



HAL
open science

Pour une évaluation visuelle des modèles de simulation, au-delà de la face validation

Robin Cura

► **To cite this version:**

Robin Cura. Pour une évaluation visuelle des modèles de simulation, au-delà de la face validation. 16th Spatial Analysis and GEomatics Conference (SAGEO 2021), May 2021, La Rochelle, France. pp.139-150. hal-03312815

HAL Id: hal-03312815

<https://hal.science/hal-03312815v1>

Submitted on 3 Aug 2021

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.

Pour une évaluation visuelle des modèles de simulation, au-delà de la *face validation*.

Robin Cura

UMR Géographie-cités, CNRS
robin.cura@parisgeo.cnrs.fr

RÉSUMÉ. Nous proposons une approche méthodologique dédiée à l'évaluation de modèles de simulation stochastiques descriptifs caractérisés par une faible possibilité de recours à des données empiriques. En l'absence de données structurées de contrôle – ou quand celles-ci sont trop peu nombreuses, incertaines ou lacunaires –, l'évaluation peut être réalisée d'après les connaissances expertes, de manière qualitative. Pour que ce processus soit mené de manière rigoureuse, robuste et reproductible, il ne doit pas être uniquement basé sur des intuitions ou premières impressions. L'évaluation peut prendre la forme d'une comparaison systématique entre les sorties du modèle et les hypothèses thématiques sous-jacentes à la modélisation, tirées des connaissances expertes. Cette comparaison est menée à l'aide d'une méthode d'« évaluation visuelle » que nous définissons et appliquons ici. L'approche méthodologique proposée doit ainsi permettre et faciliter l'évaluation de modèles de simulations par le biais de connaissances expertes explicitées, en s'appuyant sur l'utilisation d'un environnement d'exploration interactive des données de sortie de simulation.

ABSTRACT. In this talk, we propose a new methodological framework, the “visual evaluation”, that enables or facilitates the evaluation of complex and stochastic simulation models that mostly rely on expert knowledge rather than empirical data for their hypotheses. This method has been developed for the modelling of space-time processes and relies on the thorough use of qualitative visual comparison between the model's outputs and the underlying empirical knowledge. Visual evaluation is based on a remediation of face validation for massive simulation datasets, and thus requires an interactive exploratory data analysis platform dedicated to simulation outputs.

MOTS-CLÉS : Simulation multi-agents, méthodes d'évaluation, évaluation visuelle, analyse interactive de données, modélisation en interdisciplinarité

KEYWORDS: Agent-based modelling, evaluation methods, visual evaluation, face validation, Exploratory Data Analysis, interdisciplinary modelling

1 Introduction

Pour comprendre et modéliser une structure spatiale modelée par l'héritage de ses états antérieurs, il est nécessaire d'en comprendre l'émergence, souvent bien antérieure. Comme l'écrit Durand-Dastès (1999, p. 23), l'explication d'un « objet spatial » contemporain n'est en effet possible qu'en montrant « à la fois comment il est né et comment il a persisté ». La modélisation de processus spatiaux génériques en géographie implique donc de comprendre les origines et raisons de leur apparition, et donc de faire appel à des connaissances en dehors du champ strict de la modélisation en géographie. Si, dans cette discipline, la modélisation de dynamiques spatiales à l'aide de modèles de simulation multi-agents est désormais bien ancrée, la « culture modélisatrice » est moins présente dans les champs disciplinaires historiques et archéologiques. La conception d'un modèle de simulation spatial requiert alors d'une part la capacité d'intégrer les connaissances expertes nécessaires via un processus de co-construction (Cura, 2020a, pp. 26–29, 340–346)¹ et d'autre part, de manière inhérente à une démarche de modélisation, la capacité de soumettre le modèle conçu à une évaluation rigoureuse.

1.1 Contexte et cas d'étude

La proposition méthodologique ici développée se veut générique, mais a été développée dans le cadre d'un projet interdisciplinaire de modélisation de transformations des structures spatiales sur le temps long, l'ANR TRANSMONDYN. Au sein de ce projet, un groupe composé de géographes-modélisateurs, d'archéologues et d'historiens a conçu et développé un modèle de simulation, SIMFEODAL, cherchant à reproduire les dynamiques de polarisation, de hiérarchisation et de fixation de l'habitat rural de l'Europe du Nord-Ouest au tournant du 1er millénaire de notre ère, en modélisant les mécanismes de migrations et de concentration de foyers paysans. Le modèle et sa démarche de construction ont fait l'œuvre de publications spécifiques (Cura *et al.*, 2017; Tannier *et al.*, 2020), et nous en donnerons donc ici un descriptif très succinct. SIMFEODAL est un modèle de simulation à base d'agents stochastique, à visée exploratoire, qui suit l'approche « KIDS » formalisée par Edmonds et Moss (2005). Il se classe ainsi dans la catégorie des modèles descriptifs, avec de nombreux types d'agents (foyers paysans, seigneurs, églises paroissiales, pôles d'attraction, etc.) dont les interactions multiscalaires font émerger les dynamiques étudiées à l'aide de mécanismes complexes et nombreux sur une longue durée simulée de 400 ans.

1. Cet article s'appuie sur une thèse de doctorat, et plus spécifiquement sur le matériel conceptuel, méthodologique et technique qui y est développé dans les chapitres 3 et 4 (Cura, 2020a, pp. 107–180, 181–262).

1.2 Objectif

Nous proposons ici une méthode d'évaluation dédiée à des modèles de simulation à base d'agents co-construits selon un paradigme descriptif. La méthode proposée est particulièrement pensée pour l'évaluation de modèles qui s'appuient avant tout sur des connaissances à dire d'experts plus que sur des données structurées. Cette méthode d'évaluation, que nous nommons « évaluation visuelle », constitue une extension d'une étape classique de l'évaluation de modèle : la *face validation*, c'est-à-dire l'évaluation de la plausibilité des sorties d'un modèle. Cette méthode d'évaluation est fondée sur une analyse visuelle et subjective de la plausibilité du modèle, en demandant à l'évaluateur de se conforter à des intuitions sur le comportement attendu. L'utilisation de cette méthode est souvent considérée comme une mesure préalable relevant du bon sens plus que de l'évaluation strictement dite. L'évaluation visuelle étend le processus de *face validation* en systématisant l'analyse des sorties du modèle de manière qualitative. Il s'agit ainsi de définir en amont une grille d'analyse exhaustive, couvrant tous les aspects du modèle. C'est au prisme d'une comparaison systématique entre cette grille de critères attendus et des « indicateurs de sortie de simulation » (indicateurs visuels, quantitatifs ou qualitatifs) que le modèle sera évalué. Pour permettre à un public d'experts diversifié de mener cette évaluation visuelle directement, sans faire appel à un expert modélisateur ou à un analyste de données, nous proposons un outil interactif d'analyse visuelle et de comparaison des simulations effectuées par le biais de leurs données en sortie.

2 Approche conceptuelle : pour une évaluation visuelle

2.1 Évaluer un modèle

Depuis les travaux précurseurs en simulation informatique (Naylor, Finger, 1967 ; Hermann, 1967), la plupart des chercheurs mettent en avant qu'un modèle de simulation non évalué n'a ni utilité – pour Naylor et Finger (1967) notamment –, ni validité. Sans caricaturer ces écrits, on peut noter que tous cantonnent les modèles non évalués à des « jeux » ou encore, pour les plus modérés, à des outils pédagogiques. Nous ambitionnons de rendre possible l'évaluation de modèles pour lesquelles les méthodes plus classiques sont soit inadaptées, soit insatisfaisantes.

2.1.1 Différentes approches d'évaluation

Le terme d'évaluation en lui-même porte une certaine polysémie et peut être désambiguïsé, en reprenant les termes de Amblard *et al.* (2006), en « validation interne » et « validation externe ». Dans le premier cas, il s'agit de s'assurer de la bonne conception du modèle – « *building the model right* » (Balci, 1994, p. 123) – et dans le second, de montrer que le modèle est adapté à ce qu'il cherche à représenter – « *building the right model* » (Balci, 1994, p. 121). Ces

deux composantes permettent de définir une procédure générale d'évaluation qui peut être résumée entre des cycles d'allers-retours et d'itérations entre les hypothèses du modèle, la validation interne de l'implémentation qui en est faite, et la validation externe des sorties produites (Klügl, 2008). Pour chacune de ces composantes, de nombreuses méthodes existent, sont généralement préconisées dans l'une ou l'autre des étapes de l'évaluation, et peuvent être classées en différentes catégories, selon qu'elles utilisent une approche formelle, statique, dynamique ou encore informelle (Balci, 1998).

2.1.2 *Quelle évaluation pour SIMFEODAL ?*

Toutes les méthodes ne sont pas pour autant applicables à tous les modèles. Par exemple, les méthodes formelles sont souvent peu adaptées à l'évaluation de modèle de simulation à base d'agents. On emploie en effet généralement ces modèles parce qu'ils encouragent une approche anthropomorphique, plus aisément compréhensible et requérant moins de connaissances mathématiques que d'autres approches, mais aussi car il est extrêmement difficile d'exprimer des systèmes dotés de multiples interactions, qui plus est multi-scalaires, sous forme de réseaux d'équations. Les méthodes statiques ne sont aussi que partiellement applicables, et renvoient davantage à la validation interne des modèles, via une vérification logique des interactions entre les mécanismes des modèles sur un plan algorithmique. Pour la validation externe, le plus classique est de faire appel aux méthodes « dynamiques » au sein d'une étape communément nommée validation statistique. Il s'agit de confronter quantitativement les données produites par le modèle aux données empiriques qu'elles cherchent à reproduire, c'est-à-dire à s'assurer de la qualité de l'ajustement des données simulées. Ces méthodes dynamiques s'appuient donc intrinsèquement sur la présence de données empiriques directement comparables avec les sorties du modèle, chose commune dans la modélisation de phénomènes contemporains ou récents.

Dans le cas de SIMFEODAL, et plus généralement des modèles de processus anciens et inscrits dans la longue durée, les méthodes dynamiques sont difficilement mobilisables. Les données empiriques disponibles sont en effet des traces, incertaines et peu précises de par leur ancienneté, et extrêmement incomplètes (absentes et lacunaires) de par la durée de 400 ans modélisée (d'après la typologie de De Runz, 2008). SIMFEODAL est fondé avant tout sur des hypothèses thématiques et des connaissances expertes, et les méthodes dynamiques ne sont alors pas exploitables. Dans cette situation, les théoriciens de l'évaluation recommandent, à défaut, de tout de même faire appel aux méthodes informelles (Petty, 2010), telle que la « *face validation* » qui est la plus couramment mobilisée.

2.2 *De la face validation à l'évaluation visuelle*

La *face validity* a été définie dès les débuts de la simulation comme une « impression initiale du réalisme d'un [modèle] » (Hermann, 1967, p. 221). Klügl

(2008) en donne une définition plus explicite² et dresse une typologie des temps et modes d'intervention de la *face validation*, c'est-à-dire du processus d'évaluation permettant de juger de la *face validity* d'un modèle. Cette méthode a l'avantage de pouvoir être mobilisée dans toutes les étapes du cycle d'évaluation (Balci, 1998) et de systématiquement mettre en avant les connaissances expertes des modélisateurs et thématiciens sur l'objet modélisé.

2.2.1 *Limites de la face validation et propositions de résolution.*

C'est toutefois une méthode assez critiquée par les théoriciens de l'évaluation (Hermann, 1967 ; Balci, 1994), qui considèrent que la *face validation* ne peut qu'être une étape préalable à des méthodes d'évaluation plus quantitatives et formelles. Parmi ces auteurs, Hermann (1967, p. 222) en pointe plusieurs limites – à partir desquelles nous construisons notre proposition d'évaluation visuelle – que l'on peut résumer ainsi.

Un manque de connaissance experte. Pour Hermann, la *face validation* a le défaut d'être trop accessible et de pouvoir être menée par n'importe qui, y compris par des gens qui n'auraient qu'une expérience limitée des phénomènes modélisés. Cette remarque est généralisable, et s'applique d'ailleurs quelle que soit la méthode d'évaluation : une expertise ne vaut que par la qualité de l'expert impliqué. Hermann considère ainsi que c'est au modélisateur uniquement de mener cette phase d'évaluation. Il se positionne en cela dans une logique de séparation forte entre les acteurs de l'élaboration d'un modèle : les modélisateurs conçoivent et évaluent le modèle seuls, et les thématiciens n'en sont qu'utilisateurs. L'évaluation visuelle est conçue en opposition à ce principe, et ne peut être pensée sans une collaboration étroite entre les porteurs de connaissances thématiques expertes sur le phénomène modélisé et les porteurs d'une connaissance méthodologique experte sur les manières de le modéliser.

Invraisemblance de certains comportements. Hermann met ensuite en avant le fait que dans un modèle, tous les mécanismes n'ont pas vocation à être vraisemblables et qu'une recherche de plausibilité de tels mécanismes fausserait l'analyse du modèle. Cela revient à ignorer l'importance du dialogue entre modélisateur et évaluateur. Si le modélisateur connaît les « substitutions » opérées dans le modèle, il pourra se garder de chercher à juger de leur vraisemblance. À l'inverse, un expert thématicien pourrait être étonné par certains comportements microscopiques si et seulement si les correspondances entre éléments du modèle et éléments du système modélisé n'étaient pas explicitées. Plus généralement, cette limite est présente s'il n'y a pas d'explicitation des « substitutions » implémentées, c'est-à-dire de la correspondance entre le do-

2. « Face validity can be seen as the result of face validation. Under this paradigm I want to subsume all methods that rely on natural human intelligence. Examples are structured walk-through, expert assessments of descriptions, animations or results. Thus, face validity shows that processes and outcomes are reasonable and plausible within the frame of theoretic basis and implicit knowledge of system experts or stake-holder », Klügl, 2008, pp. 39–41.

maine conceptuel et le domaine du modèle implémenté : le niveau attendu de plausibilité individuelle de chaque comportement doit être décidé pendant la construction du modèle. L'explicitation des choix de modélisation, et l'assurance que leurs enjeux et effets sont bien compris de tous les membres d'un projet de modélisation, sont donc des préalables indispensables à une évaluation visuelle.

Besoin d'explicitation préalable des objectifs. Hermann note enfin que la *face validation* ne peut constituer une méthode d'évaluation adaptée si l'on ne spécifie pas, en amont, les critères qu'elle doit s'attacher à examiner. Le risque est alors d'introduire encore plus de subjectivité dans cette analyse, et en particulier de briser la capacité de reproductibilité ou de justification d'une évaluation. Quand un modèle résulte d'un travail collaboratif les évaluations d'un même résultat peuvent en effet varier, et doivent alors être explicitées autant que possible, ne serait-ce que pour se doter d'un outil de discussion et de débat commun. Il apparaît alors primordial de définir en amont de la modélisation les critères d'évaluation. Cela ne limite aucunement la nécessaire subjectivité et complémentarité des évaluateurs experts, mais permet au contraire d'inscrire leurs discours dans un référentiel de comparabilité. Dès lors, en matière d'évaluation visuelle, la mise au point d'une grille d'analyse systématique nous semble indispensable. Celle-ci doit expliciter, en amont des analyses, l'ensemble des éléments qui seront analysés dans le modèle, les « substitutions » dans le domaine empirique auxquelles ils ont trait, et dès lors, les comportements attendus pour chacun.

2.2.2 Définition et pré-requis de l'évaluation visuelle

C'est à partir de ces remarques que l'on peut proposer une définition de l'évaluation visuelle, qui consiste alors à évaluer la plausibilité du comportement d'un modèle à partir des données qu'il produit. Cette évaluation s'appuie sur une comparaison systématique menée entre 1) les sorties du modèle et 2) les connaissances expertes sur les processus modélisés. Les critères de cette comparaison sont définis en amont de l'évaluation, et doivent expliciter pour chaque sortie analysée la ou les formes qu'elle devrait prendre. Ces éléments, les critères d'évaluation, ne peuvent être formulés par n'importe qui : si le modélisateur autant qu'un expert externe peuvent les spécifier, il convient de s'assurer de l'expertise – thématique et de la connaissance des processus tels que modélisés – de l'évaluateur. L'ensemble des critères d'évaluation forment alors une grille d'analyse qualitative, et la visualisation des différentes sorties du modèle permet de la compléter. Il devient alors possible d'apprécier globalement l'écart entre le modèle et les processus empiriques qu'il représente, sans chercher pour autant à quantifier ni à mesurer cet écart.

2.2.3 Composition de l'évaluation visuelle

Pour établir la correspondance entre le domaine empirique et celui du modèle, nous proposons une dichotomie simple en systématisant l'usage d'un vo-

cabulaire souvent employé de manière plurielle. Nous caractérisons ces éléments en deux grands ensembles : les indices empiriques et les indicateurs de sortie.

Afin d'évaluer la capacité du modèle à reproduire un phénomène observé, il est nécessaire de disposer dans le domaine empirique de « points de repère », qui doivent relever de l'ensemble des échelles spatiales et temporelles que l'on choisit de mettre en scène. Chacun doit pouvoir être estimé, c'est-à-dire être en capacité d'être reproduits et comparables avec d'autres mesures. Nous les qualifions d'« **indices empiriques** » et les regroupons en deux catégories basées sur la précision avec laquelle ils peuvent être décrits : les « ordres de grandeur », mono-dimensionnels et d'une certitude variable, et les « formes stylisées », qui reposent sur l'agrégation ou la composition de plusieurs ordres de grandeur.

Dans le modèle de simulation, il est nécessaire de définir des « **indicateurs de sortie** », c'est-à-dire des variables informatiques produites par le modèle que l'on pourra ensuite comparer aux « indices empiriques » définis. Comme pour les indices empiriques, les indicateurs de sortie de simulation peuvent être distingués en des formes numériques simples (des scalaires), et en des plus complexes, les indicateurs multi-dimensionnels, qui peuvent être mis en correspondance avec les « formes stylisées » du domaine empirique.

On peut illustrer cette typologie et sa **mise en correspondance** avec un exemple tiré de SIMFEODAL . Dans le domaine empirique, on sait que la part de l'habitat rural dispersé a diminué au cours de la période étudiée, depuis environ 95% en 800 jusqu'à environ 20% en 1200. Ces deux valeurs sont des indices empiriques, de type « ordre de grandeur ». Dans le modèle de simulation, un indicateur de sortie comparable est calculé et correspond au taux de foyers paysans isolés (fig. 2, C). Le taux de foyers paysans isolés en début de simulation est un indicateur de sortie scalaire, de même que celui en fin de simulation. Dans le domaine empirique, la part de l'habitat dispersé a diminué régulièrement au cours du temps (forme stylisée), et on évaluera donc la capacité du modèle à reproduire un processus de concentration semblable. La courbe d'évolution du taux de foyers paysans isolés au cours de la simulation – constituée d'une série d'indicateurs de sortie « scalaires » – constitue alors, dans le domaine du modèle, un indicateur de sortie multi-dimensionnel que l'on pourra analyser : si la courbe part de 95% en début de simulation pour atteindre 20% en fin de simulation, avec une diminution régulière, alors l'indicateur de sortie multi-dimensionnel sera en adéquation avec l'indice empirique de type forme stylisée « évolution de la part de l'habitat rural dispersé ». En définissant en amont de la modélisation un ensemble d'indices empiriques, d'indicateurs de sortie correspondants ainsi que des propriétés attendues des ces derniers, on constitue alors une « grille d'analyse » qui permettra de mener l'évaluation visuelle du modèle.

3 Mise en œuvre pratique de la démarche d'évaluation visuelle

Dans le cas de SIMFEODAL, nous avons identifié une cinquantaine d'indices empiriques utiles à mobiliser, c'est-à-dire à comparer systématiquement avec les indicateurs de sortie de simulation correspondant (ces indices sont présentés, ainsi que leur forme attendue, dans Cura, 2020a, pp. 142–150). Cette masse s'explique notamment par le fait que, dans ce modèle descriptif, près d'une dizaine de types d'agents coexistent et interagissent. Il est dès lors important d'observer les comportements (au moins agrégés) de ces différents types d'agents. On peut citer comme exemple le suivi temporel du taux de foyers paysans dispersés, la répartition spatiale des agrégats de population (villes, villages, etc.), l'évolution du maillage des paroisses ou encore le suivi du nombre de châteaux créés. Ces différents éléments permettent d'appréhender globalement la multi-dimensionnalité de la transformation de la société entre 800 et 1200, et sont nécessaires à la bonne compréhension de sa composante spatiale qui constitue le cœur d'intérêt de SIMFEODAL. La multiplication des indicateurs de sortie, nécessaire à une bonne appréhension de la diversité des comportements d'un modèle, pose alors, en premier lieu, une difficulté concrète ne serait-ce que pour les visualiser.

3.1 *Visualiser de nombreux indicateurs de sortie de simulation*

La masse des indicateurs de sortie (voir 2.2.3) rend, en effet, peu évidente la consultation « en direct » de l'exécution du modèle. De plus, en raison de la stochasticité du modèle, il est nécessaire d'effectuer plusieurs répliques de chaque simulation afin d'en visualiser la tendance agrégée. Le nombre d'indicateurs ainsi générés est important, et nous avons choisi, pour en garantir une consultation aisée, d'organiser ces indicateurs dans un outil interactif dédié à la mise en œuvre de l'évaluation visuelle. La visualisation doit donc se faire a posteriori, et implique dès lors de mener une collecte importante des données générées.

3.1.1 *Une collecte de toutes les données de sortie de simulation*

Cette collecte doit être aussi large que possible, afin d'être en mesure de fournir une base de calcul à l'ensemble des indicateurs de sortie de simulation convenus. Au-delà même des indicateurs de sortie définis, et afin de garantir une reproductibilité des analyses, il est nécessaire de veiller à généraliser la collecte des données produites par le modèle, c'est-à-dire d'enregistrer aussi les données qui ne devraient pas être étudiées. En effet, pour faire face à de potentielles futures modifications du modèle ou assurer une rétro-compatibilité complète, il est obligatoire de disposer des données fines dont l'agrégation permet le calcul des indicateurs et leur visualisation. De plus, au fur et à mesure de l'évolution du modèle, le nombre d'exécution du modèle dont il faut pouvoir conserver les données augmente. Par conséquent, la masse de données produites augmente d'autant, et devient ainsi rapidement trop importante pour être gérée de manière manuelle. Dans le cas de SIMFEODAL par exemple, la masse totale de

données collectées sur le modèle au cours des nombreuses étapes de son évolution – plusieurs milliers de simulations – représente des dizaines de millions d’enregistrements en base de données.

3.1.2 Gestion et organisation d’une importante masse de données

Pour assurer l’universalité et la simplicité de l’interrogation de ces données massives, nous avons choisi de les organiser au sein d’une base de données relationnelle. La masse de données imposait des contraintes fortes en matière de performance des requêtes, notamment en raison de notre choix de calculer les indicateurs « à la volée ». L’interrogation et l’agrégation de millions d’enregistrements dans un temps acceptable pour une visualisation interactive nous a poussé à réaliser un *benchmark* de systèmes de gestion de bases de données (SGBD) relationnels performants (Cura, 2020a, fig. 4.10, p. 226). En raison de la généricité d’interrogation (SQL) et de l’extrême rapidité de traitement des requêtes qu’il permet, nous avons choisi de mobiliser un SGBD « analytique » libre, orienté colonne et haute-performance, OmniSciDB (Root, Mostak, 2016).

3.1.3 Mise en place d’un pipeline de visualisation

Afin de garantir une certaine reproductibilité à nos analyses, nous avons opté pour une chaîne de traitement basée sur des solutions libres et programmables plutôt que sur des outils graphiques, afin que le partage des codes-sources (et des données) permette de reproduire les visualisations de manière stricte. La production graphique des indicateurs de sortie de simulation s’inscrit ainsi dans un « pipeline de visualisation » (fig. 1) formalisé avec le langage libre R, qui orchestre l’interrogation des données, leur filtrage, la production des visualisations (carto)graphiques et tabulaires, ainsi que leur organisation et présentation au sein d’une plateforme interactive dédié, SIMEDB (fig. 2).

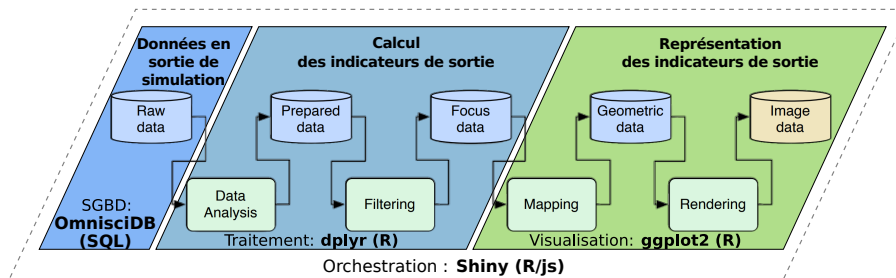


FIGURE 1 – Le « pipeline » de visualisation (d’après Keim *et al.*, 2010) et son implémentation dans SIMEDB.

3.2 Comparer visuellement des simulations

La co-construction d’un modèle de simulation s’appuie sur une itération fréquente entre évolutions du modèle conceptuel, implémentation et évaluation des sorties. Chaque version du modèle entraîne en effet des modifications dans

les valeurs de paramètres ou dans le détail des mécanismes, et il est donc important de pouvoir comparer les effets de ces changements. Pour appliquer une démarche d'évaluation visuelle, il est alors nécessaire de procéder régulièrement à une comparaison visuelle entre les simulations.

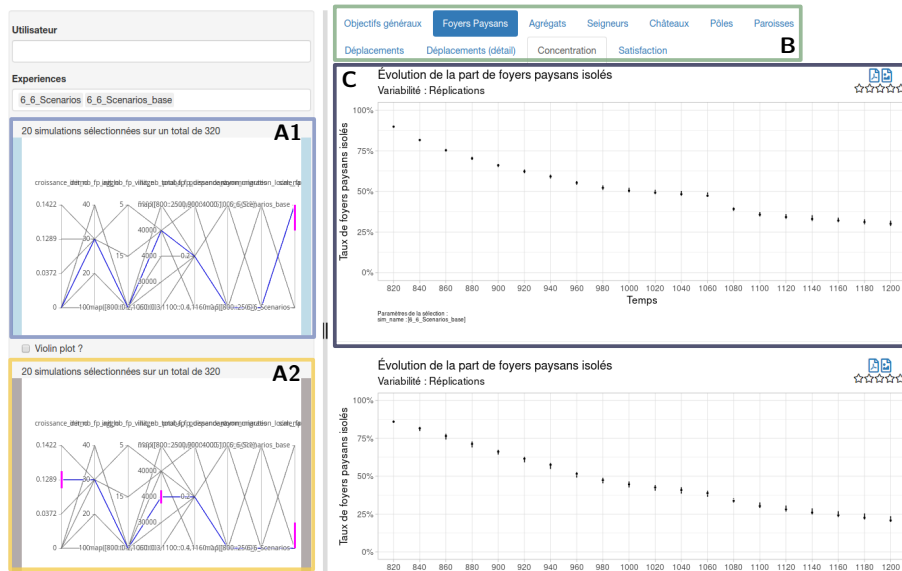


FIGURE 2 – Interface de la plateforme SIMEDB.

3.2.1 Visualiser les différences entre simulations

Dans la définition de l'évaluation visuelle (2.2.2), nous explicitons le besoin de recourir à une comparaison entre les indicateurs de sortie du modèle et des formes attendues d'indices empiriques exprimées dans une grille d'analyse. À mesure que la construction du modèle évolue (calibration notamment), les écarts entre ces valeurs diminuent, jusqu'à atteindre des valeurs trop faibles pour être jugées thématiquement significatives. Il devient alors nécessaire de mener une évaluation plus fine, en comparant directement les indicateurs de sortie des simulations les unes avec les autres (évaluation relative) plutôt qu'avec le référentiel absolu que constituent les indices empiriques. SIMEDB est donc orientée vers une approche comparative plus que normative : si une partie des indices empiriques sont figurés dans certaines visualisations, l'accent est majoritairement placé sur la visualisation des différences entre simulations.

3.2.2 De la visualisation à la comparaison interactive

Au cours de son « cycle de vie », le modèle SIMFEODAL est passé par sept versions majeures et près d'une cinquantaine de sous-versions, sans compter les essais de variations de jeux de paramètres au sein même de ces sous-versions. L'identification des configurations du modèle (versions concurrentes, valeurs

des jeux de paramètres mobilisés) devenait alors un problème en elle-même. Nous avons alors choisi de renforcer l’interactivité de SimEDB, en permettant notamment une sélection des simulations à comparer par un outil graphique. Dans l’interface de SimEDB, le choix des simulations dont on souhaite visualiser les indicateurs de sortie se fait ainsi via une sélection graphique des valeurs de paramètres à isoler, via un *brushing* sur des graphiques interactifs en coordonnées parallèles (fig. 2, A1 et A2). En répétant ces opérations de « brushing » dans l’interface, on peut ainsi comparer rapidement les indicateurs de sortie des simulations (accessibles via le double-menu d’onglets B). L’outil étant interactif, nous préférons le présenter sous forme de vidéo plutôt que d’en réaliser une description textuelle forcément moins adaptée. Une présentation de SIMEDB est ainsi disponible dans un dépôt dédié (Cura, 2020b).

4 Conclusion

Nous proposons une méthode d’évaluation de modèles de simulation stochastiques, descriptifs, et fondés sur la connaissance experte plus que sur des données empiriques. Cette méthode en elle-même a été conçue, raffinée et mise en œuvre collectivement dans le cadre de la co-construction d’un modèle de simulation ancré dans un passé lointain et sur le temps long. Au vu des évolutions du modèle SIMFEODAL que l’évaluation visuelle a accompagné, nous pouvons dire que le résultat de cette démarche est concluant. Le modèle produit des résultats comparables aux attentes que nous en avons et donc conforme aux connaissances empiriques sur lesquelles il s’appuie. L’exécution systématique de ce processus d’évaluation visuelle a facilité la réalisation d’allers-retours nombreux et réguliers entre le modèle conceptuel, son implémentation générale, le détail technique des mécanismes et les valeurs de paramètres testées. Les dernières versions du modèle, les mieux calibrées, nous permettent ainsi de valider la plausibilité des hypothèses thématiques que le modèle cherchait à expérimenter. La robustesse de SIMFEODAL a enfin été testée par l’application d’une analyse de sensibilité, elle aussi menée de manière visuelle (Cura, 2020a, pp. 290–322). Il nous semble donc que l’évaluation visuelle, enrichissant l’étape préalable et peu formalisée qu’est la *face validation*, est une méthode tout à fait valable pour l’évaluation de modèles de simulations qui se prêteraient mal aux méthodes plus formelles. Plus généralement, nous gageons que la construction et l’évaluation de tous les types de modèles de simulation gagneraient à mobiliser plus fréquemment des méthodes visuelles.

Bibliographie

- Amblard F., Rouchier J., Bommel P. (2006). Evaluation et validation de modèles multi-agents. *Modélisation et simulation multi-agents. Applications pour les Sciences de l’Homme et de la Société*, Hermès, p. 103–140.

- Balci O. (1994). Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of operations research*, vol. 53, n° 1, p. 121–173.
- Balci O. (1998). Verification, validation, and accreditation. In *Proceedings of the 1998 WSC*, vol. 1, p. 41–48.
- Cura R. (2020a). *Modéliser des systèmes de peuplement en interdisciplinarité : co-construction et exploration visuelle d'un modèle de simulation*. Thèse de doctorat non publiée, Université Panthéon-Sorbonne. (<https://tel.archives-ouvertes.fr/tel-02944405>)
- Cura R. (2020b). *Présentation de SimEDB*. DOI: 10.5281/zenodo.4554338.
- Cura R., Tannier C., Leturcq S., Zadora-Rio E., Lorans E., Rodier X. (2017). Transition 8 : 800-1100. Fixation, polarisation et hiérarchisation de l'habitat rural en Europe du Nord-Ouest. In L. Sanders (Ed.), *Peupler la terre - De la préhistoire à l'ère des métropoles*, p. 301–332. Presses universitaires François Rabelais.
- De Runz C. (2008). *Imperfection, temps et espace : modélisation, analyse et visualisation dans un SIG archéologique*. Thèse de doctorat non publiée, Université de Reims - Champagne Ardenne. (<https://tel.archives-ouvertes.fr/tel-00560668>)
- Durand-Dastès F. (1999). Jamais deux fois... Ou : de quelques précautions à prendre avec le temps. *Travaux de l'Institut de Géographie de Reims*, vol. 26, n° 101, p. 5–23.
- Edmonds B., Moss S. (2005). From KISS to KIDS – An ‘Anti-simplistic’ Modelling Approach. In P. Davidsson, B. Logan, K. Takadama (Eds.), *International workshop on multi-agent systems and agent-based simulation*, p. 130–144. Springer Berlin Heidelberg.
- Hermann C. (1967). Validation Problems in Games and Simulations with Special Reference to Models of International Politics. *Behavioral Science*, vol. 12, n° 3.
- Keim D., Kohlhammer J., Ellis G., Mansmann F. (2010). *Mastering the Information Age Solving Problems with Visual Analytics*. Eurographics Association.
- Klügl F. (2008). A Validation Methodology for Agent-based Simulations. In *Proceedings of the 2008 ACM Symposium on Applied Computing*, p. 39–43. New York, NY, USA, ACM.
- Naylor T. H., Finger J. M. (1967). Verification of Computer Simulation Models. *Management Science*, vol. 14, n° 2, p. B92-B106.
- Petty M. D. (2010). Verification, validation, and accreditation. *Modeling and simulation fundamentals: Theoretical underpinnings and practical domains*, p. 325–372.
- Root C., Mostak T. (2016). MapD: A GPU-powered Big Data Analytics and Visualization Platform. In *ACM SIGGRAPH 2016 Talks*, p. 73:1–73:2. New York, NY, USA, ACM.
- Tannier C., Cura R., Leturcq S., Zadora-Rio E. (2020). An agent-based model for exploring the combined effects of social and demographic changes on the concentration and hierarchy of rural settlement patterns in North-Western Europe during the Middle Ages (800–1200 CE). *Journal of Anthropological Archaeology*, vol. 59, p. 101204.