HAL open science

Assessment of the single-ensemble method applied to hydrodynamic simulations

J. Prats, P.A. Danis

To cite this version:

J. Prats, P.A. Danis. Assessment of the single-ensemble method applied to hydrodynamic simulations. Limnetica, 2017, 36 (1), pp.1-14. 10.23818/limn.36.01 . hal-01556452

HAL Id: hal-01556452

https://hal.science/hal-01556452v1

Submitted on 5 Jul 2017

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.

Limnetica, 36 (1): 1-14 (2017). DOI: 10.23818/limn.36.01
© Asociación Ibérica de Limnología, Madrid. Spain. ISSN: 0213-8409

Assessment of the single-ensemble method applied to hydrodynamic simulations

Jordi Prats[1,*] and Pierre-Alain Danis[2]

[1] Irstea, UR RECOVER, Pôle AFB-Irstea "Hydroécologie Plans d'eau", centre of Aix-en-Provence, 3275 Route de Cézanne, F-13182 Aix-en-Provence, France.
[2] Agence Française pour la Biodiversité, Pôle AFB-Irstea "Hydroécologie Plans d'eau", UR RECOVER, centre of Aix-en-Provence, 3275 Route de Cézanne, F-13182 Aix-en-Provence, France.

[*] Corresponding author: jordi.prats@irstea.fr

Received: 05/11/15 Accepted: 21/09/16

ABSTRACT

Assessment of the single-ensemble method applied to hydrodynamic simulations

Averaging the results of different models and/or model realizations has been suggested as a way to improve the estimations of environmental models, especially in meteorology and climate sciences. We applied ensemble modelling to study the hydrodynamic and thermal behaviour of the Mediterranean reservoir of Bimont (France). We used a single-model ensemble composed of three realizations of the same hydrodynamic model (Dynamic Reservoir Simulation Model, DYRESM) with different calibration parameter values as follows: 1) default parameter values, 2) parameter values obtained by manual calibration, and 3) parameter values corresponding to the best performing member of a set of behavioural parameter sets obtained using the GLUE (Generalized Likelihood Uncertainty Estimation) method. The results confirmed the good performance of the ensemble average for lake water temperature simulations. A loss of signal was present because of the smoothing induced by model averaging, especially for the simulation of surface temperature. Single ensembles are a good alternative to obtain consistently better simulation results in hydrodynamic simulations, as well as providing a qualitative estimation of uncertainty. However, awareness of its limitations is necessary when interpreting the results.

Key words: Hydrodynamic model, single-model ensemble, water temperature.

RESUMEN

Evaluación del método de modelización por conjuntos unimodelo aplicado a simulaciones hidrodinámicas

Para mejorar las estimaciones de los modelos ambientales, se ha sugerido la utilización de la media de las simulaciones de diferentes modelos y/o de diferentes realizaciones de un modelo, especialmente en meteorología y ciencias del clima. En este estudio aplicamos la modelización de conjuntos para estudiar el comportamiento térmico e hidrodinámico del embalse mediterráneo de Bimont (Francia). Utilizamos un conjunto unimodelo formado por tres realizaciones de un mismo modelo hidrodinámico (DYRESM) con diferentes valores de los parámetros de calibración: 1) valores de los parámetros por defecto, 2) valores de los parámetros obtenidos por calibración manual, y 3) valores de los parámetros correspondientes al miembro más eficiente de un conjunto de combinaciones aceptables de parámetros obtenido utilizando la técnica GLUE (Generalized Likelihood Uncertainty Estimation). Los resultados confirman el buen comportamiento de la media de conjunto a la hora de estimar la temperatura del agua en masas de agua continentales. Se dio una cierta pérdida de la señal a causa del suavizado inducido por el promediado de los diferentes miembros del conjunto, especialmente para el caso de la simulación de la temperatura superficial. Los modelos de conjuntos son una buena alternativa para obtener resultados de las simulaciones consistentemente mejores, así como para obtener una estimación qualitativa de su incertidumbre. Sin embargo, es necesario ser conscientes de las limitaciones del método al interpretar los resultados.

Palabras clave: *Modelo hidrodinámico, modelización por conjuntos, temperatura del agua.*

INTRODUCTION

On April 22, 2016, world leaders signed the Paris climate agreement, by which they agreed to limit global warming to between 1.5 °C and 2.0 °C (Tollefson, 2016). Predicting the ecological consequences of such climate warming through models is necessary for optimal management. The simulation of the evolution of a lentic ecosystem depends on the good simulation of hydrodynamics and water temperature since errors in the prediction of the physical properties of the system propagate to the prediction of biological variables (Rigosi & Rueda, 2012). The available lake models simulate water temperatures with reasonable accuracy, usually with overall root mean square errors of 0.5-2.5 °C (Fang *et al.*, 2012; Rigosi & Rueda, 2012; Stepanenko *et al.*, 2014), even when using default parameter values (Read *et al.*, 2014). However, punctual discrepancies in certain moments of the simulations and/or certain points of the water column can attain several degrees (Stepanenko *et al.*, 2013; Read *et al.*, 2014). The errors made in the simulation of hypolimnion temperatures are often higher than those of the epilimnion (Read *et al.*, 2014; Prats & Danis, 2015), probably reflecting difficulties in the simulation of mixing processes (Stepanenko *et al.*, 2013). In relation to this, difficulties also arise in the prediction of certain hydrodynamic properties such as the thermocline depth, the end of the stratification period or the duration of the ice cover, which have a direct influence on the functioning of the ecosystem. It is thus essential to improve the quality of hydrodynamic predictions based on models and to take into account the uncertainty of the simulations when making forecasts.

Combining forecasts, or ensemble modelling, has beensuggested as a way to improve forecasting accuracy since the 1960s (Clemen, 1989). The success of model ensembles has been empirically demonstrated in many fields (Clemen, 1989; Armstrong, 2001) and has been recommended in ecological modelling (Mooij *et al.*, 2010). The combination of multiple individual estimations is usually more accurate than individual estimations, and simple combination methods such as model averaging often outperform more complex methods (e.g., weighted averaging) (Clemen, 1989; Armstrong, 2001; Hagedorn *et al.*, 2005; Knutti *et al.*, 2010). However, model ensembles were only recently applied to aquatic ecosystems (Gal *et al.*, 2014; Trolle *et al.*, 2014).

Although Clemen (1989) defended the unnecessary need to further justify the methodology of combining forecasts, its validity is still contested (Armstrong, 2001; Hagedorn *et al.*, 2005). Some objections are based on the assumption that there is a right way to make predictions or that it would be preferable to develop a more comprehensive model that takes into account other relevant processes (Armstrong, 2001). Instead, ensemble modelling is based on the assumption that no model can perfectly identify all of the relevant processes and that different models capture different aspects of the studied system (Clemen, 1989).

There is another objection regarding the statistical interpretation of ensemble modelling results (Armstrong, 2001; Stephenson *et al.*, 2012). In particular, a solid theoretical basis, in the form of an appropriate statistical framework, for multimodel ensembles is lacking (Stephenson *et al.*, 2012). Knutti *et al.* (2010) mentioned other challenges in combining model projections, such as loss of signal.

Such objections make it necessary to analyse the strengths and weaknesses of ensembles when applied to aquatic ecosystem modelling. Thus, this study analysed the performance of a single-model ensemble in the simulation of the hydrodynamic behaviour of a small Mediterranean reservoir. The single-model ensemble consisted of different realizations of simulations made with the model DYRESM and three sets of calibration parameter values. The unweighted average simulation results were compared to the individual simulation results of water temperature in the vertical dimension and in the temporal dimension.

METHODS

Study area

The reservoir of Bimont is located in the commune of Saint-Marc-Jaumegarde in the Provence region (South of France). It is part of the water management system of the Société du Canal de Provence (SCP) that distributes drinking and irrigation water to the region of Provence. The reservoir is also used for flood lamination and hydroelectric production.

This valley reservoir is delimited by an arch dam, with a maximum height of 87.5 m and length of 355 m. The reservoir has two main tributaries, an artificial and a natural one. The catchment basin area is $41 \cdot 10^6$ m^2, of which $27 \cdot 10^6$ m^2 correspond to the main natural tributary, La Cause stream. In the artificial inlet structure, water is released into the reservoir through a Ø 1.3 m outlet at an elevation of 315 m.a.s.l. after passing through a reversible pump-turbine, or it is diverted through a bypass gate to a chute at 331 m.a.s.l. The dam has two bottom outlets at 288 m.a.s.l. (Ø 1.5 m) and 287 m.a.s.l. (Ø 0.5 m) and a spillway at 336 m.a.s.l. At the normal exploitation level of 329.5 m.a.s.l., the water depth is 55 m and the reservoir contains a volume of $14 \cdot 10^6$ m^3. The residual volume below the bottom outlets is just 8 m^3.

From 1992 to 2012, the mean flow of the artificial tributary was 1.5 m^3/s, and the mean flow of the main natural tributary was 0.15 m^3/s, so that residence time of the reservoir was approximately 3 months. To maintain a constant water level, the outflow volumes usually equalled the inflow volumes minus evaporation. The reservoir is located in an area of karstic geology and suffers from infiltration problems quantified as 0.2-0.3 m^3/s (Société du Canal de Provence, 2013).

Hydrodynamic model

We used version v.4.0.0-b2 of the hydrodynamic lake model DYRESM (Dynamic Reservoir Simulation Model) developed by the Water Research Centre of the University of Western Australia (Antenucci & Imerito, 2000), which is a good performing model that has been applied to many lakes and reservoirs around the world (e.g., Han *et al.*, 2000; Gal *et al.*, 2003; Perroud *et al.*, 2009). DYRESM is a one-dimensional turbulence model that uses a Lagrangian approach with a variable layer thickness. The model simulates temperature, salinity and density by considering the most important physical processes that affect the vertical density structure of lakes and reservoirs: exchanges of heat, mass and momentum at the water-air interface, surface and bottom water mixing, and inflow and outflow dynamics. The model is applicable when the mixing processes in the horizontal dimension are more important than those in the vertical dimension.

Input data

The necessary meteorological input data to run DYRESM is as follows: air temperature and relative humidity at 2 m above soil level, wind speed at 10 m above soil level, precipitation, solar radiation and cloud cover. Meteorological data for the period 2006-2014 were provided by the French meteorological service, Météo-France, for the meteorological station of Aix-en-Provence (14 km west of the reservoir, 173 m.a.s.l., 43°32′N 5°25′E). An adiabatic gradient correction of −0.006 °C/m was initially applied to the air temperature measurements to take into account the difference in altitude between the Aix-en-Provence meteorological station and the Bimont reservoir. Since cloud cover data were not available for the Aix-en-Provence station, cloud cover data were obtained from the Marseille meteorological station (43°19′N 5°29′E).

The SCP provided measured daily inflow and outflow data. Infiltration was considered constant and was adjusted to close the hydrologic budget. The SCP also provided water temperature data for the artificial tributary. Water temperature at the Cause stream was estimated through a sigmoid air temperature-water temperature regression model (Mohseni *et al.*, 1998) fitted with field data measured in 2009-2010.

The lake bathymetry between the elevations 274.9 m.a.s.l. and 341 m.a.s.l. was derived from

the aggregation of two data sources: 1) raster data measured by the Irstea Hydrobiology Unit using a depth recorder up to an elevation of 329 m.a.s.l. and 2) a 5 m resolution MNT raster obtained from Spot-5 satellite images and provided by the Regional Centre of Geographic Information (CRIGE). Water quality profiles (water temperature, oxygen, and conductivity)

Table 1. List of parameter values for the different ensemble members: uncalibrated, manually calibrated and GLUE-calibrated. The parameters used in calibration and values that differed from the uncalibrated simulation are highlighted in grey. *Lista de valores de los parámetros para los diferentes miembros del conjunto: sin calibrar, calibración manual y calibración mediante el método GLUE. Se resaltan en gris los parámetros usados en la calibración del modelo y los valores que difieren de los valores sin calibrar.*

Parameters (units)	Symbols	Uncalibrated values	Manually calibrated values	GLUE calibrated values
Configuration parameters				
Atmospheric Stability Switch		TRUE	TRUE	TRUE
Maximum layer thickness (m)	Δz_{max}	3	3	3
Minimum layer thickness (m)	Δz_{min}	0.25	0.25	0.25
Output time (h:min)		15:00	15:00	15:00
Time step (s)	Δt	10800	10800	10800
Physical constants				
Bulk aerodynamic transport coefficient	$C_D(= C_S = C_L)$	$1.3 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$
Critical Area (m^2)	A_C	10^7	$9 \cdot 10^4$	10^7
Critical wind speed (m/s)	U_{crit}	3	3	3
Mean albedo (–)	α	0.08	0.08	0.08
Potential Energy Mixing Efficiency (–)	η_p	0.2	0.2	0.2
Shear mixing efficiency (–)	η_k	0.06	0.06	0.06
Vertical Mixing Coefficient (–)	C	200	200	200
Water emissivity (–)	ε_w	0.97	0.97	0.97
Wind Stirring Efficiency (–)	η_s	0.4	0.3	0.4
Morphometry parameters				
Depth of infiltration (m)	z_{inf}	275	285	275
Drag coefficient 1 (–)	$C_{D,in1}$	0.015	0.015	0.015
Drag coefficient 2 (–)	$C_{D,in2}$	0.015	0.015	0.015
Drag coefficient 3 (–)	$C_{D,in3}$	0.015	0.015	0.015
Slope 1 (–)	φ_1	0.03	0.03	0.03
Slope 2 (–)	φ_2	0.5	0.5	0.5
Slope 3 (–)	φ_3	0.0225	0.0225	0.0225
Stream half-angle 1 (°)	α_{in1}	45	45	45
Stream half-angle 2 (°)	α_{in2}	85	85	85
Stream half-angle 3 (°)	α_{in3}	85	85	85
Input data correction coefficients				
Air temperature correction (°)	C_{Ta}	0	1.5	1.89
Artificial inflow temperature correction (°)	C_{Tai}	0	–1.5	0
Natural inflow temperature correction (°)	C_{Tni}	0	0	1.93
LEC coefficient (–)	C_{LEC}	1.7	1.7	1.7
Shading coefficient (–)	C_{HS}	1	0.8	0.60
Wind speed correction coefficient 1 (m/s)	$C_{1,W}$	0	1	–0.74
Wind speed correction coefficient 2 (–)	$C_{2,W}$	1	1.5	3.17

and Secchi depth (*SD*) were periodically measured in the Bimont reservoir since June 2009. Measurements were made twice a month during the period from March 2010 to February 2011, and once a month the rest of the time. Profiles were usually measured in the afternoon between 2 p.m. and 4 p.m. A homogeneous initial thermal profile was assumed. The initial water temperatures were 8 °C for the calibration period and 9.5 °C for the validation period.

Experimental setup

A single-model ensemble was created by grouping the simulations made with three different sets of parameter values: the default set of parameter values, a set of parameter values obtained by manual calibration, and the set of parameter values of the best performing acceptable simulation obtained through the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven & Freer, 2001). DYRESM has different types of parameters: configuration parameters, physical constants and morphometry data. We also used some calibration coefficients to calibrate the input data. In this section, we first present the different types of parameters. Then, we present the different members of the ensemble. Finally, we describe the performance statistics used to assess the simulations.

Model parameters

The list of model parameters and values used for each simulation are presented in Table 1.

Model configuration

The study period ran from January 2009 to August 2014. Forty-two measured water temperature profiles from June 2009 to December 2011 were used for sensitivity analysis and calibration. Thirty-one water temperature profiles from January 2012 to August 2014 were used for model validation. In all of the simulations, minimum layer thickness was set to 0.25 m, maximum layer thickness was set at 3 m (Perroud *et al.*, 2009) and the calculation time step was set to 3 hours. Simulation results were output daily at 3 p.m. in correspondence with profile measurement time. The non-neutral atmospheric stability correction was switched off.

Physical constants

These were the constants used in the calculation of the physical processes. These constants were albedo, water emissivity, bulk aerodynamic transport coefficient, critical wind speed, shear mixing efficiency, potential energy mixing efficiency, wind stirring efficiency, critical area and vertical mixing coefficient. The meaning of these parameters was explained in Antenucci and Imerito (2000), and the mixing coefficients were studied in Yeates and Imberger (2003). The mean annual albedo was assumed to be $\alpha = 0.08$, and the water emissivity was assumed to be 0.97. The albedo was not calibrated since the model was not very sensitive to it and it would interact with the solar radiation correction coefficient (see below). Water emissivity was not calibrated either since this physical parameter can be considered constant (Henderson-Sellers, 1986).

Morphometry parameters

Morphometry parameters refer to the morphometric characteristics of the tributaries (slope, stream half-angle, and drag coefficient) and outlets (depth of extraction). Stream half-angle, slope of tributaries and outlet depths were obtained from charts. Infiltration was considered as an additional outlet, with extraction taking place at the bottom of the reservoir following Dutordoir (2010). Only the parameters for which more uncertainty existed were considered in the sensitivity analysis and calibration process.

Input data correction coefficients

The light extinction coefficient was estimated from the *SD* as follows:

$$LEC = \frac{C_{LEC}}{SD} \tag{1}$$

where the coefficient C_{LEC} is usually 1.7. A constant SD equal to the mean of the study period (4.99 m) was used, and the coefficient C_{LEC} was considered a calibration parameter that can vary in the range of 0.61-1.9 (Margalef, 1983). Calibration coefficients were also applied to the meteorological forcing to account for differences in microclimatic conditions and shading. An additive coefficient was also applied to the temperature of the inflows to analyse the sensitivity of the model to inflow temperatures.

$$T_{a,\,lake} = T_a + C_{Ta} \tag{2}$$

$$HS_{lake} = C_{HS} * HS \quad C_{HS} \in [0, 1] \tag{3}$$

$$W_{lake} = C_{1,W} + C_{2,W}\, W \tag{4}$$

$$T_{wx,\,lake} = T_{wx} + C_{Twx} \tag{5}$$

The variables T_a, HS, W and T_{wx} are measured air temperature, solar radiation, wind speed and water temperature of inflow x, respectively. $T_{a,lake}$, HS_{lake}, W_{lake} and $T_{wx,\,lake}$ are air temperature, solar radiation, wind speed and inflow temperature applied at the lake, respectively. C_{Ta}, C_{HS}, $C_{1,W}$, $C_{2,W}$ and C_{Twx} are correction coefficients.

Ensemble members

UNCALIBRATED SIMULATION

For the first member of the ensemble, the uncalibrated model simulation, the suggested default values of the physical parameters were used, as well as uncalibrated morphometry parameters. The value of the forcing coefficients were such that they did not affect the modified variable (additive coefficients were set to zero, and multiplicative coefficients were set to one). The rationale for using this set of parameter values was that, according to Antenucci & Imerito (2000), no calibration should be necessary for DYRESM. Additionally, the LakeMIP project assessed the performance of 1D hydrodynamic models based on their uncalibrated versions (Stepanenko *et al.*, 2010).

MANUALLY CALIBRATED SIMULATION

For the second member of the ensemble, we obtained a set of manually calibrated parameter values by trying to minimize average RMSE aided by the results of a local sensitivity analysis. The sensitivity analysis was used to identify influential parameters by varying the value of individual parameters in a plausible range while keeping the value of the other parameters constant. According to the local sensitivity analysis, the model was sensitive to most input correction coefficients except C_{LEC} and C_{Twn}, to the infiltration depth, to the wind stirring efficiency and to the critical area.

GLUE-CALIBRATED SIMULATION

For the third member of the ensemble, we used the Generalized Likelihood Uncertainty Estimation (GLUE) methodology to obtain a set of behavioural parameter value sets (Beven & Binley, 1992). For the sensitivity analysis stage of the GLUE method, we ran 2000 Monte Carlo simulations. The initial prior parameter distribution was the uniform distribution. We used the Nash-Sutcliffe efficiency of the simulated volume-averaged temperature as the likelihood measure L (Beven & Binley, 1992):

$$\mathcal{L} = 1 - \frac{\sum_{i=1}^{m} \left(T_{po,i} - T_{ps,i}\right)^2}{\sum_{i=1}^{m} \left(T_{po,i} - \bar{T}_{po}\right)^2} \tag{6}$$

where $T_{po,i}$ is the volume-averaged observed temperature, $T_{ps,i}$ is the volume-averaged simulated temperature, and $\bar{T}_{po}$ is the temporal average of $T_{po,i}$. Acceptable simulations were defined as those such that $\mathcal{L} > 0.90$, while unacceptable simulations where those with $\mathcal{L} \leq 0.90$ or those that stopped before the end of the simulation period. The sensitive parameters were identified by comparing the posterior distributions of acceptable and non-acceptable simulations for each parameter. The model was deemed sensitive to

a parameter if these distributions were different according to the Kolmogorov-Smirnov (K-S) test (Hornberger & Spear, 1980). A Bonferroni bound was applied to the significance value α of the test to avoid multiple testing issues (Efron, 2010), so that $\alpha = 0.05/\mathrm{NP}$, where NP is the number of parameters.

According to the results of the GLUE analysis, the model was sensitive to C_{Ta}, C_{HS}, $C_{2,W}$, C_{Tni}, and C_D, which we used for calibration. We also kept $C_{1,W}$ for calibration because of its relationship to $C_{2,w}$. For the calibration stage, we ran 20 000 Monte Carlo simulations in ten iterations of 2000 simulations each. We used the posterior distributions of the six parameters to be calibrated obtained in the sensitivity analysis as the initial prior distributions. For the other parameters, the initial values were used. After each iteration of 2000 simulations, the prior distributions were updated by using the posterior distributions of the previous iteration to accelerate convergence (Lepage, 1978). The values of the parameter set that provided the highest likelihood in the last iteration were selected as the calibration values for the third member of the ensemble.

Performance evaluation

The performance of the average of the single-model ensemble in predicting the temporal behaviour of surface temperature, hypolimnion temperature and volume-averaged temperature was compared to that of individual members of the ensemble. We defined the surface temperature as the temperature of the first simulated layer, and the hypolimnion temperature was defined as the temperature at the depth of the outlet (at 288 m.a.s.l.). The volume-averaged temperature T_p at the i-th time step was calculated as follows:

$$T_{p,i} = \frac{\sum_j V_j T_{i,j}}{\sum_j V_j} \tag{7}$$

where T is the water temperature, V is the volume and j indicates the layer number.

It was expected that the performance of the uncalibrated model would be worse than that of the two calibrated members. However, experience in the application of ensemble modelling showed that the inclusion of poor models in the ensemble could add useful information (Weigel *et al.*, 2008). To test whether this was the case, the performance of the full ensemble was compared to the performance of the three possible ensembles formed by two members: uncalibrated + manually calibrated members, uncalibrated + GLUE calibrated members, and manually calibrated and GLUE calibrated members.

To test whether model averaging induces a smoothing of the temperature time series, we calculated the lag-one autocorrelation (LOAC) (Alciaturi *et al.*, 2005). The higher the smoothness of the time series, the higher the LOAC would be.

RESULTS

Individual simulations

The simulations issued by the three members of the ensemble showed different error patterns. The simulation with the parameters by default showed a clear vertical and temporal error pattern. It tended to overestimate the near-surface temperatures in the stratification period. It also tended to overestimate the water temperature in the mixing period and underestimate it in the stratification period below the first 5-10 metres. The manually calibrated model tended to overestimate the water temperature in the metalimnion. A seasonal, less clear, error pattern was also present in the second and third ensemble member simulations. All of the ensemble members had difficulties in predicting the temperature below the outlet depth. This could be related to three possible causes: uncertainty in the natural catchment inflow, uncertainty in the volume and depth of infiltration, and the outflow algorithm used by the model. However, the incidence of this error on the general simulation results was reduced because of the small volume held below the outlet depth ($3.4 \cdot 10^5$ m^3, approximately 2% of the volume).

The modelling performance of a given ensemble member varied with the selected period and variable. The most important divergences between the different model simulations occurred during the stratification period, which reflected the difficulty of appropriately simulating the vertical thermal structure. It should be noted that some uncertainties existed regarding the existence of infiltrations, the flow of the natural catchment and the wind speed and direction. Three- and two-dimensional effects were discarded because the field measurements did not reveal important spatial temperature variations.

For the surface temperature (see Fig. S1, available at www.limnetica.com, Table 2), the automatically calibrated model outperformed the manually calibrated model, especially in the validation period. For the hypolimnion temperature (see Fig. S2, available at www.limnetica.com, Table 2), in general, the GLUE calibrated model performed better than the other ensemble members. For the volume−averaged temperature (see Fig. S3, available at www.limnetica.com, Table 2), the manually calibrated model performed better in the calibration period, but the ensemble members performed comparably well in the validation period. The uncalibrated model predicted higher temperatures than the other model versions but matched better measured surface temperatures in the summer of 2009.

Ensemble simulations

When comparing the performance of the ensemble average to that of individual ensemble members, in general, the ensemble average performed best or second best, depending on the variable and performance indicator considered. The ensemble mean was the best performer in the estimation of the surface temperature in both periods and for almost all of the performance indicators, with the exception of MAE and RMSE in the calibration period. In the estimation of the hypolimnion temperature, the ensemble av-

Table 2. Model performance statistics: mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (r). The best performing model for each performance indicator is indicated with a grey background. *Estadísticos de comportamiento de los modelos: error medio (ME), error absoluto medio (MAE), raíz cuadrada del error cuadrático medio (RMSE) y coeficiente de correlación (r).*

Variable	Perf. Statistic	Calibration period				Validation period			
		Default cal.	Manual cal.	GLUE cal.	Ens. mean	Default cal.	Manual cal.	GLUE cal.	Ens. mean
Surface temp.	ME	2.47	−0.81	−0.50	0.39	2.52	−1.33	−1.15	0.02
	MAE	6.71	4.94	3.61	3.82	5.07	4.99	3.02	2.09
	RMSE	2.98	1.75	1.28	1.29	2.88	1.97	1.48	0.88
	r	0.9841	0.9720	0.9805	0.9842	0.9879	0.9727	0.9876	0.9907
Hypolimnion temp.	ME	−0.15	0.54	−0.31	0.03	−1.17	−0.03	−0.88	−0.69
	MAE	4.26	5.91	2.92	3.53	5.14	3.03	3.45	3.03
	RMSE	1.69	1.80	1.03	1.04	2.19	0.98	1.32	1.14
	r	0.9549	0.9482	0.9801	0.9804	0.9125	0.9788	0.9757	0.9793
Volume−averaged temp.	ME	0.93	−0.04	−0.43	0.15	0.14	−0.47	−1.01	−0.44
	MAE	3.14	1.76	3.47	1.91	2.34	1.92	2.10	1.82
	RMSE	1.58	0.97	1.06	0.94	1.15	1.10	1.23	0.96
	r	0.9849	0.9950	0.9810	0.9897	0.9771	0.9910	0.9886	0.9876

erage showed the best correlation, while its performance was best or slightly poorer than best according to the other parameters. In the estimation of the volume-averaged temperature, the ensemble average was most often the second best performer.

The comparison of the performance of the full ensemble with that of the two-member ensembles (Table 3) showed that the full ensemble usually outperformed the other ones. However, this was not always the case for all indicators and for all variables considered. In fact, for the surface temperature, better correlation values were obtained for the two-member ensemble formed by the uncalibrated and the GLUE-calibrated models. For the estimation of the hypolimnion temperature, the full ensemble performed best in the calibration period, but the ensemble formed by the two model calibrations performed best in the validation period.

Uncertainty of simulation results

In this study, we used the ensemble spread as a qualitative estimation of uncertainty. The ensemble spread was wider during the stratification period, both for the surface and hypolimnion temperatures (see Figs. S1-S2, available at www.limnetica.com). For the surface temperature, there was a spread of more than 3-4 °C during the entire stratification period. For the hypolimnion temperature, the spread was narrower, and important differences among the models were limited to the hottest months of the year (May to September). Uncertainty was less important for the volume-averaged temperature (Figure S3), where the ensemble spread amounted to just a few degrees in general (~ 1.5 °C on average), with small differences among seasons. In the summer of 2009, the uncertainty of the volume-averaged temperature was somewhat higher, with a spread larger than 4 °C. These results seemed

Table 3. Performance statistics of the full ensemble and the three two-member ensembles: mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (r). The best performing model for each performance indicator is indicated with a grey background. *Estadísticos de comportamiento del conjunto completo y de los tres subconjuntos formados por dos miembros: error medio (ME), error absoluto medio (MAE), raíz cuadrada del error cuadrático medio (RMSE) y coeficiente de correlación (r). El mejor modelo para cada indicador se señala con un fondo gris.*

Variable	Perf. Statistic	Calibration period				Validation period			
		Full Ens.	Default + Manual	Default + GLUE	Manual + GLUE	Full Ens.	Default + Manual	Default + GLUE	Manual + GLUE
Surface temp.	ME	0.39	0.83	0.98	−0.65	0.02	0.60	0.69	−1.24
	MAE	3.82	4.58	4.51	3.89	2.09	3.07	2.34	3.58
	RMSE	1.29	1.61	1.55	1.45	0.88	1.18	1.09	1.66
	r	0.9842	0.9843	0.9859	0.9780	0.9907	0.9910	0.9921	0.9823
Hypolimnion temp.	ME	0.03	0.19	−0.23	0.12	−0.69	−0.60	−1.02	−0.45
	MAE	3.53	3.84	3.59	3.17	3.03	3.48	3.64	1.98
	RMSE	1.04	1.18	1.27	1.10	1.14	1.15	1.69	0.77
	r	0.9804	0.9777	0.9713	0.9766	0.9793	0.9752	0.9535	0.9901
Volume−averaged temp.	ME	0.15	0.45	0.25	−0.24	−0.44	−0.16	−0.43	−0.74
	MAE	1.91	2.31	2.37	1.95	1.82	2.06	1.94	1.90
	RMSE	0.94	1.18	1.03	0.83	0.96	1.02	0.98	1.05
	r	0.9897	0.9912	0.9850	0.9910	0.9876	0.9863	0.9840	0.9909

to indicate that the overall heat budget was well calculated, given the low uncertainty of the volume-averaged temperature. The simulation of the thermal stratification seemed to be more difficult, given the low concordance among ensemble members.

Smoothing caused by averaging

In the vertical dimension, model averaging resulted in a smoothing of the simulated thermal profile. For example, on July 12, 2013 (see Fig. S4, available at www.limnetica.com), the simulated column water temperatures showed a rather sharp thermocline somewhere between the surface and 10 m. However, the different models disagreed about the position of the thermocline. This explained the important errors observed some metres below the surface, since the measured thermocline was more gradual. However, the ensemble average showed a more gradual temperature profile by compensating the errors among the different estimations. This was equivalent to a change in the physical properties of the simulated system, e.g., to an increased diffusivity in the water column. Regarding the properties of the time series, the LOAC values for the ensemble simulation were higher than the LOAC values for the individual simulations or equal to the maximum of the individual LOAC values. The smoothing was most important for the surface temperatures.

DISCUSSION

The ensemble member that gave a better match to observations varied for different periods. For example, in the summer of 2009, the uncalibrated model simulations of the surface temperature were nearer to the surface temperature observations than the other models. In addition, the hypolimnion temperature was predicted better by the GLUE-calibrated model in the calibration period but was calibrated better by the manually calibrated model in the validation period. This may indicate that the hypothesis of constant model coefficients was not optimal. A temporal dependence of parameter sensitivity in ecological models was previously shown (Song *et al.*, 2013). Due to meteorological variability and changes in management, different processes may be more important in different periods. In the case of the summer of 2009, the hydrology of the reservoir was significantly different since the reservoir was partially emptied for maintenance and there was almost no inflow from June 26 to July 22. There were also differences in the reservoir management between the calibration and validation periods. In the calibration period, inflow was more irregular, and the artificial inflow was often diverted through the bypass gate at 331 m.a.s.l. instead of passing through the submerged inlet at 315 m.a.s.l. Additionally, as noted by one of the reviewers, "[t]here is a seasonal dependence of processes which is not reflected in the parameterizations of the driving factors, e.g., changing wind direction and thus seasonal fetch realizations." In addition, when using the RMSE as an objective function, as in the manual calibration, a few significant events may have a greater influence in determining the calibration values (Berthet *et al.*, 2010).

As a result, even if past performance information of different models is available, it can be difficult to predict, which will be the best and poorest models in another time period (Hagedorn *et al.*, 2005). It is not assured that the best models in the calibration period will continue to be the same in the other periods (Van Straten & Keesman, 1991; Andréassian *et al.*, 2012). Miscalibration and overcalibration can cause a reduced performance outside of the calibration period (Andréassian *et al.*, 2012). Miscalibration occurs when the calibration algorithm identifies a secondary optimum, while some causes for overcalibration are measurement error and noise, the limited volume of available field data and uncertain model structure (Beck, 1981; Andréassian *et al.*, 2012). The use of calibration methods that explore the parameter space (Monte Carlo methods, Monte Carlo Markov chains, genetic algorithms, etc.) can address miscalibration, but not necessarily overcalibration. Using an ensemble that includes models calibrated with different methods and/or objective functions is a way to guard against both issues.

In general, the performance of the ensemble average was better than that of the ensemble members or just slightly worse, in accordance with the results of other authors (Hagedorn *et al.*, 2005). Trolle *et al.* (2014) found that a multi-model ensemble provided more accurate predictions of phytoplankton dynamics than its individual members. In our case, more often than not, the two- and three-member ensembles were superior to the individual simulations. For example, the surface temperatures of the simulation results of the manually calibrated model were improved by simply averaging them with those of the uncalibrated model. According to Hagedorn *et al.* (2005), a poor model cannot add value to an ensemble if it is consistently worse than average in all considered aspects. This was not the case of the uncalibrated model, which had a performance comparable to that of the other ensemble members for some indicators and periods.

Using model averaging is especially appropriate when there are important uncertainties regarding the modelling results (Armstrong, 2001). Ensemble simulations provide a simple way to estimate the uncertainty of the simulation results at a low computational cost. However, only qualitative assessments of uncertainty are possible with ensemble modelling (Gal *et al.*, 2014), given the difficulty to assign a probability to each member (Stephenson *et al.*, 2012). In contrast, other computationally intensive methods, such as Monte Carlo simulations or GLUE when applied in its full breadth (McIntyre *et al.*, 2002), allow making quantitative estimations of the uncertainty of predictions. Ensemble models seem to be at a disadvantage against quantitative uncertainty methods, but it should be kept in mind that in both cases we are assuming that the model structure is correct. Notwithstanding the apparent exactness of quantitative uncertainty estimations, they are nonetheless inexact since structural uncertainty is not taken into account and because of the limitations of calibration data on which their results depend. For example, the 5th and 95th percentiles of 900 Monte Carlo simulations made with the model PCLake by Nielsen *et al.* (2014) enclosed between 70% and 90% of the observed variability. When the user is very confident in the correctness and completeness of the model, a quantitative uncertainty estimation method seems justified. Otherwise, a more economic method that offers qualitative uncertainty estimations might be preferable. While providing a more reduced exploration of the parameter space, qualitative uncertainty estimates underline the uncertain nature of the simulation results and might be less prone to misguiding the final users.

The need to take into account structural uncertainty is especially important in aquatic modelling, where there is important model diversity. Janssen *et al.* (2015) listed 42 aquatic ecosystem models, 10 of which included a hydrodynamic module. To explore structural uncertainty, as well as parametric uncertainty, a multi-model ensemble may then be used. In that case, since models share some structural similarity, their results are not independent from each other, and weights should be added to the different models (Knutti *et al.*, 2010). However, even models with very similar structures can give quite different answers with the same input data, e.g., GLM and DYRESM in Gal *et al.* (2016).

Finally, averaged model simulations not necessarily share the same physical and dynamic characteristics as the system of study, as shown by our results and others (Du *et al.*, 1997; Knutti *et al.*, 2010). The characteristics of the time series could be different from those of the original models, as exemplified by the increase in the LOAC, and the magnitude of peaks of the simulated variables could be reduced. A reason for the smoothing of the averaged simulations may be the cancellation of aleatory errors from the different models by averaging them. Another reason may be the loss of signal when temperature variations in the different models are of different signs at a certain point in time or they are not exactly synchronous (Knutti *et al.*, 2010). While the cancellation of aleatory errors may be a desirable property for the derived temperature series, the loss of signal could affect the estimation of extreme temperatures.

The dynamic properties of the system may also be modified. Where a bifurcation between different states exists in the original system, the

model average may predict an implausible average state (Knutti *et al.*, 2010). One such example in lakes is the transition between a stratified water column and a mixed water column. Wind induced mixing events can be quite sharp, but they can appear to be more gradual in the averaged simulation if the individual models react differently to the wind forcing.

CONCLUSIONS

Ecological models are useful representations of ecosystems that can be used to investigate which processes take place and as management tools. However, they are just approximations of reality and have limitations (Boschetti *et al.*, 2011). Improving model results through more sophisticated calibration methods can take the modeller just that far. Limitations in data availability and in the identification of the model structure can cause overcalibration so that model performance degrades outside the calibration period. Using an ensemble formed by different model versions calibrated with different methods and function objectives can help us obtain consistently better simulations, as well as obtain a qualitative estimation of uncertainty. Such an approach is especially useful when there is high uncertainty regarding the model and/or input data. While the ensemble average can provide more accurate estimations, it should be noted that it can result in a loss of signal or may lack a physical meaning. Caution is then necessary in the interpretation of the results, especially regarding extreme values, such as maximum temperatures, and singular events, such as wind-induced overturns. The method can be applied to other types of lakes and for more complex lake models.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers for their comments to improve the manuscript. This paper also benefitted from discussions with Alexey A. Voinov and Björn Reineking. The authors thank Météo-France for providing the meteorological data, and the Société du Canal de Provence for providing hydrological and water quality data. The authors thank Nathalie Reynaud and Yann Le Coarer for their help in preparing the reservoir bathymetry, the members of the Onema-Irstea consortium Martin Daufresne, Tiphaine Péroux, Julien Dublon and Delphine Rebière for the acquisition of water temperature and Secchi depth data, and Marie-José Salençon and Magali Gant from EdF R&D for their comments. The funding was provided by the research programme of the joint Irstea-EdF research group HYNES.

REFERENCES

ALCIATURI, C.E., M.E. ESCOBAR & I. ESTÉVES. 2005. The use of the autocorrelation function in modeling of multivariate data. *Analytica Chimica Acta*, 553 (1–2): 134–140.

ANDRÉASSIAN, V., N. LE MOINE, C. PERRIN, M.-H. RAMOS, L. OUDIN, T. MATHEVET, J. LERAT & L. BERTHET. 2012. All that glitters is not gold: the case of calibrating hydrological models. *Hydrological Processes*, 26(14): 2206–2210.

ANTENUCCI, J. & A. IMERITO. 2000. *The CWR Dynamic Reservoir Simulation Model DYRESM. Science Manual*. The University of Western Australia. Perth, Australia.

ARMSTRONG, J. S. 2001. Combining forecasts. In: *Principles of Forecasting: A Handbook for Researchers and Practitioners*. J. S. Armstrong (ed.): 417–439. Kluwer Academic Publishers, Norwell, MA.

BECK, M.B. 1981. Hard or soft environmental systems? *Ecological Modelling*, 11(4): 233–251.

BERTHET, L., V. ANDRÉASSIAN, C. PERRIN & C. LOUMAGNE. 2010. How significant are quadratic criteria? Part 2. On the relative contribution of large flood events to the value of a quadratic criterion. *Hydrological Sciences Journal*, 55 (6): 1063–1073.

BEVEN, K. & A. BINLEY. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3): 279–298.

BEVEN, K. & J. FREER. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental sys-

tems using the GLUE methodology. *Journal of Hydrology*, 249 (1-4): 11–29.

BOSCHETTI, F., N.J. GRIGG & I. ENTING. 2011. Modelling = conditional prediction. *Ecological Complexity*, 8(1): 86–91.

CLEMEN, R.T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4): 559–583.

DU, J., S.L. MULLEN & F. SANDERS. 1997. Short-Range Ensemble Forecasting of Quantitative Precipitation. *Monthly Weather Review*, 125(10): 2427–2459.

DUTORDOIR, S. 2010. *Modélisation thermodynamique de la retenue Bimont à l'aide du modèle DYRESM* (Master's Report). Pôle Onema-Irstea, Université Pierre et Marie Curie, Mines ParisTech, AgroParisTech. Paris, France.

EFRON, B. 2010. *Large-Scale Inference. Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press. Cambridge.

FANG, X., S.R. ALAM, H.G. STEFAN, L. JIANG, P.C. JACOBSON & D.L. PEREIRA. 2012. Simulations of water quality and oxythermal cisco habitat in Minnesota lakes under past and future climate scenarios. *Water Quality Research Journal of Canada*, 47 (3-4): 375–388.

GAL, G., J. IMBERGER, T. ZOHARY, J. ANTENUCCI, A. ANIS & T. ROSENBERG. 2003. Simulating the thermal dynamics of Lake Kinneret. *Ecological Modelling*, 162 (1-2): 69–86.

GAL, G., V. MAKLER-PICK & N. SHACHAR. 2014. Dealing with uncertainty in ecosystem model scenarios: Application of the single-model ensemble approach. *Environmental Modelling & Software*, 61: 360–370.

GAL, G., D. SCBLING, Y. GILBOA & N. SHACHAR. 2016. The simulated impact of increased frequency of extreme climatic conditions on a sub-tropical lake ecosystem. 33rd SIL Congress. 31 July-5 August. Turin.

HAGEDORN, R., F.J. DOBLAS-REYES & T.N. PALMER. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting –I. Basic concept. *Tellus A*, 57(3): 219–233.

HAN, B.-P., J. ARMENGOL, J.C. GARCIA, M. COMERMA, M. ROURA, J. DOLZ & M. STRASKRABA. 2000. The thermal structure of Sau Reservoir (NE: Spain): a simulation approach. *Ecological Modelling*, 125 (2-3): 109–122.

HENDERSON-SELLERS, B. 1986. Calculating the surface energy balance for lake and reservoir modeling: A review. *Reviews of Geophysics*, 24(3): 625–649.

HORNBERGER, G.M. & R.C. SPEAR. 1980. Eutrophication in peel inlet –I. The problem-defining behavior and a mathematical model for the phosphorus scenario. *Water Research*, 14(1): 29–42.

JANSSEN, A.B.G., G.B. ARHONDITSIS, A. BEUSEN, K. BOLDING, L. BRUCE, J. BRUGGEMAN, R.M. COUTURE, A.S. DOWNING, J. ALEX ELLIOTT, M.A. FRASSL, G. GAL, D.J. GERLA, M.R. HIPSEY, F. HU, S.C. IVES, J.H. JANSE, E. JEPPESEN, K.D. JÖHNK, D. KNEIS, X. KONG, J.J. KUIPER, M.K. LEHMANN, C. LEMMEN, D. ÖZKUNDAKCI, T. PETZOLDT, K. RINKE, B.J. ROBSON, R. SACHSE, S.A. SCHEP, M. SCHMID, H. SCHOLTEN, S. TEURLINCX, D. TROLLE, T.A. TROOST, A.A. VAN DAM, L.P.A. VAN GERVEN, M. WEIJERMAN, S.A. WELLS & W.M. MOOIJ. 2015. Exploring, exploiting and evolving diversity of aquatic ecosystem models: a community perspective. *Aquatic Ecology*, 49(4): 513–548.

KNUTTI, R., R. FURRER, C. TEBALDI, J. CERMAK & G.A. MEEHL. 2010. Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate*, 23(10): 2739–2758.

LEPAGE, P.G. 1978. A new algorithm for adaptive multidimensional integration. *Journal of Computational Physics*, 27(2): 192–203.

MARGALEF, R. 1983. *Limnología*. Omega. Barcelona.

MCINTYRE, N., H. WHEATER & M. LEES. 2002. Estimation and propagation of parametric uncertainty in environmental models. *Journal of Hydroinformatics*, 4(3): 177–198.

MOHSENI, O., H.G. STEFAN & T.R. ERICKSON. 1998. A nonlinear regression model for weekly stream temperatures. *Water Resources Research*, 34(10): 2685–2692.

MOOIJ, W., D. TROLLE, E. JEPPESEN, G. ARHONDITSIS, P.V. BELOLIPETSKY, D.B. R. CHITAMWEBWA, A.G. DEGERMENDZHY, D.L. DEANGELIS, L.N. DE SENERPONT DOMIS, A.S. DOWNING, J.A. ELLIOTT, C.R. FRAGOSO, JR., U. GAEDKE, S.N. GENOVA, R.D. GULATI, L. HÅKANSON, D.P. HAMILTON, M.R. HIPSEY, J.T. HOEN, S. HÜLSMANN, F.H. LOS, V. MAKLER-PICK, T. PETZOLDT, I.G. PROKOPKIN, K. RINKE,

S.A. SCHEP, K. TOMINAGA, A.A. VAN DAM, E.H. VAN NES, S.A. WELLS & J.H. JANSE. 2010. Challenges and opportunities for integrating lake ecosystem modelling approaches. *Aquatic Ecology*, 44(3): 633–667.

NIELSEN, A., D. TROLLE, R. BJERRING, M. SØNDERGAARD, J.E. OLESEN, J.H. JANSE, W.M. MOOIJ & E. JEPPESEN. 2014. Effects of climate and nutrient load on the water quality of shallow lakes assessed through ensemble runs by PCLake. *Ecological Applications*, 24(8): 1926–1944.

PERROUD, M., S. GOYETTE, A. MARTYNOV, M. BENISTON & O. ANNEVILLE. 2009. Simulation of multiannual thermal profiles in deep Lake Geneva: A comparison of one-dimensional lake models. *Limnology and Oceanography*, 54(5): 1574–1594.

PRATS, J. & P.-A. DANIS. 2015. *Optimisation du réseau national de suivi pérenne in situ de la témperature des plans d'eau : apport de la modélisation et des données satellitaires*. Onema. Aix-en-Provence, France.

READ, J.S., L.A. WINSLOW, G.J.A. HANSEN, J. VAN DEN HOEK, P.C. HANSON, L.C. BRUCE & C.D. MARKFORT. 2014. Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecological Modelling*, 291: 142–150.

RIGOSI, A. & F.J. RUEDA. 2012. Propagation of uncertainty in ecological models of reservoirs: From physical to population dynamic predictions. *Ecological Modelling*, 247(0): 199–209.

SOCIÉTÉ DU CANAL DE PROVENCE. 2013. *Étude de l'optimisation de la cote de RN du barrage de Bimont* (Project Report). Société du Canal de Provence. Le Tholonet, France.

SONG, X., B.A. BRYAN, A.C. ALMEIDA, K.I. PAUL, G. ZHAO & Y. REN. 2013. Time-dependent sensitivity of a process-based ecological model. *Ecological Modelling*, 265: 114–123.

STEPANENKO, V., K.D. JÖHNK, E. MACHULSKAYA, M. PERROUD, Z. SUBIN, A. NORDBO, I. MAMMARELLA & D. MIRONOV. 2014. Simulation of surface energy fluxes and stratification of a small boreal lake by a set of one-dimensional models. *Tellus A*, 66 (21389): 1–18.

STEPANENKO, V.M., S. GOYETTE, A. MARTYNOV, M. PERROUD, X. FANG & D. MIRONOV. 2010. First steps of a Lake Model Intercomparison Project: LakeMIP. *Boreal Environment Research*, 15(2): 191–202.

STEPANENKO, V.M., A. MARTYNOV, K.D. JÖHNK, Z.M. SUBIN, M. PERROUD, X. FANG, F. BEYRICH, D. MIRONOV & S. GOYETTE. 2013. A one-dimensional model intercomparison study of thermal regime of a shallow, turbid midlatitude lake. *Geoscientific Model Development*, 6(4): 1337–1352.

STEPHENSON, D.B., M. COLLINS, J.C. ROUGIER & R.E. CHANDLER. 2012. Statistical problems in the probabilistic prediction of climate change. *Environmetrics*, 23(5): 364–372.

TOLLEFSON, J. 2016. Paris climate deal: what comes next. *Nature News*. Nature Publishing Group.

TROLLE, D., J.A. ELLIOTT, W.M. MOOIJ, J.H. JANSE, K. BOLDING, D.P. HAMILTON & E. JEPPESEN. 2014. Advancing projections of phytoplankton responses to climate change through ensemble modelling. *Environmental Modelling & Software*, 61(0): 371–379.

VAN STRATEN, G.T. & K.J. KEESMAN. 1991. Uncertainty propagation and speculation in projective forecasts of environmental change: A lake-eutrophication example. *Journal of Forecasting*, 10(1-2): 163–190.

WEIGEL, A.P., M.A. LINIGER & C. APPENZELLER. 2008. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134(630): 241–260.

YEATES, P.S. & J. IMBERGER. 2003. Pseudo two-dimensional simulations of internal and boundary fluxes in stratified lakes and reservoirs. *International Journal of River Basin Management*, 1(4): 297–319.