P, T, RH, U, R_S
Unica	Cerknica	569	45.7956	14.3634	P, S, NS
Unica	Postojna	533	45.7661	14.1932	P, T, RH, S, NS

Table 6.6: Origin of the meteorological data (i) P, (ii) T, (iii) R_{SO}, (iv) RH, (v) U, (vi) AET, (vii) R_S, (viii) S and (ix) NS refer to (i) precipitation, (ii) temperature, (iii) clear-sky solar radiation, (iv) relative humidity, (v) wind speed, (vi) actual evapotranspiration, (vii) solar radiation, (viii) snow and (ix) new snow, respectively.

6.C Calculation details for the snow routine

Accounting for snow accumulation and melting in hydrological modelling can greatly improve model results, especially for regions with high snow volumes. Chen et al. (2018) successfully simulated spring discharge of a mountainous karst system strongly influenced by snow accumulation and melting. They applied a modified version of the HBV snow routine Bergström (1992) proposed by (Hock, 1999). We used this snow routine as an external KarstMod module (i.e. without internal calibration).

The snow routine simulates snow accumulation and melting over different subcatchments defined according to altitude ranges. The input data consist of three time series (temperature, precipitation and potential clear-sky solar radiation) and five parameters (temperature threshold, melt coefficient, refreezing coefficient, radiation coefficient and water holding capacity of snow). The potential clear sky solar radiation time series and radiation coefficient are only used when working at an hourly timescale to simulate a more refined snowmelt by considering sun exposure. The parameters values were estimated by model calibration.

Precipitation is considered as snow when the air temperature is below the temperature threshold. Snowmelt begins when the temperature is above the threshold according to a degree-day expression, where snowmelt is a function of the melt coefficient, solar radiation and degrees above the threshold. Runoff starts when the liquid water holding capacity of snow is exceeded. The refreezing coefficient allows one to consider the refreezing processes of liquid water in the snow if snowmelt is interrupted (Bergström, 1992). The output of the snow routine is a time series of redistributed precipitation.

6.D Calibration scores of the reservoir models

6.E Examples of wave-like behaviour produced by the ANN model

The periods were selected in such a way that the influence of snow precipitation and snowmelt is zero or almost zero. Precipitation input corresponds to either direct observa-

Spring	Calibration period		Validation period	
	ANN	Reservoir	ANN	Reservoir
Aubach	/	0.40	**0.26**	0.63
Gato Cave	/	1.53	1.53	**1.07**
Lez	/	8.63	**4.53**	8.84
Qachqouch	/	1.17	7.78	**2.87**
Unica	/	81.57	**123.51**	152.92

Table 6.7: Scores of both modelling approaches over the calibration and validation periods with MSE. No results are available in the calibration period for the ANN models, as this corresponds to the learning period of the models. Each component of the composite objective function MSE(Q, Z) of the Lez spring has been normalised.

tions from a meteorological station, or preprocessed observations with Thiessen's polygon interpolation (Appendix 6.B) if there are several meteorological stations.

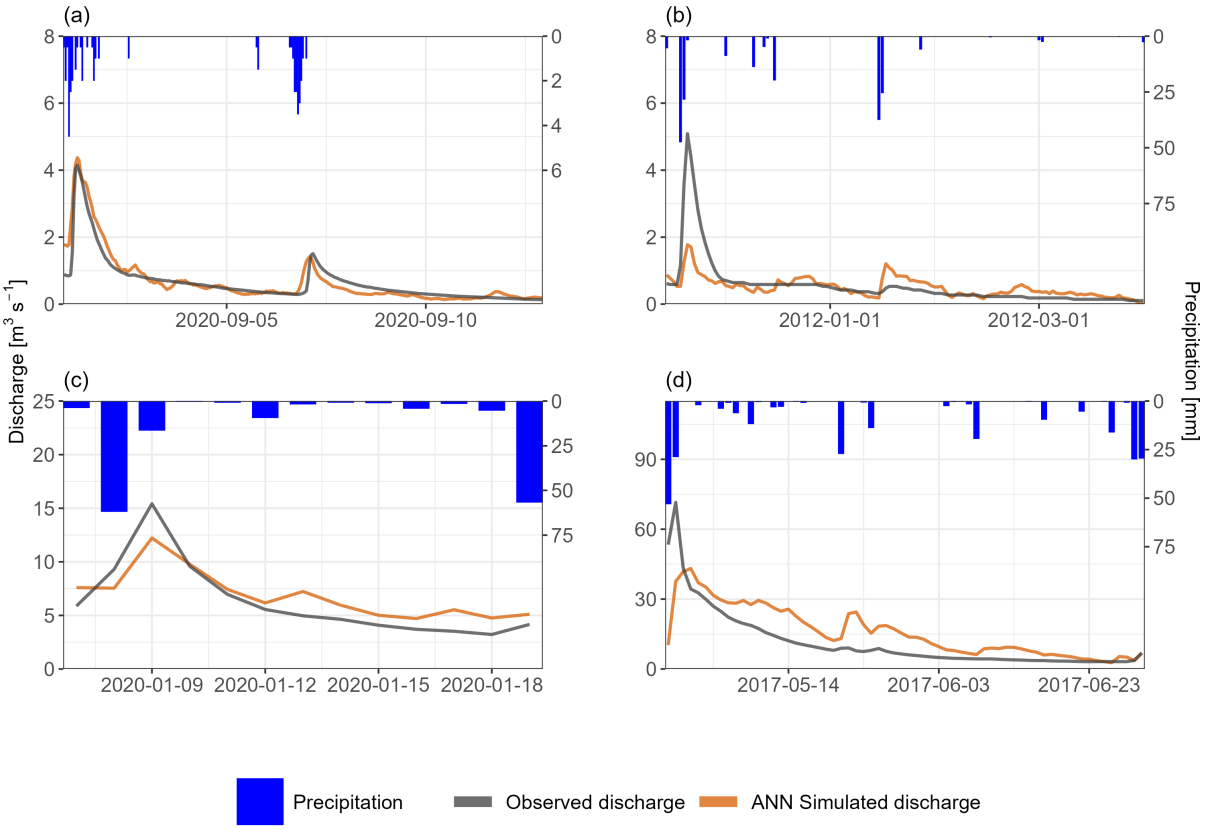


Figure 6.8: Examples of wave-like behaviour produced by the ANN model on (a) Aubach, (b) Gato Cave, (c) Qachqouch and (d) Unica springs.

6.7 References for Chapter 6

Aalto, J., Riihimäki, H., Meineri, E., Hylander, K., & Luoto, M. (2017). Revealing topoclimatic heterogeneity using meteorological station data. *Intern. J. Climatol.*, *37*(S1), 544–556. https://doi.org/10.1002/joc.5020 (cit. on p. 158).

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.* (Cit. on p. 162).

Addor, N., & Melsen, L. A. (2019). Legacy, Rather Than Adequacy, Drives the Selection of Hydrological Models. *Water Resources Research*, *55*(1), 378–390. https://doi.org/10.1029/2018WR022958 (cit. on p. 139).

Allaire, J. J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., & Iannone, R. (2021). *Rmarkdown: Dynamic documents for r.* (Version 2.22, R package). (Cit. on p. 162).

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., & FAO (Eds.). (1998). *Crop evapotranspiration: Guidelines for computing crop water requirements.* Food and Agriculture Organization of the United Nations. (Cit. on pp. 143, 144).

Andreo, B., Vadillo, I., Carrasco, F., Neukum, C., Jiménez, P., Goldscheider, N., Hötzl, H., Vías, J., Pérez, I., & Göppert, N. (2004). Precisiones sobre el funcionamiento hidrodinámico y la vulnerabilidad a la contaminación del acuífero kárstico de la

sierra de líbar (provincias de málaga y cádiz, sur de españa) a partir de un ensayo de trazadores. *Rev. Soc. Geol. Esp.*, *17*(3-4), 187–198 (cit. on p. 142).

Andreo, B., Goldscheider, N., Vadillo, I., Vías, J., Neukum, C., Sinreich, M., Gavilán, P., Brechenmacher, J., Carrasco, F., Hötzl, H., Perles, M., & Zwahlen, F. (2006). Karst groundwater protection: First application of a Pan-European Approach to vulnerability, hazard and risk mapping in the Sierra de Líbar (Southern Spain). *Sci. Total Environ.*, *357*, 54–73. https://doi.org/10.1016/j.scitotenv.2005.05.019 (cit. on p. 142).

ARSO. (2021a). *Archive of meteorological data, Ministry of the Environment and Spatial Planning, Slovenian Environment Agency, Archive of hydrological data.* Retrieved March 6, 2023, from http://vode.arso.gov.si/hidarhiv/ (cit. on pp. 162, 163).

ARSO. (2021b). *Archive of meteorological data, Ministry of the Environment and Spatial Planning, Slovenian Environment Agency, Archive of meteorological data.* Retrieved June 6, 2023, from http://www.meteo.si/ (cit. on pp. 162, 163).

Bakalowicz, M. (2005). Karst groundwater: A challenge for new resources. *Hydrogeol. J.*, *13*(1), 148–160. https://doi.org/10.1007/s10040-004-0402-9 (cit. on p. 139).

Bergström, S. (1992). The HBV model - its structure and applications. *SMHI*, *4*, 32 (cit. on p. 164).

Bittner, D., Parente, M. T., Mattis, S., Wohlmuth, B., & Chiogna, G. (2020). Identifying relevant hydrological and catchment properties in active subspaces: An inference study of a lumped karst aquifer model. *Adv. Water. Resour.*, *135*, 103472. https://doi.org/10.1016/j.advwatres.2019.103472 (cit. on p. 139).

Bittner, D., Richieri, B., & Chiogna, G. (2021). Unraveling the time-dependent relevance of input model uncertainties for a lumped hydrologic model of a pre-alpine karst system. *Hydrogeol. J.*, *29*(7), 2363–2379. https://doi.org/10.1007/s10040-021-02377-1 (cit. on p. 158).

Çallı, S. S., Çallı, K. Ö., Tuğrul Yılmaz, M., & Çelik, M. (2022). Contribution of the satellite-data driven snow routine to a karst hydrological model. *Journal of Hydrology*, *607*, 127511. https://doi.org/10.1016/j.jhydrol.2022.127511 (cit. on p. 154).

Chen, Z., & Goldscheider, N. (2014). Modeling spatially and temporally varied hydraulic behavior of a folded karst system with dominant conduit drainage at catchment scale, Hochifen–Gottesacker, Alps. *J. Hydrol.*, *514*, 41–52 (cit. on p. 142).

Chen, Z., Hartmann, A., Wagener, T., & Goldscheider, N. (2018). Dynamics of water fluxes and storages in an Alpine karst catchment under current and potential future climate conditions. *Hydrol. Earth Syst. Sci.*, *22*(7), 3807–3823. https://doi.org/10.5194/hess-22-3807-2018 (cit. on pp. 140, 142, 154, 164).

Cheng, S., Qiao, X., Shi, Y., & Wang, D. (2020). *Comparison of Machine Learning Methods for Predicting Karst Spring Discharge in North China.* (Cit. on p. 139).

Chollet, F. (2015). *Keras.* Retrieved June 6, 2023, from https://github.com/keras-team/keras (cit. on p. 162).

Cinkus, G., & Wunsch, A. (2022). *Busemorose/ANN-Reservoir_model-code: Model code release.* https://doi.org/10.5281/zenodo.7242077 (cit. on pp. 144, 163).

Cinkus, G., Wunsch, A., Mazzilli, N., Liesch, T., Chen, Z., Ravbar, N., Doummar, J., Fernández-Ortega, J., Barberá, J. A., Andreo, B., Goldscheider, N., & Jourde, H. (2023). Comparison of artificial neural networks and reservoir models for simulating karst spring discharge on five test sites in the Alpine and Mediterranean re-

gions. *Hydrol. Earth Syst. Sci.*, *27*(10), 1961–1985. https://doi.org/10.5194/hess-27-1961-2023 (cit. on p. 137).

Conradt, T., Wechsung, F., & Bronstert, A. (2013). Three perceptions of the evapotranspiration landscape: Comparing spatial patterns from a distributed hydrological model, remotely sensed surface temperatures, and sub-basin water balances. *Hydrol. Earth Syst. Sci.*, *17*(7), 2947–2966. https://doi.org/10.5194/hess-17-2947-2013 (cit. on p. 142).

Consejería de Agricultura, Pesa, Agua y Desarrollo rural. (2022). *Archive of meteorological data*. Retrieved October 13, 2022, from http://www.redhidrosurmedioambiente.es/saih/presentacion (cit. on p. 163).

Cuenca Mediterránea Andaluza. (2022). *Archive of hydrological data*. Retrieved October 13, 2022, from https://hispagua.cedex.es/instituciones/confederaciones/andalucia (cit. on p. 163).

Dausse, A., Leonardi, V., & Jourde, H. (2019). Hydraulic characterization and identification of flow-bearing structures based on multi-scale investigations applied to the Lez karst aquifer. *J. Hydrol. Reg. Stud.*, *26*, 100627. https://doi.org/10.1016/j.ejrh.2019.100627 (cit. on p. 156).

Di Baldassarre, G., & Montanari, A. (2009). Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.*, *13*(6), 913–921. https://doi.org/10.5194/hess-13-913-2009 (cit. on p. 158).

Doummar, J., Hassan Kassem, A., & Gurdak, J. J. (2018). Impact of historic and future climate on spring recharge and discharge based on an integrated numerical modelling approach: Application on a snow-governed semi-arid karst catchment area. *J. Hydrol.*, *565*, 636–649. https://doi.org/10.1016/j.jhydrol.2018.08.062 (cit. on p. 143).

Dubois, E. (2017). *Analysis of high resolution spring hydrographs and climatic data: Application on the Qachqouch spring (Lebanon)* [Doctoral dissertation, American University of Beirut]. (Cit. on p. 143).

Dubois, E., Doummar, J., Pistre, S., & Larocque, M. (2020). Calibration of a lumped karst system model and application to the Qachqouch karst spring (Lebanon) under climate change conditions. *Hydrol. Earth Syst. Sci.*, *24*(9), 4275–4290. https://doi.org/10.5194/hess-24-4275-2020 (cit. on pp. 139, 143).

DWD. (2022). *DWD Opendata*. Retrieved October 12, 2022, from https://opendata.dwd.de/ (cit. on p. 163).

Fleury, P., Ladouche, B., Conroux, Y., Jourde, H., & Dörfliger, N. (2009). Modelling the hydrologic functions of a karst aquifer under active water management – The Lez spring. *J. Hydrol.*, *365*(3-4), 235–243. https://doi.org/10.1016/j.jhydrol.2008.11.037 (cit. on pp. 139, 143).

Fleury, P., Plagnes, V., & Bakalowicz, M. (2007). Modelling of the functioning of karst aquifers with a reservoir model: Application to Fontaine de Vaucluse (South of France). *J. Hydrol.*, *345*(1), 38–49. https://doi.org/10.1016/j.jhydrol.2007.07.014 (cit. on p. 139).

Ford, D., & Williams, P. (2007). Karst Hydrogeology. In *Karst Hydrogeology and Geomorphology* (pp. 103–144). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118684986.ch5 (cit. on p. 139).

Gabrovšek, F., Kogovšek, J., Kovačič, G., Petrič, M., Ravbar, N., & Turk, J. (2010). Recent Results of Tracer Tests in the Catchment of the Unica River (SW Slovenia).

Acta Carsologica, *39*(1), 27–37. https://doi.org/10.3986/ac.v39i1.110 (cit. on p. 143).

Gohel, D. (2021). *Flextable: Functions for tabular reporting* (Version 0.9.1, R package). (Cit. on p. 162).

Goldscheider, N. (2005). Fold structure and underground drainage pattern in the alpine karst system Hochifen-Gottesacker. *Eclogae Geol. Helv.*, *98*(1), 1–17. https://doi.org/10.1007/s00015-005-1143-z (cit. on p. 142).

Goldscheider, N. (2015). Overview of Methods Applied in Karst Hydrogeology. In Z. Stevanović (Ed.), *Karst Aquifers—Characterization and Engineering* (pp. 127–145). Springer International Publishing. https://doi.org/10.1007/978-3-319-12850-4_4 (cit. on p. 139).

Goldscheider, N., Chen, Z., Auler, A. S., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Stevanovic, Z., & Veni, G. (2020). Global distribution of carbonate rocks and karst water resources. *Hydrogeol J*, *28*(5), 1661–1677. https://doi.org/10.1007/s10040-020-02139-5 (cit. on p. 141).

Gondwe, B. R. N., Merediz-Alonso, G., & Bauer-Gottwein, P. (2011). The influence of conceptual model uncertainty on management decisions for a groundwater-dependent ecosystem in karst. *J. Hydrol.*, *400*(1), 24–40. https://doi.org/10.1016/j.jhydrol.2011.01.023 (cit. on p. 139).

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, *377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003 (cit. on p. 149).

Hargreaves, G. H., & Samani, Z. A. (1985). Reference Crop Evapotranspiration from Temperature. *Appl. Eng. Agric.*, *1*(2), 96–99. https://doi.org/10.13031/2013.26773 (cit. on p. 142).

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., & Weiler, M. (2014). Karst water resources in a changing world: Review of hydrological modeling approaches. *Rev. Geophys.*, *52*(3), 218–242. https://doi.org/10.1002/2013RG000443 (cit. on pp. 139, 145).

Hartmann, A., Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proc. Natl. Acad. Sci.*, *114*(11), 2842–2847. https://doi.org/10.1073/pnas.1614941114 (cit. on p. 158).

Hartmann, A., Lange, J., Aguado, A. V., Mizyed, N., Smiatek, G., & Kunstmann, H. (2012). A multi-model approach for improved simulations of future water availability at a large Eastern Mediterranean karst spring. *J. Hydrol.*, *468–469*, 130–138. https://doi.org/10.1016/j.jhydrol.2012.08.024 (cit. on p. 139).

Hock, R. (1999). A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. *J. Glaciol.*, *45*(149), 101–111. https://doi.org/10.3189/S0022143000003087 (cit. on p. 164).

Hohmann, C., Kirchengast, G., O, S., Rieger, W., & Foelsche, U. (2021). Small Catchment Runoff Sensitivity to Station Density and Spatial Interpolation: Hydrological Modeling of Heavy Rainfall Using a Dense Rain Gauge Network. *Water*, *13*(10), 1381. https://doi.org/10.3390/w13101381 (cit. on p. 158).

Hu, C., Hao, Y., Yeh, T.-C. J., Pang, B., & Wu, Z. (2008). Simulation of spring flows from a karst aquifer with an artificial neural network. *Hydrol. Process.*, *22*(5), 596–604. https://doi.org/10.1002/hyp.6625 (cit. on p. 139).

Huang, Y., Bárdossy, A., & Zhang, K. (2019). Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data. *Hydrol. Earth Syst. Sci.*, *23*(6), 2647–2663. https://doi.org/10.5194/hess-23-2647-2019 (cit. on p. 158).

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55 (cit. on p. 162).

Ioffe, S., & Szegedy, C. (2015). *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. (Cit. on p. 144).

Jeannin, P.-Y., Artigue, G., Butscher, C., Chang, Y., Charlier, J.-B., Duran, L., Gill, L., Hartmann, A., Johannet, A., Jourde, H., Kavousi, A., Liesch, T., Liu, Y., Lüthi, M., Malard, A., Mazzilli, N., Pardo-Igúzquiza, E., Thiéry, D., Reimann, T., … Wunsch, A. (2021). Karst modelling challenge 1: Results of hydrological modelling. *J. Hydrol.*, *600*, 126508. https://doi.org/10.1016/j.jhydrol.2021.126508 (cit. on pp. 139, 140, 158).

Jukić, D., & Denić-Jukić, V. (2009). Groundwater balance estimation in karst by using a conceptual rainfall–runoff model. *J. Hydrol.*, *373*(3), 302–315. https://doi.org/10.1016/j.jhydrol.2009.04.035 (cit. on p. 139).

Katz, B. G., Sepulveda, A. A., & Verdi, R. J. (2009). Estimating Nitrogen Loading to Ground Water and Assessing Vulnerability to Nitrate Contamination in a Large Karstic Springs Basin, Florida1. *JAWRA Journal of the American Water Resources Association*, *45*(3), 607–627. https://doi.org/10.1111/j.1752-1688.2009.00309.x (cit. on p. 158).

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011 (cit. on p. 149).

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrol. Earth Syst. Sci.*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019 (cit. on p. 150).

Kogovšek, J., Knez, M., Mihevc, A., Petrič, M., Slabe, T., & Šebela, S. (1999). Military training area in Kras (Slovenia). *Environ. Geol.*, *38*(1), 69–76. https://doi.org/10.1007/s002540050402 (cit. on p. 143).

Kong-A-Siou, L., Fleury, P., Johannet, A., Borrell Estupina, V., Pistre, S., & Dörfliger, N. (2014). Performance and complementarity of two systemic models (reservoir and neural networks) used to simulate spring discharge and piezometry for a karst aquifer. *J. Hydrol.*, *519*, 3178–3192. https://doi.org/10.1016/j.jhydrol.2014.10.041 (cit. on p. 139).

Kong-A-Siou, L., Johannet, A., Borrell, V., & Pistre, S. (2011). Complexity selection of a neural network model for karst flood forecasting: The case of the Lez Basin (southern France). *J. Hydrol.*, *403*(3), 367–380. https://doi.org/10.1016/j.jhydrol.2011.04.015 (cit. on p. 139).

Kong-A-Siou, L., Johannet, A., Borrell Estupina, V., & Pistre, S. (2015). Neural networks for karst groundwater management: Case of the Lez spring (Southern France). *Environ. Earth Sci.*, *74*(12), 7617–7632. https://doi.org/10.1007/s12665-015-4708-9 (cit. on p. 139).

Kovačević, M., Ivanišević, N., Dašić, T., & Marković, L. (2018). Application of artificial neural networks for hydrological modelling in Karst. *Građevinar*, *70*(1), 1–10. https://doi.org/10.14256/JCE.1594.2016 (cit. on p. 139).

Kovačič, G. (2010). Hydrogeological study of the Malenščica karst spring (SW Slovenia) by means of a time series analysis. *Acta Carsologica*, *39*(2), 201–215. https://doi.org/10.3986/ac.v39i2.93 (cit. on p. 143).

Kovács, A., & Sauter, M. (2007). Modelling karst hydrodynamics. In N. Goldscheider & D. Drew (Eds.), *Methods in Karst Hydrogeology* (pp. 201–222). Taylor & Francis. (Cit. on p. 139).

Kurtulus, B., & Razack, M. (2007). Evaluation of the ability of an artificial neural network model to simulate the input-output responses of a large karstic aquifer: The La Rochefoucauld aquifer (Charente, France). *Hydrogeol. J.*, *15*(2), 241–254. https://doi.org/10.1007/s10040-006-0077-5 (cit. on p. 139).

Kurtulus, B., & Razack, M. (2010). Modeling daily discharge responses of a large karstic aquifer using soft computing methods: Artificial neural network and neuro-fuzzy. *J. Hydrol.*, *381*(1), 101–111. https://doi.org/10.1016/j.jhydrol.2009.11.029 (cit. on p. 139).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539 (cit. on p. 144)
Bandiera_abtest: a Cg_type: Nature Research Journals Primary_atype: Reviews Subject_term: Computer science;Mathematics and computing Subject_term_id: computer-science;mathematics-and-computing.

Lehmann, P., Stauffer, F., Hinz, C., Dury, O., & Flühler, H. (1998). Effect of hysteresis on water flow in a sand column with a fluctuating capillary fringe. *J. Contam. Hydrol.*, *33*(1), 81–100. https://doi.org/10.1016/S0169-7722(98)00066-7 (cit. on p. 145).

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P., & Loumagne, C. (2014). When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events. *Hydrol. Earth Syst. Sci.*, *18*(2), 575–594. https://doi.org/10.5194/hess-18-575-2014 (cit. on p. 158).

Makropoulos, C., Koutsoyiannis, D., Stanić, M., Djordjević, S., Prodanović, D., Dašić, T., Prohaska, S., Maksimović, Č., & Wheater, H. (2008). A multi-model approach to the simulation of large scale karst flows. *J. Hydrol.*, *348*(3), 412–424. https://doi.org/10.1016/j.jhydrol.2007.10.011 (cit. on p. 139).

Martín-Algarra, A. (1987). *Evolucion geologica alpina del contacto entre las zonas internas y las zonas externas de la cordillera Bética (sector centro-occidental)* [Doctoral dissertation, Universidad de Granada]. (Cit. on p. 142).

Mayaud, C., Gabrovšek, F., Blatnik, M., Kogovšek, B., Petrič, M., & Ravbar, N. (2019). Understanding flooding in poljes: A modelling perspective. *J. Hydrol.*, *575*, 874–889. https://doi.org/10.1016/j.jhydrol.2019.04.092 (cit. on p. 143).

Mayaud, C., Kogovšek, B., Gabrovšek, F., Blatnik, M., Petrič, M., & Ravbar, N. (2022). Deciphering the water balance of poljes: Example of Planinsko Polje (Slovenia). *Acta Carsologica*, *51*(2), 155–177. https://doi.org/10.3986/ac.v51i2.11029 (cit. on p. 157).

Mazzilli, N., & Bertin, D. (2019). *KarstMod User Guide - version 2.2*. Retrieved June 6, 2023, from https://hal.science/hal-01832693v1 (cit. on p. 146).

Mazzilli, N., Guinot, V., & Jourde, H. (2012). Sensitivity analysis of conceptual model calibration to initialisation bias. Application to karst spring discharge models. *Adv. Water Resour.*, *42*, 1–16. https://doi.org/10.1016/j.advwatres.2012.03.020 (cit. on p. 139).

Mazzilli, N., Guinot, V., Jourde, H., Lecoq, N., Labat, D., Arfib, B., Baudement, C., Danquigny, C., Soglio, L. D., & Bertin, D. (2019). KarstMod: A modelling platform for rainfall - discharge analysis and modelling dedicated to karst systems. *Environ. Model. Softw.*, *122*, 103927. https://doi.org/10.1016/j.envsoft.2017.03.015 (cit. on pp. 140, 145, 146).

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. https://doi.org/10.25080/Majora-92bf1922-00a (cit. on p. 162).

McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. *WIREs Water*, *5*(6), e1319. https://doi.org/10.1002/wat2.1319 (cit. on p. 158).

Meng, X., Yin, M., Ning, L., Liu, D., & Xue, X. (2015). A threshold artificial neural network model for improving runoff prediction in a karst watershed. *Environ. Earth Sci.*, *74*(6), 5039–5048. https://doi.org/10.1007/s12665-015-4562-9 (cit. on p. 139).

Moges, E., Demissie, Y., Larsen, L., & Yassin, F. (2021). Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water*, *13*(1), 28. https://doi.org/10.3390/w13010028 (cit. on p. 158).

Nash, J., & Sutcliffe, J. (1970). River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.*, *10*(3), 282–290 (cit. on p. 149).

Nogueira, F. (2014). *Bayesian Optimization: Open source constrained global optimization tool for Python*. Retrieved June 6, 2023, from https://github.com/fmfn/BayesianOptimization (cit. on pp. 144, 162).

Ollivier, C., Mazzilli, N., Olioso, A., Chalikakis, K., Carrière, S. D., Danquigny, C., & Emblanch, C. (2020). Karst recharge-discharge semi distributed model to assess spatial variability of flows. *Sci. Total Environ.*, *703*, 134368. https://doi.org/10.1016/j.scitotenv.2019.134368 (cit. on pp. 158, 159).

Onyutha, C. (2019). Hydrological Model Supported by a Step-Wise Calibration against Sub-Flows and Validation of Extreme Flow Events. *Water*, *11*(2), 244. https://doi.org/10.3390/w11020244 (cit. on p. 159).

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *J. Hydrol.*, *303*(1), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026 (cit. on p. 143).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2018). *Scikit-learn: Machine Learning in Python*. (Cit. on p. 162).

Pelletier, P. M. (1988). Uncertainties in the single determination of river discharge: A literature review. *Can. J. Civ. Eng.*, *15*(5), 834–850. https://doi.org/10.1139/l88-109 (cit. on p. 158).

Pereira, D. d. R., Martinez, M. A., Almeida, A. Q. de, Pruski, F. F., Silva, D. D. da, & Zonta, J. H. (2014). Hydrological simulation using SWAT model in headwater basin in Southeast Brazil. *Eng. Agríc.*, *34*, 789–799. https://doi.org/10.1590/S0100-69162014000400018 (cit. on p. 158).

Perrin, J., Jeannin, P.-Y., & Zwahlen, F. (2003). Epikarst storage in a karst aquifer: A conceptual model based on isotopic data, Milandre test site, Switzerland. *J.*

Hydrol., *279*(1), 106–124. https://doi.org/10.1016/S0022-1694(03)00171-9 (cit. on p. 139).

Petric, M. (2010). Chapter 10.3 - Case Study: Characterization, exploitation, and protection of the Malenščica karst spring, Slovenia. In N. Kresic & Z. Stevanovic (Eds.), *Groundwater Hydrology of Springs* (pp. 428–441). Butterworth-Heinemann. https://doi.org/10.1016/B978-1-85617-502-9.00021-9 (cit. on p. 143).

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* Retrieved June 6, 2023, from https://www.R-project.org/ (cit. on p. 162).

Ravbar, N., Barberá, J. A., Petrič, M., Kogovšek, J., & Andreo, B. (2012). The study of hydrodynamic behaviour of a complex karst system under low-flow conditions using natural and artificial tracers (the catchment of the Unica River, SW Slovenia). *Environ. Earth Sci.*, *65*(8), 2259–2272. https://doi.org/10.1007/s12665-012-1523-4 (cit. on p. 143).

Reback, J., jbrockmendel, McKinney, W., Bossche, J. V. den, Augspurger, T., Cloud, P., Hawkins, S., Roeschke, M., gfyoung, Sinhrks, Klein, A., Petersen, T., Hoefler, P., Tratner, J., She, C., Ayd, W., Naveh, S., Garcia, M., Darbyshire, J. H. M., … Seabold, S. (2021). *Pandas-dev/pandas: Pandas 1.3.5*. https://doi.org/10.5281/zenodo.5774815 (cit. on p. 162).

Refsgaard, J. C., van der Sluijs, J. P., Højberg, A. L., & Vanrolleghem, P. A. (2007). Uncertainty in the environmental modelling process – A framework and guidance. *Environ. Model. Softw.*, *22*(11), 1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004 (cit. on p. 145).

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrol. Earth Syst. Sci.*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018 (cit. on p. 149).

Schwemmle, R., Demand, D., & Weiler, M. (2021). Technical note: Diagnostic efficiency – specific evaluation of model performance. *Hydrol. Earth Syst. Sci.*, *25*(4), 2187–2198. https://doi.org/10.5194/hess-25-2187-2021 (cit. on pp. 149, 150).

Seibert, J. (2003). Reliability of Model Predictions Outside Calibration Conditions: Paper presented at the Nordic Hydrological Conference (Røros, Norway 4-7 August 2002). *Hydrol. Res.*, *34*(5), 477–492. https://doi.org/10.2166/nh.2003.0019 (cit. on p. 159).

Sezen, C., Bezak, N., Bai, Y., & Šraj, M. (2019). Hydrological modelling of karst catchment using lumped conceptual and data mining models. *J. Hydrol.*, *576*, 98–110. https://doi.org/10.1016/j.jhydrol.2019.06.036 (cit. on pp. 139, 140).

Sidle, R. C. (2006). Field observations and process understanding in hydrology: Essential components in scaling. *Hydrol. Process.*, *20*(6), 1439–1445. https://doi.org/10.1002/hyp.6191 (cit. on p. 158).

Sidle, R. C. (2021). Strategies for smarter catchment hydrology models: Incorporating scaling and better process representation. *Geosci. Lett.*, *8*(1), 24. https://doi.org/10.1186/s40562-021-00193-9 (cit. on p. 158).

Singh, S. K., & Bárdossy, A. (2012). Calibration of hydrological models on hydrologically unusual events. *Adv. Water Resour.*, *38*, 81–91. https://doi.org/10.1016/j.advwatres.2011.12.006 (cit. on p. 159).

SNO KARST. (2019). *Time series of type hydrology-hydrogeology in Le Lez (Méditerranée) basin - MEDYCYSS observatory - KARST observatory network - OZCAR Critical Zone network Research Infrastructure* (CSV). CSV. OSU OREME. https:

//doi.org/10.15148/CFD01A5B-B7FD-41AA-8884-84DBDDAC767E (cit. on pp. 162, 163).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.*, *15*(56), 1929–1958 (cit. on p. 144).

Stevanović, Z. (2019). Karst waters in potable water supply: A global scale overview. *Environ. Earth Sci.*, *78*(23), 662. https://doi.org/10.1007/s12665-019-8670-9 (cit. on p. 139).

Thoen, E. (2021). *Padr: Quickly get datetime data ready for analysis.* (Version 0.6.2, R package). (Cit. on p. 162).

Tritz, S., Guinot, V., & Jourde, H. (2011). Modelling the behaviour of a karst system catchment using non-linear hysteretic conceptual model. *J. Hydrol.*, *397*(3-4), 250–262. https://doi.org/10.1016/j.jhydrol.2010.12.001 (cit. on pp. 139, 145).

Van, S. P., Le, H. M., Thanh, D. V., Dang, T. D., Loc, H. H., & Anh, D. T. (2020). Deep learning convolutional neural network in rainfall–runoff modelling. *J. Hydroinformatics*, *22*(3), 541–561. https://doi.org/10.2166/hydro.2020.095 (cit. on p. 140).

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Comput. Sci. Eng.*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37 (cit. on p. 162).

van Rossum, G. (1995). *Python Tutorial.* CWI. (Cit. on p. 162).

Wendling, U., & Müller, J. (1984). Entwicklung eines Verfahrens zur rechnerischen Abschätzung der Verdunstung im Winter. *Z. Meteorol.*, *34*, 82–85 (cit. on p. 142).

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A., & Freer, J. (2016). Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resources Research*, *52*(3), 1847–1865. https://doi.org/10.1002/2015WR017635 (cit. on p. 158).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *J. Open Source Softw.*, *4*(43), 1686. https://doi.org/10.21105/joss.01686 (cit. on p. 162).

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* (Version 1.1.1, R package). (Cit. on p. 162).

Wilks, D. (2011). Statistical Forecasting. In D. Wilks (Ed.), *International Geophysics* (pp. 215–300, Vol. 100). Elsevier. https://doi.org/10.1016/B978-0-12-385022-5.00007-5 (cit. on p. 159).

Willmott, C. J., Robeson, S. M., & Matsuura, K. (2012). A refined index of model performance. *Intern. J. Climatol.*, *32*(13), 2088–2094. https://doi.org/10.1002/joc.2419 (cit. on p. 149).

Wu, Q., Xu, H., & Pang, W. (2008). GIS and ANN coupling model: An innovative approach to evaluate vulnerability of karst water inrush in coalmines of north China. *Environ. Geol.*, *54*(5), 937–943. https://doi.org/10.1007/s00254-007-0887-3 (cit. on p. 139).

Wunsch, A., Liesch, T., & Broda, S. (2021). Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous

input (NARX). *Hydrol. Earth Syst. Sci.*, *25*(3), 1671–1687. https://doi.org/10.5194/hess-25-1671-2021 (cit. on p. 140).

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., & Goldscheider, N. (2022). Karst spring discharge modeling based on deep learning using spatially distributed input data. *Hydrol. Earth Syst. Sci.*, *26*(9), 2405–2430. https://doi.org/10.5194/hess-26-2405-2022 (cit. on pp. 139, 140, 160, 161).

Xie, Y., Allaire, J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC. (Cit. on p. 162).

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman and Hall/CRC. (Cit. on p. 162).

Yin, D., Shu, L., Chen, X., Wang, Z., & Mohammed, M. E. (2011). Assessment of Sustainable Yield of Karst Water in Huaibei, China. *Water Resour. Manage.*, *25*(1), 287–300. https://doi.org/10.1007/s11269-010-9699-4 (cit. on p. 139).

Zambrano-Bigiarini, M. (2020). *hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series*. https://doi.org/10.5281/zenodo.839854 (cit. on p. 162).

Zhang, J. L., Li, Y. P., Huang, G. H., Wang, C. X., & Cheng, G. H. (2016). Evaluation of Uncertainties in Input Data and Parameters of a Hydrological Model Using a Bayesian Framework: A Case Study of a Snowmelt–Precipitation-Driven Watershed. *Journal of Hydrometeorology*, *17*(8), 2333–2350. https://doi.org/10.1175/JHM-D-15-0236.1 (cit. on p. 158).

Zhou, B.-Q., Yang, Z., Hu, R., Zhao, X.-J., & Chen, Y.-F. (2021). Assessing the impact of tunnelling on karst groundwater balance by using lumped parameter models. *J. Hydrol.*, *599*, 126375. https://doi.org/10.1016/j.jhydrol.2021.126375 (cit. on p. 139).

Conclusions

This thesis aimed to develop and improve numerical methods and hydrological models for characterising the functioning of karst systems in response to climatic and anthropogenic changes, with a focus on conceptual reservoir modelling.

Tools. This work used different databases and approaches for characterising karst systems hydrological functioning:

i. The SNO KARST and World Karst Spring hydrograph (WoKaS) databases were used for the analysis of spring discharge time series (Chapter 2, Chapter 3 and Chapter 4).

ii. Numerical methods such as recession curves, signal and statistical analyses were used as a basis for the classification of the hydrological functioning of karst systems (Chapter 2 and Chapter 3).

iii. The framework of the KARMA project allowed to work on several karst systems across the mediterranean region and to benefit from local expert knowledge (Chapter 5 and Chapter 6).

iv. Lumped parameter approaches, and especially conceptual reservoir models, were used to gain insights into the internal dynamics of karst systems (Chapter 6).

Research questions. Numerical methods and hydrological models have been used to answer the initial research questions:

(i) *How to characterise the functioning of a karst system in data-scarce contexts?*

A data-scarce context implies the study of a system with limited data, whether in terms of spatial coverage, temporal scope, resolution, or diversity. Chapter 2 shows the information about a system functioning that can be retrieved from the analysis of the spring discharge time series. The analysis of recession curves can give relevant insight into different aspects of a karst system hydrological functioning: capacity of dynamic storage, draining dynamic of the capacitive function and variability of the hydrological functioning. The consideration of the variability of the hydrological functioning in the proposed classification is a novel aspect, allowing for a more accurate differentiation of karst systems. The use of multivariate analyses and the application of the classification on large databases have made it possible to identify an optimal number of classes of functioning – six providing a good compromise between the gain in knowledge, relevance, and accessibility of the method. Chapter 2 demonstrates that this classification is relevant given a minimum number of observations that depends on the system dynamics and the variability of climate, which generally remains compatible with a data-scarce context. Chapter 6 shows that reservoir models can provide relevant results with few input data (one-dimensional precipitation and temperature) and, especially, short length time series. Using reservoir models can therefore provide a benefit in a context of data scarcity or help gaining preliminary insights into the functioning of a poorly known karst system (e.g. water partitioning between matrix and conduits and internal flow dynamics).

(ii) *On which aspects can simple, one-dimensional models be further improved?*

One-dimensional models can be improved on several aspects: input data, calibration and evaluation. Chapter 4 proposes several additions to the KarstMod modelling platform on such aspects. In the new version 3.0 of this software, the definition of the parameters for preprocessing input data can be realised by model calibration using optional modules (snow or evapotranspiration), resulting in a more adapted input data. The objective function now supports more than two components, which can help to calibrate a model with respect to several observation variables (spring discharge, piezometry, temporary spring discharge), and using one or several performance criteria, as demonstrated in the Lez case study. Additional tools are also proposed for model evaluation, which can be realised on different aspects of the functioning of a system, i.e. the good reproduction of volumes, variability and correlation. Chapter 5 underlines the importance of choosing relevant performance criteria. They all have their flaws and a one-dimensional simulation can be improved by good modelling practices alone.

(iii) *Can performance criteria be trusted for the calibration and evaluation of hydrological models?*

Chapter 5 evidences the impact of counterbalancing errors using performance criteria for the calibration and evaluation of a hydrological model. Performance criteria cannot be fully trusted as they all have different limitations. They should be carefully selected in regard to the aim of a model, and preferably in a multi-criteria framework.

(iv) *What are the advantages and drawbacks of different lumped parameter modelling approaches in karst hydrology?*

Chapter 6 proposes a comparison of artificial neural networks (ANN) and reservoir modelling approaches. ANN models, a branch of machine learning, show a high flexibility regarding input data and are efficient for reproducing high flows. Reservoir models can work with relatively short time series and seem to accurately reproduce low flows. Both approaches can help characterising different aspects of a system: (i) ANN models for the relevance of input data, catchment delineation, and (ii) reservoir models for the internal functioning of the aquifer.

Key results. The key points of the work are the following:

i. Despite the high heterogeneity and complexity of karst systems, a preliminary characterisation of their hydrological functioning is possible even in data-scarce context. The analysis of discharge time series or the use of one-dimensional lumped parameter models can give a lot of information on system characteristics and internal functioning.

ii. Hydrological models should be applied thoughtfully for the interpretation to be relevant. Each step of a modelling process requires a thorough review of the available options, whether it is for the choice of the modelling approach, the definition of the model structure, the use of input data, the calibration procedure, the objective function, the performance criteria for model evaluation, or the sensitivity analysis of the model parameters. In particular, this thesis shows that performance criteria all have limitations and should be selected with respect to the objectives of the hydrological model and the aim of the study.

Perspectives. Future work may focus on:

i. The impact of climate change on karst systems in the Mediterranean using a consistent methodology across multiple sites with different characteristics in terms of locations, climatic conditions, hydrogeological properties and available data. A current work (recently started during this thesis) takes advantage of the framework of the KARMA project to study the large scale future trends of water dynamics and availability in karst systems in this region, using reservoir models. To account for uncertainties in climate projection, 8 coupled GCM/RCM climate models are considered with two emission scenarios (RCP 4.5 and RCP 8.5). Such study could bring interesting general insights in the response of karst systems to changes in precipitation patterns and under significant warming and drying conditions.

ii. The possible link between the hydrological functioning class of a karst system and the reservoir model structure considered in the simulation of spring discharge and/or water level of karst systems. The identification of structure components (compartments, transfer functions) that may be specific to certain classes could simplify the modelling process while adding value to the classification proposed in Chapter 2.

iii. The use of indicators of functioning (e.g. recession dynamics) for the calibration of hydrological models. Constraining the model to reproduce key aspects of the functioning of the system (capacity of dynamic storage, draining dynamic of the capacitive function, and variability of the hydrological functioning) could result in more relevant model structure and parameters.

iv. The proposition of an indicator to assess the extent of counterbalancing errors in a simulation, which could be used as a tool to evaluate the relevance of the model. Another perspective would be to develop an alternative performance criterion not subject to counterbalancing errors, in order to avoid undesired overestimations and underestimations in the simulations.

v. The coupling of the ANN and reservoir modelling approaches to leverage the advantages of each approach. A possible way would be to use ANN models to bypass the limitations of reservoir models concerning the processing of input data, and then use a reservoir model accounting for the output of the ANN model. This would allow benefiting from the information derived from the model structure and parameters of the reservoir model.

Abstract: As a result of erosion due to infiltration and circulation of acidic waters, karst systems have a particular geomorphological structure which generally consists of a substantial and structured heterogeneity conditioned by past flow conditions. Understanding the functioning of flow and storage processes in karst aquifers thus represents a major challenge – especially since these resources provide drinking water to approximately 9 % of the global population. This thesis aims to develop and improve numerical methods and hydrological modelling for characterising the functioning of karst systems in response to climatic and anthropogenic changes.

A classification of karst systems hydrological functioning is first proposed. Based on the analysis of recession curves, the method allows to categorise a karst system into 6 classes with different characteristics in terms of storage capacity, draining dynamic and hydrological variability. In parallel, a software – KarstID – is developed to perform various discharge analyses as well as the aforementioned classification.

Another aspect of the thesis is dedicated to the evaluation and improvement of lumped parameter models. Benefiting from the framework of the KARMA European project, the accessible data facilitates large-scale and extensive research based on multiple karst systems in the Mediterranean region. Firstly, a comparative study of the performance of two modelling approaches – artificial neural networks and reservoir models – is carried out. The objective is to identify the advantages and disadvantages of each approach, depending on the available data and the objectives of a study. Secondly, the counterbalancing error mechanism on the Kling-Gupta Efficiency and its variants is extensively studied: when calibrating or evaluating a model, counterbalancing errors can artificially lead to a higher criterion score that is not associated with an increase in model relevance. Further work is also focused on the development of new functionalities (related to input data and performance criteria) for KarstMod, a software dedicated to rainfall-runoff modelling for karst systems.

Résumé : Les systèmes karstiques, principalement constitués par des roches carbonatées, présentent une structure géomorphologique particulière résultant de leur érosion par l'infiltration et la circulation d'eaux acides. La compréhension du fonctionnement des processus d'écoulement et de stockage des eaux des aquifères karstiques représente un défi majeur – d'autant plus que ces ressources alimentent en eau potable environ 9 % de la population à l'échelle mondiale. Cette thèse vise à développer et à améliorer les méthodes numériques et la modélisation hydrologique pour caractériser le fonctionnement des systèmes karstiques en réponse aux changements climatiques et anthropiques.

Une classification du fonctionnement des systèmes karstiques est proposée dans un premier temps. Cette classification est basée sur l'analyse des courbes de récession et permet de caractériser un système karstique en six classes sur la base de trois indicateurs, qui reflètent sa capacité de stockage, sa dynamique de drainage et sa variabilité de fonctionnement hydrologique. En parallèle, un logiciel a été développé, permettant de réaliser différentes analyses de débit ainsi que la classification proposée.

Un autre aspect de cette thèse est consacré à l'évaluation et à l'amélioration des modèles à paramètres globaux. Bénéficiant du cadre du projet européen KARMA, les données accessibles facilitent les études large échelle basées sur plusieurs systèmes karstiques dans la région méditerranéenne. Dans un premier temps, une étude comparative de la performance de deux approches de modélisation – les modèles réservoir et à réseaux de neurones artificiels – est réalisée. L'objectif est d'identifier les avantages et les inconvénients de chaque approche, en fonction des données disponibles et des objectifs d'une étude. Dans

un second temps, le mécanisme de compensation des erreurs sur le Kling-Gupta Efficiency (KGE) et ses variantes est étudié : lors de la calibration ou de l'évaluation d'un modèle, les erreurs de compensation peuvent artificiellement conduire à un score de critère plus élevé qui n'est pas nécessairement associé à une meilleure pertinence du modèle. Des travaux supplémentaires ont également porté sur le développement de nouvelles fonctionnalités (liées aux données d'entrée et aux critères de performance) pour KarstMod, un logiciel dédié à la modélisation réservoir pour les systèmes karstiques.