A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network

Louise C. Bruce a, *, Marieke A. Frassl b, c, George B. Arhonditsis d, Gideon Gal e, David P. Hamilton f, Paul C. Hanson g, Amy L. Hetherington h, i, John M. Melack j, Jordan S. Read k, Karsten Rinke b, Anna Rigosi l, Dennis Tolle m, Luke Winslow n, Rita Adrian o, Ana I. Ayala p, Serghei A. Bocaniov b, aa, Bertram Boehner b, Casper Boon a, Justin D. Brookes l, Thomas Bueche q, Brendan D. Busch a, Diego Copetto r, Alicia Cortés s, Elvira de Eyto t, J. Alex Elliott u, Nicole Gallina v, Yael Gilboa w, Nicolas Guyennon x, Lei Huang y, Onur Kerimoglu z, John D. Lenters g, Sally MacIntyre i, ab, Vardit Makler-Pick ac, Chris G. McBride i, Santiago Moreira ad, Deniz Öz Kundakci ae, o, Marco Pilotti af, Francisco J. Rueda b, ag, James A. Rusak ah, Nihar R. Samal ai, Martin Schmid aj, Tom Shatwell ak, Craig Snorthwell al, Frédéric Soulignac az, al, Giulia Valerio af, Leon van der Linden am, Mark Vetter an, Brigitte Vinçon-Leite a, Junbo Wang y, Michael Weber b, Chaturangi Wickramaratne a, R. Iestyn Woolway ao, Huaxia Yao ah, Matthew R. Hipsey a

a Aquatic EcoDynamics Group, UWA School of Agriculture and Environment, The University of Western Australia, Australia
b Department of Lake Research, Helmholtz Centre for Environmental Research (UFZ), Germany
c Limnological Institute, University of Konstanz, Germany
d Department of Physical and Environmental Sciences, University of Toronto, Canada
e Kinneret Limnological Laboratory, Israel Oceanographic & Limnological Research, Israel
f Environmental Research Institute, The University of Waikato, New Zealand
gh Center for Limnology, University of Wisconsin-Madison, USA
i Department of Biological Science, Virginia Tech, USA
j Department of Natural Resources, Cornell University, USA
k Bren School of Environmental Science and Management, University of California, Santa Barbara, USA
l Center for Integrated Data Analytics, U.S. Geological Survey, USA
m Water Research Centre, The Environment Institute, School of Biological Sciences, The University of Adelaide, Australia
n Department of Bioscience, Århus University, Denmark
o Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
p Department of Ecosystem Research, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Germany
q Department of Civil Engineering, University of Granada, Spain
r Department of Geography, Ludwig-Maximilians-University of Munich, Germany
s National Research Council, Water Research Institute (IRSA-CNR), Italy
t Marine Science Institute, University of California Santa Barbara, USA
u Marine Institute, Ireland
v Centre for Ecology and Hydrology, Lancaster, United Kingdom
w Group of Aquatic Physics, Department F-A. Forel for Environmental and Aquatic Sciences, Institute for Environmental Sciences, University of Geneva, Switzerland
x Faculty of Civil and Environmental Engineering, Technion - Israel Institute of Technology, Israel
y National Research Council, Water Research Institute (IRSA-CNR), Rome, Italy
z Key Laboratory of Tibetan Plateau Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), China
a INRA, UMR CARRET, Université de Savoie Mont Blanc, Thonon les Bains, France
b INRIA, UMR CARRET, Université de Savoie Mont Blanc, Thonon les Bains, France
c Key Laboratory of Tibetan Plateau Environmental Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences (CAS), China
d INRA, UMR CARRET, Université de Savoie Mont Blanc, Thonon les Bains, France
e Institute of Coastal Research, Helmholtz-Zentrum Geesthacht (HZG), Geesthacht, Germany
f Department of Ecology, Evolution and Marine Biology, University of California Santa Barbara, USA
g Granim College, Israel
h Laboratoire des Sciences du Climat et de l’Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, France
i Science and Strategy, Waikato Regional Council, New Zealand

* Corresponding author. School of Earth & Environment (M004), The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.
1. Introduction

Vörösmarty et al. (2000) urged the international “water sciences community” to work together in the collation and dissemination of hydrological data and modelling techniques to improve our understanding of freshwater ecosystems and “secure a more complete picture of future water vulnerabilities”. Lakes, in particular, are highly valued ecosystems as they provide important water and food resources, and numerous other ecosystem services (Wilson and Carpenter, 1999). Human activities such as fresh water diversion and increased nutrient loading, in addition to indirect pressures from climate change, have led to an increased vulnerability of lakes on a global scale (Folke et al., 2004). These challenges have given rise to international networks of scientists such as the Global Lake Ecological Observatory Network (GLEON: gleon.org). Collaborative networks can take advantage of shared data, techniques, and expertise to enable scientists to address the ecological challenges facing lakes globally (Eigenbrode et al., 2007; Adams, 2012; Goring et al., 2014). GLEON was initiated in 2005 as a grassroots science community with a vision to observe, understand and predict freshwater systems at a global scale (Weathers et al., 2013).

Collaboration between scientists and synthesis of data collected through international networks has led to advances in our understanding of how lake ecosystems respond to external changes and contribute to effective lake management on a local (Gal et al., 2009), regional (Read et al., 2014; Trolle et al., 2015) and global scale (O’Reilly et al., 2015). Analyses based on data from a broad spectrum of lakes across the globe have provided insight into metabolism and carbon cycling in lakes (Hanson et al., 2013; Solomon et al., 2013), the role of wind and heat exchange in lake physics (Read et al., 2012), the impact of climate change (Adrian et al., 2009), response and recovery of lakes to extreme events (Jennings et al., 2012; Klug et al., 2012), incorporation of high frequency data for model validation (Hamilton et al., 2015) and assisted in development of models (Staehr et al., 2010; Read et al., 2011; Kara et al., 2012; Hipsey et al., 2017). Further interrogation of the emerging multi-lake datasets offers the potential to advance our understanding of how lakes respond to pressures such as climate or land use change from the individual to global scales.

The collaborative network also creates opportunities for developing and testing modelling tools. Aquatic ecosystem models are recognized as essential instruments to improve understanding of processes, analyse relationships, test hypotheses and predict the state of a system (Trolle et al., 2012). These models have evolved since the first attempts in the early 1920s, with a recent review of aquatic ecosystem models revealing the diversity of existing models from simple 0-D to complex 3-D coupled hydrodynamic-biogeochemical models (Janssen et al., 2015). This diversity creates challenges for integration and synthesis of model approaches (Mooij et al., 2010). The Aquatic Ecosystem Modelling Network (AEMON: https://sites.google.com/site/aquaticmodelling/home) originated to foster collaboration and improve model development, predictability, transparency and reliability. One of the major challenges facing modellers is how to develop generic models that can capture the diversity of ecosystems while allowing prediction with confidence of the processes of each system. In order to undertake analytical synthesis across multiple sites, there is a need to assess the transferability of the underlying model and standardise its structure, parameterisation, development and examination. While the need to develop a set of standards for model assessment and reporting is widely recognized (Bennett et al., 2013; Grimm et al., 2014), the ability to test these standards across multiple systems and highlight both strengths and limitations of a particular model remains a challenge.

For lakes and reservoirs in particular, one-dimensional (1-D) models that resolve vertical profiles of temperature and density have found widespread use due to their computational efficiency and minimal calibration requirements. The reduced complexity of 1-D models is advantageous whenever greater computational efficiency is needed, e.g., in ensemble modelling (Trolle et al., 2014), model inter-comparison projects such as LakeMIP (http://www.unige.ch/climate/lakemip) (Stepanenko et al., 2010; Thiery et al., 2014), probabilistic studies (Schlaibing et al., 2014), long-term scenario analysis (Gilboa et al., 2014) or when linking lake models to global climate models (Balsamo et al., 2012) or catchment models (Hipsey et al., 2014).
et al., 2015). Moreover, lake managers and reservoir operators prefer models having a simpler application and often rely on 1-D models for this reason (Kerimoglu and Rinke, 2013; Weber et al., 2017).

Here we introduce the Multi-Lake Comparison Project (MLCP) undertaken within AEMON. The MLCP is a community driven project, where teams of modellers simulate lakes using common approaches for model setup, assessment and analysis. The underlying purpose of the project was to bring together an international network of scientists and modellers with diverse experience in order to improve our ability to predict how lake ecosystems respond to external drivers. In the first stage, the MLCP took advantage of GLEON and AEMON member data from numerous, diverse lakes to stress test the recently developed General Lake Model (GLM) (Hipsey et al., 2017). GLM is a 1-D hydrodynamic model for use in a broad spectrum of enclosed aquatic ecosystems such as lakes, reservoirs and wetlands. The model is simple in nature and is based on assumptions that are common to previous model applications (Imberger and Patterson, 1989; Hamilton and Schladow, 1997; Coats et al., 2006). The model conducts a lake mass and energy balance to compute vertical profiles of temperature, salinity and density while accounting for the effect of inflows and outflows, surface heating and cooling, mixing and ice cover on the lake. GLM can be coupled with biogeochemical models to explore the impact of temperature, stratification, and vertical mixing on the dynamics of lake ecology (e.g. Snortheim et al., 2017).

This paper summarises the first phase of the MLCP to develop and stress-test GLM. The stress-test involved applying a single standardised procedure for model set-up, simulation, performance testing and analysis to 32 lakes from across the global network. The main objective of this study was to undertake comparative analysis of model performance using an unprecedented diversity of lake types in order to advance our understanding of limnology and contemporary modelling practices. The specific aims of the study were to:

1. Ascertain levels of model performance and relate it to model input uncertainty;
2. Identify lake attributes (e.g. depth, inflows, and climate) that correspond with high (or low) prediction accuracy;
3. Relate sensitivity of model output variables to changes in surface exchange, heating and mixing parameters that characterise 1-D lake models;
4. Document the transferability of the model without recalibration of individual parameters among lakes, even where these lakes may strongly differ in their properties; and
5. Provide guidance to lake modellers as to how to focus data collation and model application efforts to improve predictions for lake ecosystems.

To ease readability, this main section of the paper includes all text as well as tables and figures relevant to the major methodology and results from the study. Additional data have been provided in the following four appendices as supplementary material to the main study:

A Describing uncertainty error associated with the model set up;
B Extended results describing model performance;
C Extended results of the sensitivity analysis; and
D A summary of acknowledgements for each lake.

2. Methods

2.1. Study site selection

Lakes were not chosen a priori based on their attributes, but rather AEMON and GLEON members were invited to participate in the MLCP by volunteering details of their candidate lake to the group (shared via open access spreadsheet). The requirement for inclusion of a lake was based on the following three conditions:

1. Sufficient temperature data were available for validation (at least 2 years of monthly/regular thermistor chain and/or profile data);
2. High-resolution meteorological forcing data from an on-lake buoy or local terrestrial based station were available; and
3. Gauged or well-estimated inflows and outflows were available over the simulation period to form a reliable lake water balance.

Participants were also required to have a basic knowledge of lake modelling. Instructions as to how to set-up the GLM test cases, and a common binary executable (GLM v2.2.0) were made available for download from the Aquatic Ecosystem Dynamics (AED) website (https://github.com/AquaticEcoDynamics/GLM). Pre- and post-processing MATLAB scripts were provided to all participants to ensure a common model setup and assessment approach (https://github.com/AquaticEcoDynamics/GLMn), and all GLM lake setups were available to other members via a cloud-based, shared folder.

A total of 32 lakes was chosen for the analysis, with an alphabetic listing of the lakes and their physical characteristics in Table 1. Each lake is associated with a two letter abbreviated code, and for brevity when presenting model results, the lakes are frequently referred to by this code. To illustrate the range of sizes in the lakes included in this study, lake outlines have been drawn to scale in Fig. 1. With the exception of lakes Geneva and Kinneret, all lake simulations were run for two years, with the start year and date indicated in Table A3. For Lake Geneva and Lake Kinneret, analyses were performed separately for two alternative 2-year time periods with significant differences in climate and inflows. For Lake Geneva, 2003 to 2004 had higher than average summer air temperatures, precipitation and inflows as well as an uncharacteristically high winter inflow in early 2004. In contrast, 2001 to 2002 experienced closer to the “normal” seasonal cycles of climate and inflows (Anneville et al., 2010). These simulations are referred to as Geneva03 and Geneva01 respectively. For Lake Kinneret, 1997 to 1998 had generally average climatic conditions (Bruce et al., 2006). In contrast, 2003 to 2004 had a rainy winter (Feb-Mar 2003, Jan-Feb 2004), large changes to lake level and lower than normal water temperatures (Berger and Telzch, 2005). These simulations are referred to as Kinneret97 and Kinneret03, respectively.

Lake depths ranged from 2.4 to 440 m, and lake surface areas from 104,000 m² to 579,000,000 m² (Table 1). A comparative plot of the hypsographic curves for each of the 32 lakes shows diversity in lake size and bed slope (Fig. A1). Annual average inflows ranged from 0 to 3.3 × 10⁷ m³ d⁻¹ and residence times from 1 month to 67 years (Table A3). Lake elevation ranged from 209 m below to 4718 m above sea level (Table A1). Annual average air temperature ranged from below freezing (—9.1 °C) to 22.4 °C (Table A3). While the majority of the lakes in the MLCP are mid-latitude (both northern and southern hemisphere), two lakes are located in the Arctic (Emaiksoun and Toolik).

2.2. GLM set-up

GLM has several configuration options for simulating surface heating, mixing and inflow and outflow (Hipsey et al., 2017). For this assessment, model set-ups were configured based on the site-specific conditions (e.g. hypsographic curve and number of inflows and outflows), but all simulations adopted the same model algorithms and parameters for mixing, surface heat fluxes, and ice cover. Default parameters adopted are summarised in Table 2.
All simulations were run for 2 years or 730 days starting with initial conditions in the winter or when the lake was most nearly well mixed. For the northern hemisphere lakes the start date was the 1st of January and for lakes located in the southern hemisphere the start date was set at 1st July. The initial conditions were taken from the closest field profile measurements to the start date. The standardised start date was chosen to simplify cross lake comparisons. For the majority of the lakes in the MLCP, mid-winter was also associated with complete mixing thus reducing error associated with complete mixing.

In an attempt to assess the errors associated with input data limitations, a qualitative weighting system was used to assess each input variable or constant, where a minimum score is associated with best available input or observation data (Table A1).

Box plots are used to present monthly means and range of input data, lakes where input data had to be estimated were still selected based on the best available input or observation data (Table A2a). This set of measures of model performance was used to evaluate model performance with the best available input or observation data (Table A1).

Table A2a lists the method of determining the hypsographic curve, distance from lake and frequency of meteorological data and observed data and method of determining inflow and outflow and extinction coefficient for each lake in the MLCP. This information is used to determine the relative error scale associated with boundary forcing and observed data for each lake (Table A2b), where low refers to low uncertainty in forcing data and high indicates a higher level of error associated with model input. Error associated with the determination of long wave radiation was not included in the error scaling method.

Measures of model fit used to evaluate model performance included five alternatives listed below. This set of measures of model fit enabled us to standardise comparisons among lakes, track estimated for three lakes (Alexandrina, Muggelsee and Woods).

All meteorological data for short wave radiation, air temperature, relative humidity, wind speed and precipitation were supplied either from an on lake station or the closest meteorological station to the lake. Long wave radiation was either measured directly (net or incident) or calculated by GLM using cloud cover data.

In an attempt to assess the errors associated with input data limitations, a qualitative weighting system was used to assess each input variable or constant, where a minimum score is associated with the best available input or observation data (Table A1).
trends in deviations from observed data (Bennett et al., 2013) and to compare with similar lake modelling studies previously published (e.g. Rigosi et al., 2010).

Measures of model fit were calculated as:

1) Root mean square error ($RMSE$):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i - O_i)^2}{N}}$$  \hspace{1cm} (2-1)

2) Model Efficiency ($MEFF$; Murphy, 1988; Nash and Sutcliffe, 1970):

$$MEFF = 1 - \frac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2}$$  \hspace{1cm} (2-2)

3) Correlation coefficient ($r$):

$$r = \frac{\sum_{i=1}^{N}(P_i - \bar{P})(O_i - \bar{O})}{\left[ \sum_{i=1}^{N}(P_i - \bar{P})^2 \sum_{i=1}^{N}(O_i - \bar{O})^2 \right]^{1/2}}$$  \hspace{1cm} (2-3)
Table 2
Description, symbols and initial values of the parameters used in the sensitivity analysis.

| Symbol | Description                                      | Reference                  | Initial value |
|--------|--------------------------------------------------|----------------------------|---------------|
| $C_h$  | Bulk aerodynamic coefficient for sensible heat transfer | (Fischer et al., 1979)   | 0.0013        |
| $C_e$  | Bulk aerodynamic coefficient for latent heat transfer | (Fischer et al., 1979)   | 0.0013        |
| $C_d$  | Bulk aerodynamic momentum transfer coefficient   | (Fischer et al., 1979)   | 0.0013        |
| $C_C$  | Mixing efficiency - convective overturn          | (Yeates and Imberger, 2003)| 0.2           |
| $C_w$  | Mixing efficiency - wind stirring                | (Spigel et al., 1986)     | 0.23          |
| $C_T$  | Mixing efficiency - unsteady turbulence (acceleration) | (Sherman et al., 1978)   | 0.3           |
| $C_s$  | Mixing efficiency - shear production             | (Sherman et al., 1978)   | 0.51          |
| $C_{KH}$ | Mixing efficiency - Kelvin-Helmholtz turbulent billows | (Sherman et al., 1978)   | 0.3           |
| $C_{hyp}$ | Mixing efficiency of hypolimnetic turbulence   | (Weinstock, 1981)        | 0.5           |

Fig. 2. Time series of monthly mean values across all lakes for (a) short wave radiation, (b) relative humidity, (c) net longwave radiation, (d) wind speed, (e) air temperature and (f) precipitation. For each box, horizontal lines represent median, 25th and 75th percentile, whiskers <1.5 times the interquartile, and outliers (+) values > 1.5 times the interquartile range. Note that lakes from the Southern Hemisphere start with a shift of 6 months relative to the Northern Hemisphere lakes.
for each lake (the river slope is indicated as other out).

absolute so that under and over estimations do not cancel each
to the mean, enabling like comparisons between variables and is
predictions (Bennett et al., 2013). Finally,
ated from observed values and can be used to determine the bias in
Brett, 2004).
The correlation coef
provides equal predictive skill as the mean of the observed data.
that one indicates perfect
fi
t and zero indicates that the model

initial manual calibration focused on re

and model predicted data and

measured data.
A further advantage of calculating alternative measures of
model fit is that different methods of model evaluation highlight
different aspects of model performance (Bennett et al., 2013). RMSE
is a standard measure of the average deviation of simulated values
from observations with values near zero indicating a close match
and units that correspond to those of the variable. MEFF is the
square of the deviation of simulated values from observations,
normalised to the standard deviation of the observed data, such
that one indicates perfect fit and zero indicates that the model
provides equal predictive skill as the mean of the observed data.
The correlation coefficient r gives an indication of the linear relation-
ship between observed and predicted data and is the most
common measure for assessing aquatic models (Arhonditsis and Brett, 2004). PRE is a measure of the relative deviation of simu-
lated from observed values and can be used to determine the bias in
predictions (Bennett et al., 2013). Finally, NMAE is both normalised
to the mean, enabling like comparisons between variables and is
absolute so that under and over estimations do not cancel each other out.

Initial manual calibration focused on refining input data by
adjusting the wind scaling factor and river inflow slope parameters
for each lake (the river slope is indicated as ϕinf in Hipsey et al.
(2017), and they are denoted as wind_factor and strmbslope in
the configuration file, respectively). Wind factor adjustment was
required where wind stations were located some distance from the
lake and/or to account for wind sheltering effects (Markfort et al.,
2010). River inflow slope was adjusted to correct the magnitude
of momentum and entrainment associated with plunging inflows.
For lakes where few or no light attenuation or Secchi depth read-
ings were available, Ken was also adjusted until simulated thermo-
cline depth matched that of observed data. Initial calibration was
carried out until an RMSE (calculated for all observed temperature
data over the simulation period) of less than 2°C was achieved.

4) Percent relative error (PRE):

\[ \text{PRE} = \frac{\sum_{i=1}^{N} (P_i - O_i) / O_i}{N} \times 100 \]  

(2-4)

5) Normalised mean absolute error (NMAE):

\[ \text{NMAE} = \frac{\sum_{i=1}^{N} |P_i - O_i| / O_i}{N} \]  

(2-5)

where \( N \) is the number of observations, \( O_i \) and \( P_i \), the “ith” observed
and model predicted data and \( \overline{O} \) and \( \overline{P} \) the mean observed
and model predicted data, respectively.

We chose a range of thermal metrics to assess model perfor-
ance at each site: observed full profile temperature data; epilimnion temperature; hypolimnion temperature; thermocline depth and Schmidt Stability (Idso, 1973). Schmidt Stability (\( S_T \)) and thermocline depth (\( \text{thermD} \)) were calculated for both model output
and observed thermistor data using Lake Analyzer (http://
lakeanalyzer.gleon.org/), an open source software tool that com-
putes indices of mixing and stratification for lakes and reservoirs
(Read et al., 2011). The comparison of \( \text{thermD} \) calculations was
included in the analysis as it is a simple, widely-used metric of
mixed layer depth, while acknowledging the calculation of \( \text{thermD} \)
can be challenging for weakly stratified and polymictic lakes. Also,
the approach used in Lake Analyzer identifies the strongest thermal
gradient, and may miss important thermal structure. \( S_T \) represents
resistance to mechanical mixing due to the potential energy
inherent in the stratification of the water column, calculated as:

\[ S_T = \frac{g}{A_s} \int_0^{z_0} (z - z_s) \rho_z A_z d z \]  

(2-6)

where \( g \) is the acceleration due to gravity, \( A_s \) is the surface area of
the lake, \( A_z \) is the area of the lake at depth \( z \), \( z_0 \) is the maximum
depth of the lake, and \( z_s \) is the depth to the centre of volume of the
lake, and \( \rho_z \) is the water density at depth \( z \). While not used as a
direct gauge of model performance, the daily Lake Number (\( LN \))
output as a GLM diagnostic parameter was also used in the cross
lake comparison analysis as a measure of the validity of the one-
dimensional assumption of the model. \( LN \) balances the strength of
stratification to wind induced mixing across the thermocline and is
a measure of the potential for mixing across the thermocline
(Imberger and Patterson, 1989).

\[ LN = S_T (z_t + z_h) \]  

(2-7)

where \( z_t \) and \( z_h \) are the depths to the top and bottom of the met-
alimnion, respectively, \( \rho_o \) is the average density of the hypolimnion
and \( u_c \) is the surface friction velocity.

2.4. Sensitivity analysis

Sensitivity of model output to nine parameters of mixing
and heat exchange was evaluated for each lake. Three of the parameters
influence surface heat and momentum exchange: bulk aerody-
namic coefficient for sensible heat transfer (\( C_H \)), bulk
aerodynamic coefficient for latent heat transfer \( (C_l) \) and coefficient of wind drag \( (C_D) \). The remaining six parameters control surface and hypolimnetic mixing: mixing efficiency for convective overturn \( (C_r) \), mixing efficiency of wind stirring \( (C_W) \), mixing efficiency of shear production \( (C_s) \), mixing efficiency of unsteady turbulence \( (C_T) \), mixing efficiency of Kelvin-Helmholtz turbulent billows \( (C_{K-H}) \), and mixing efficiency of hypolimnetic turbulence \( (C_{hyp}) \) (Table 2).

To gauge a response to parameter change, the one-at-a-time (OAT) method (Bruce et al., 2008) was adopted for the first stage of the MLCP, where the model was first run with the model default value to each parameter and then run again increasing and decreasing parameter values by 20%.

Sensitivity to changes in parameter values for each of the five lake variables used in the model assessment described above (temperature of the full water column, epilimnion, hypolimnion, thermD, and ST) was analysed. Normalised sensitivity coefficients \( (S_{ij}) \) to assess the relative sensitivity of variable \( i \) to parameter \( j \) were calculated according to:

\[
S_{ij} = \frac{\Delta C_i/C_{kj}}{\Delta \beta_j/\beta_{kj}}
\]

where \( \Delta C_i \) is the change in output variable \( i \), averaged over the simulation period, from the standard or reference value \( C_{kj} \) (Table 2) and \( \Delta \beta_j \) is the change in parameter \( j \) from the reference value \( \beta_{kj} \) (Fasham et al., 1990).

Sensitivity coefficients were then compared relative to ten characteristics describing the morphometry, climatic conditions and trophic state of the lakes. These properties were the maximum depth, lake volume, ratio of area to maximum depth, ratio of length to width, annual average inflow, residence time, mean air temperature, mean short wave radiation, mean wind speed and extinction coefficient (Table A3).

### 3. Results

#### 3.1. Model performance

Using the simulated results from running GLM with the standard set of parameters, five model fit metrics (RMSE, MEFF, \( r \), PRE and NMAE) were calculated for five data sets (full profile, epilimnion, hypolimnion temperature, thermD and ST) for each lake. The full set of results is provided in Appendix B (Table B1) with NMAE results given in Table 3. A comprehensive description of model performance for each lake can be found in the plots of modelled versus observed temperature data included in Appendix B.

An analysis of model performance in the prediction of temperature profiles (full profile) demonstrated a robust fit for GLM across the selected metrics, with an average RMSE of 1.34 °C, MEFF of 0.88, \( r \) of 0.96, PRE of –0.16 and NMAE of 0.11 (Table B1). The lakes with the lowest RMSE included Feagh, Tarawera and Emaiksoun. The highest RMSE values were calculated for Ravn, Ammersee and Woods. Ammersee also recorded the lowest values for MEFF along with NamCo and Toolik. All values of \( r \) were >0.9, with the exception of Toolik. The PRE values ranged from +18% for NamCo to –15%.

| Lake         | Full Prof. Temp. \( (\text{°C}) \) | Epi. Temp. \( (\text{°C}) \) | Hyp. Temp. \( (\text{°C}) \) | thermD \( (\text{m}) \) | ST          |
|--------------|-----------------------------------|--------------------------|---------------------------|------------------|-------------|
| Alexandrina  | 0.07                              | 0.07                     | N/A                       | N/A              | N/A         |
| Ammersee     | 0.19                              | 0.20                     | 0.13                      | 0.40             | 0.17        |
| Bielham      | 0.12                              | 0.13                     | 0.31                      | 0.18             | 0.45        |
| Bourget      | 0.08                              | 0.11                     | 0.07                      | 0.32             | 0.09        |
| Cannonsville | 0.10                              | 0.05                     | 0.15                      | 0.39             | 0.12        |
| Como         | 0.10                              | 0.17                     | 0.06                      | 0.64             | 0.19        |
| Constance    | 0.08                              | 0.09                     | 0.07                      | 0.11             | 0.16        |
| ElGergal     | 0.08                              | 0.06                     | 0.07                      | 0.30             | 0.27        |
| Emaiksoun    | 0.08                              | 0.08                     | N/A                       | N/A              | N/A         |
| Esthwaite    | 0.13                              | 0.11                     | 0.35                      | 0.15             | 0.24        |
| Feagh        | 0.06                              | 0.04                     | 0.09                      | 0.14             | 0.30        |
| Geneva01     | 0.09                              | 0.11                     | 0.04                      | 0.41             | 0.22        |
| Geneva03     | 0.08                              | 0.05                     | 0.04                      | 0.52             | 0.20        |
| GrosseDhunn  | 0.07                              | 0.05                     | 0.09                      | 0.37             | 0.09        |
| Harp         | 0.18                              | 0.12                     | 0.27                      | 0.68             | 0.19        |
| Iseo         | 0.08                              | 0.10                     | 0.07                      | 0.76             | 0.16        |
| Kinnet03     | 0.07                              | 0.07                     | 0.07                      | 0.28             | 0.20        |
| Kinnet97     | 0.05                              | 0.06                     | 0.05                      | 0.15             | 0.21        |
| Mendota      | 0.11                              | 0.10                     | 0.11                      | 0.30             | 0.23        |
| MiBold       | 0.08                              | 0.08                     | 0.06                      | 0.25             | 0.43        |
| Muggelsee    | 0.07                              | 0.06                     | N/A                       | N/A              | N/A         |
| NamCo        | 0.23                              | 0.17                     | 0.22                      | 0.28             | 0.35        |
| Noneida      | 0.04                              | 0.03                     | 0.06                      | 0.19             | 0.86        |
| Pusiano      | 0.14                              | 0.11                     | 0.26                      | 0.24             | 0.19        |
| Rappbode     | 0.14                              | 0.08                     | 0.12                      | 0.23             | 0.23        |
| Rassnitzersee| 0.17                              | 0.15                     | 0.23                      | 0.15             | 0.17        |
| Ravn         | 0.19                              | 0.14                     | 0.21                      | 0.27             | 0.34        |
| Rotorua      | 0.07                              | 0.08                     | 0.08                      | 0.09             | 0.43        |
| Stechlin     | 0.13                              | 0.11                     | 0.11                      | 0.33             | 0.14        |
| Tarawera     | 0.04                              | 0.04                     | 0.03                      | 0.27             | 0.10        |
| Toolik       | 0.25                              | 0.26                     | 0.25                      | 0.61             | 0.43        |
| Windermere   | 0.14                              | 0.23                     | 0.26                      | 0.22             | 0.21        |
| Woods        | 0.17                              | 0.17                     | N/A                       | N/A              | N/A         |
| Zurich       | 0.12                              | 0.09                     | 0.16                      | 0.42             | 0.17        |
| **Mean**     | **0.11**                          | **0.10**                 | **0.14**                  | **0.32**         | **0.25**    |
| **Median**   | **0.09**                          | **0.09**                 | **0.10**                  | **0.28**         | **0.20**    |
for Rassnitzersee. Because lakes had both positive and negative PRE (representing a temperature bias, warm and cold respectively) the mean PRE was −0.16%. The lowest absolute PRE was for Grossedhuenn (0.33%) which also performed well on all five measures of model fit.

In general, the model performance predicting the epilimnion temperatures was of similar magnitude to the full-profile temperatures (RMSE mean = 1.62 °C). By analysing the PRE, it is clear that the GLM tended to produce both warm and cold temperature biases in the epilimnion, slightly favouring a cold bias (mean $PRE = −0.84%$). For most lakes, model performance metrics were similar for the epilimnion as the full profile with the exception of Windermere and Zurich which performed worse and Oneida which performed better in the computation of epilimnion temperatures.

For the hypolimnetic temperature simulations, average RMSE and NMAE values were relatively low, 1.31 °C and 0.14 respectively. Typically small seasonal variation across all lakes led to greater percentage error between model and simulated data with both warm and cold temperature biases and a tendency to a warm bias (mean $PRE = 1.97%$). The mean $r$ value of 0.73 was the lowest of the three temperature-associated properties. Lakes with the highest model performance for hypolimnion temperature included Geneva01, Geneva03 and Como with the lowest being Rassnitzersee, Esthwaite and Blelham. Model efficiency values for the calculation of hypolimnion temperatures were poor with less than a third greater than 0.5 and 44% of lakes recording a value of less than zero.

Thermocline depth ($thermD$) was a difficult parameter to model with the poorest PRE and NMAE values (Table 3 & Table B1). Measures of model performance comparing calculations of observed and simulated $thermD$ ranged in value across the lakes with PRE values from −16% to +52% and NMAE ranging from 0.10 to 0.76 (Table 3 & Table B1). The PRE values indicate a bias towards over prediction of $thermD$ by the model compared to the observed data. This was most apparent in Lake Geneva over the winter months when GLM predicted full mixing (i.e. $thermD =$ lake depth) and the field data recorded a shallow $thermD (<5$ m). As the lake depth was >300 m this resulted in large relative error of greater than 6000%, leading to unfavourable mean measures of fit.

The NMAE values for calculation of $S_T$ were generally low. The higher values of NMAE were associated with lakes such as Ammersee, Oneida and Pusiano which all had relatively low $S_T$ during the simulated period. The mean $MEFF$ and $r$ were both quite high (0.83 and 0.96, respectively) indicating that the general seasonal patterns for $S_T$ prediction across the majority of lakes were well simulated by the model.

Analysis of the relationship between indices of model fit and input quality showed some correlation for the prediction of full profile, epilimnion and hypolimnion temperatures and $thermD$ (Table B2). Analysis of measures of PRE indicated a cold bias in prediction of both full profile and hypolimnion temperatures when input uncertainty is greatest (Fig. 4b). In addition, for lakes where the meteorological measurement station was near or at the lake edge, there was a warm bias and for lakes where meteorological input was sourced from further away, there was a cold bias (Fig. 4a). Similarly, there was a warm bias for the prediction of hypolimnion temperatures for lakes with high frequency meteorological data and a cold bias for lakes with daily meteorological data (Fig. 4c). Lakes with lowest input uncertainty associated with the estimation of $K_w$ corresponded with lowest values of $r$ with respect to the prediction of full-profile temperatures (Fig. 4d) and similarly lakes that had close to ideal ranking of overall input uncertainty scored the lowest values of $r$ for epilimnion temperatures (Fig. 4e). This would be attributed to the use of $K_w$ as a calibration parameter for
lakes where there were no measurements for light attenuation. High frequency observed data also correlated with high NMAE scores for the prediction of hypolimnion temperatures (Fig. 4f).

Analysis of model performance revealed a number of significant correlations linking model performance to lake characteristics (Table B3). For comparison of absolute model performance, the RMSE metric was used for temperatures and PRE for thermD and $S_f$. Whilst measurements of PRE can be a deceptive measure of model performance for lake variables where under and over-prediction occurs in equal measure, they are useful to observe patterns of bias in model prediction. A number of significant correlations between lake characteristics and model error are illustrated in Fig. 5 and Fig. 6 and described below.

The RMSE error associated with the prediction of both full profile and hypolimnion temperatures was generally higher for lakes with high light extinction ($K_o > 0.8 \text{ m}^{-1}$) and lower for clear lakes ($K_o < 0.3 \text{ m}^{-1}$) (Fig. 5a&b). A correlation was observed between the RMSE associated with the prediction of hypolimnion temperatures and lake depth (Fig. 5c), with deep lakes (>100 m) having the lowest values of RMSE (<1 °C). In terms of relative measures of model performance, for lakes with both low inflows (<10$^5$ m$^3$s$^{-1}$) and low levels of incident short wave radiation averaged over the entire simulation period (<120 Wm$^{-2}$) there was a cold bias in prediction of full profile and epilimnion temperatures, respectively (Fig. 5c&d). Whilst correlation was relatively low, there was some indication that for lakes with low residence time there was a cold bias in the GLM-predicted hypolimnetic temperatures (Fig. 5f).

For prediction of $S_f$ the lake depth, residence time and extinction coefficient all had a significant impact on model performance (Fig. 6a, b & c). Generally, clear deep lakes (>100 m), with residence times $> 2$ years recorded the lowest values of NMAE. A reverse pattern of correlation was observed for the prediction of thermD, with deep lakes having the highest values of NMAE and shallow lakes (<40 m) showing highest levels of thermD predictive accuracy (Fig. 6d). There was a small but significant trend where GLM over estimated $S_f$ in lakes with high incident short wave radiation (>200 Wm$^{-2}$) (Fig. 6e). For prediction of thermD, GLM tended towards over-prediction which was more pronounced in colder lakes (air temperature < 10 °C) (Fig. 6f).

Model performance for the prediction of thermD and $S_f$ was better for lakes when mean $L_N > 10$, while these lakes tended to record reduced measures of model fit for the prediction of epilimnion and hypolimnion temperatures (Fig. 7a,c,e,g). Conversely, for the small number of lakes with a significant proportion of the stratification period under a regime of $L_N < 1$, prediction of epilimnion and hypolimnion temperatures improved but thermD and $S_f$ decreased (Fig. 7b,d,f,h).

### 3.2. Sensitivity analysis

The sensitivity analysis (SA) on each of the nine surface exchange and mixing parameters highlighted differences both between lakes and thermal properties (Fig. 8a–e). For all three temperature metrics (full profile, epilimnion or hypolimnion) there was little sensitivity to perturbations in physical parameters, when the SA was averaged over the 2 year simulation period. There was some degree of sensitivity to changes in $C_D$ in the calculation of hypolimnion temperatures and to $C_e$ in the calculation of epilimnion temperatures. Sensitivity index ($S_I$) for prediction of both thermD and $S_f$, were significant (>1) across a broader range of lakes (Fig. 8d–e). While there was some variability across the lakes and parameters, model output for both thermD and $S_f$ had a consistent level of sensitivity to perturbations of $C_D$ and $C_e$. Additionally, for $S_f$ there was a consistent level of sensitivity to perturbations of $C_e$. 

---

**Fig. 5.** GLM model performance metrics for prediction of full profile temperature (a&d), epilimnion temperature (e) and hypolimnion temperature (b,c&f) against lake characteristics. Refer to Table 1 for lake acronyms.
The sensitivity of each parameter was compared to a gradient of physical and climate lake properties (Table C1-5) and a number of significant correlations were observed. For each thermal metric, the three most significant correlations to lake characteristics were compared (Fig. 9). A common significant (p < 0.05) trend was recorded for maximum lake depth (Fig. 9e,m). For the prediction of full profile and epilimnion temperatures, deeper and larger lakes were more sensitive to changes in $CK_H$ than small, shallow lakes (Fig. 9). Similarly, for the prediction of $thermD$, deeper lakes were more sensitive to changes in $Cc$, $Cw$ and $CK_H$ than shallow lakes (Fig. 9).

A significant correlation with air temperature indicated that lakes with low air temperatures were more sensitive to changes in $Ch$, $Cs$ and $CK_H$ than lakes in warm climates (mean air temperature > 10°C) for the prediction of full profile temperature (Fig. 9c), epilimnion (Fig. 9d) and hypolimnion temperatures (Fig. 9g). Lakes with low inflows were more sensitive to changes in $Cs$ for the prediction of hypolimnion temperatures than those with larger inflows (Fig. 9i). Finally, lakes with highest wind speed recorded greatest SI to $Ce$ in the prediction of $ST$ (Fig. 9m).

4. Discussion

Historically, lake modellers have adopted simple methods to justify model performance and suitability, rarely reporting statistical measures of model fit (Arhonditsis and Brett, 2004; Arhonditsis et al., 2006). For individual lake applications, these have been adequate to undertake scenario simulations and further our understanding of site specific dynamics. However, a common approach to model assessment, both in terms of metrics that should be applied and identification of a commonly agreed level of model performance, is necessary to further enhance model development (Bennett et al., 2013). Undertaking a standardised method of assessment of the community lake model, GLM, over a diversity of lakes has led to an improved level of understanding of the strengths and weaknesses in the predictive capacity of simple 1-D lake models. By first ascertaining an acceptable model error, we were able to elucidate the relation between model performance and data input uncertainty or lake characteristics (Fig. 4; Fig. 5).

The quality of input data was not as significantly related to model performance as expected. Lakes modelled using daily meteorological input, rather than hourly, did have the largest values of $NMAE$ in the prediction of full profile temperature and $thermD$ (Fig. 4), which is not surprising given the importance of diurnal forcing in 1-D model predictive capability. The greater the meteorological observation distance to the lake tended to result in both cold-biased temperatures and under prediction of $ST$ (Fig. 4). The cause of warm-biased temperatures and over-prediction of lake stability when meteorological observations were obtained near or on-lake requires further investigation (Fig. 4). The strong correlation between accuracy of $Kw$ measurements and model performance in the prediction of both full profile temperature and $thermD$ (Fig. 4) emphasises both the importance of light extinction in the determination of thermocline depth and the need to include measurements of $Kw$ in routine lake monitoring. The GLM can be coupled to water quality models such as the Aquatic EcoDynamics Model (AED: Hipsey et al., 2013) such that seasonal changes in $Kw$ would feedback in the model to potentially improve model prediction particularly in relation to $thermD$: this link is expected to further improve model accuracy in most circumstances.

The 1-D nature of the model implicitly assumes that the mixing within the lake can be constrained by processes acting in the vertical and that processes which vary in the horizontal, such as the degree of upwelling of the thermocline, have minimal impact on vertical transport. This assumption is quantified by computation of the Lake Number (Imberger and Patterson, 1989, eq. 2.7). As the $LN$ is a relative measure of the strength of stratification to surface wind energy, the 1-D model assumption is said to hold true for $LN >> 1$.

Fig. 6. GLM model performance metrics for prediction of thermocline depth (d,f) and Schmidt stability (a,b,c,e) against lake characteristics. Refer to Table 1 for lake acronyms.

Fig. 6.
Over the past three decades, the 1-D model approach has been applied to a wide diversity of sites due to its simplicity and tractability relative to 3-D models. However, given that LN can be highly variable, it has remained unclear what significance the 1-D assumption has on model prediction error for various lake attributes and under what conditions this assumption would no longer hold. The strong correlation ($r^2 = 0.70, 0.82$) between the percent of time LN $< 1$ during Fig. 7. GLM model performance metrics for prediction of epilimnion temperature (a,b), hypolimnion temperature (c,d), thermocline depth (e,f) and Schmidt stability (g,h) against Lake Number and %LN $< 1$. Refer to Table 1 for lake acronyms.
the stratified period and the model performance of both \textit{thermD} and \textit{ST} endorses the use of $L_N$ as an indicator of the validity of the 1-D model assumption, and should be considered when modellers are deciding on model suitability.

A comparison of \textit{PRE} against $L_N$ for the calculation of simulated versus observed $S_T$ indicated that lakes with mean $L_N < 1$ tended to underestimate $S_T$. For these lakes, the 1-D assumption as defined by $L_N$ does not hold. One would expect mixing to be underestimated and $S_T$ to be higher, unless the resulting warmer near surface temperatures led to greater heat losses by evaporation. Yeates and Imberger (2003) demonstrated that for lakes where deep mixing is important, a 1-D lake model mixing scheme similar to that used in

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure8}
\caption{Sensitivity indices for a) full profile temperature, b) epilimnion temperature, c) hypolimnion temperature, d) thermocline depth and e) Schmidt stability. The colour bar has been limited to a value of 1 so that any sensitivity index (SI) greater than one (indicating the percent response in thermodynamic metric is greater than the change in physical parameter) has been highlighted. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)}
\end{figure}
Fig. 9. Significant correlation between sensitivity indices of GLM physical parameters for the prediction of: full profile temperatures and (a) surface area, (b) lake depth and (c) air temperature; epilimnion temperature and (d) air temperature, (e) lake depth and (f) residence time; hypolimnion temperatures and (g) air temperature, (h) short wave radiation and (i) inflow; thermocline depth as a function of lake depth (j–l); and Schmidt stability as a function of wind speed (m).
GLM tended to overmix the water column and thus underestimate lake stability and therefore $S_f$. A solution put forward by Yeates and Imberger (2003) included a pseudo two-dimensional algorithm in the 1-D model DYRESM to parameterise internal and boundary fluxes. Similarly, Gaudard et al. (2016) proposed a method of adding a seasonal component in the parameterisation of internal seiches that led to improved accuracy in the prediction of deep mixing in the 1-D model SIMSTRAT. Whilst compromising computational efficiency, lake modellers could consider a similar approach when conditions for improved deep mixing accuracy are necessary. For example, this approach could be valuable where upwelling or internal nutrient loading is deemed important or when specific distribution phenomena such as deep chlorophyll maxima are the focus of the modelling study.

Further exploration of how individual lake properties relate to measures of model performance indicated the strongest correlations against $K_w$ and lake depth (Fig. 5; Fig. 6). Lakes with high $K_w$ (>0.5), recorded greatest error in the prediction of lake temperatures particularly in the hypolimnion. Whilst there was no significant correlation between the accuracy in prediction of epilimnion temperatures and lake depth, there was a strong positive correlation for measures of model performance in prediction of hypolimnion temperatures and lake depth. This is, for deeper lakes (>40 m) where surface mixing dynamics have less influence on hypolimnion temperatures, GLM predicts hypolimnion temperatures with greater accuracy. This suggests that while the surface thermodynamics are better represented by the model, prediction of rates of mixing across the metalimnion requires attention and further development to enable more confident prediction across the diversity of lake types. Relatively shallow, well-mixed lakes, such as Feeagh and Emaiksuon, had the highest overall model performance. These lakes are dominated by surface exchange with no thermocline and associated deepening.

The prediction of the lake thermocline depth proved harder to achieve than the lake temperatures. Particularly in moderately deep lakes, small relative deviations in predictions can result in large changes to error magnitude. As the GLM-predicted $\text{thermD}_{\text{avg}}$ was both deeper and shallower than the observed $\text{thermD}_{\text{avg}}$ in different lakes, there does not appear to be a consistent bias in the mixing algorithms, and rather, it may be driven by high sensitivity to input parameter uncertainty and require site specific calibration. The positive correlation between NMAE of thermocline prediction and lake depth was significant with best fit occurring for lakes less than 50–80 m deep (Fig. 6). A tendency to over-predict thermocline depth in the majority of lakes could be attributed to an over-prediction of penetrative heat and may be related to both the application of a standard minimum layer thickness for all lakes and the use of a single average $K_w$ value over 2 annual seasonal cycles. The positive correlation with $K_w$ indicates that a single $K_w$ for all seasonal conditions is not appropriate, particularly for lakes with high mean or seasonally variable $K_w$ values. A consideration for using a $K_w$ weighted towards the summer stratified period could be a solution or coupling to a water quality model with explicit light extinction feedback properties could improve thermocline prediction particularly in lakes with high light extinction ($K_w > 0.5$) (Shatwell et al., 2016).

The absence of strong sensitivity to parameterisation of surface exchange and mixing algorithms in the prediction of temperature profiles (Fig. 8) is indicative of the dominance of surface boundary conditions in the thermal budget of individual lakes and negative feedbacks in the surface heating sub-model. In contrast, the prediction of thermocline depth and Schmidt Stability were more sensitive to changes in parameterisation. In particular, the model was sensitive to the shear mixing efficiency and wind drag coefficient parameters. Both parameters are directly related to the transfer of wind energy to mixing. The errors in computing these terms again points to the need for more effort in parameterizing the processes operative when $L_0$ is low and shear increases across the thermocline. Additionally, wind increases in magnitude as it flows across a lake. This effect is important for small and large lakes and is not included when wind is modelled with bulk drag coefficients. Care should be taken in both the accuracy of wind speed measurements as well as the parameterisation and classification of these parameters in relation to lake characteristics to improve model performance across a wide variety of lake properties.

In general, simulations of deep lakes with large volumes and residence times were most sensitive to changes in mixing efficiency parameters (as measured by changes in $\text{thermD}_{\text{avg}}$ and $S_f$) (Fig. 9), which was expected since larger lakes require greater efficiency in transfer of surface momentum input to thermocline deepening and subsequent mixing. Lakes with low $K_w$ were most sensitive to changes in surface exchange parameters. This sensitivity is logical given that in lakes with low $K_w$, light will penetrate deeper causing a deeper thermocline. Processes which moderate depth of mixing in the epilimnion, such as convection, become important. Being able to model changing dynamics of lakes as $K_w$ changes with modified hydrology and altered loading of chromophoric dissolved organic matter is critical for quantifying the changes associated with climate variability (Zhang and Grun, 2000).

An appealing alternative to the minimal calibration presented here (i.e., input data refinement, wind factor and river inflow slope adjustment) will be the relaxation of the assumption of globally common parameter values for the core hydrodynamic parameters and the adoption of a Bayesian hierarchical calibration framework that reflects the more realistic notion that each lake (or group of lakes) is peculiar but shares some commonality of behavior with other lakes (Zhang and Arhonditsis, 2009; Cheng et al., 2010; Shimoda and Arhonditsis, 2015). The proposed approach represents a pragmatic compromise between system- and group-specific and globally common parameter estimates and may be a conceptually sound strategy to accommodate within- and among-lake variability in the context of model application within the global observational network (Fig. 10). Recent work has shown that the delineation of more homogeneous subsets of lakes with respect to their morphological characteristics/hydraulic regimes and their subsequent integration with hierarchical frameworks may give models with better predictive capacity (Cheng et al., 2010; Shimoda and Arhonditsis, 2015). In particular, sensitivity analysis patterns identified in this study could be used to identify groups with similarities in behavior (e.g., deep versus shallow lakes, high versus low water transparency) as well as to identify the candidate parameters for the calibration exercise. The prior distributions of the hyperparameters (or global priors) can be easily formulated on the basis of existing knowledge (e.g., field observations, laboratory studies, and information from the modelling literature) of the relative plausibility of their values. Moreover, the proposed incorporation of mathematical models into Bayesian hierarchical frameworks can also assist the effective modelling of systems with limited knowledge by enabling the transfer of information across systems. With the hierarchical model configuration, we can potentially overcome problems of insufficient local data by “borrowing strength” from well-studied lakes on the basis of distributions that connect systems in space (Zhang and Arhonditsis, 2009). Another advantage of a Bayesian calibration configuration will be the ability to express the input uncertainty in the form of probability density functions which can then be propagated through the model structure and may ultimately shape the moments of the posterior predictive distributions.

Through international collaboration, this work allowed us to test and to improve the process and performance of a 1-D open
source model by simulating thermal structure in lakes with varying physical and climatic characteristics. Initial efforts in setting up a collaborative network of lake modellers were rewarded with improved user support and feedback, refinements and testing to the development team. From its initiation as v1.0 in the MLCP, using feedback and re-coding by network members, the GLM evolved through numerous improvements to the current v2.2 described in this study. The study also identified the most sensitive parameters related to surface exchange and mixing that affect model prediction and therefore performance for each individual lake. These sensitivities could then be correlated to lake characteristics such as residence time, meteorological conditions and trophic status. Additionally, this work opens a new challenge for the community of limnologists involved in ecosystem modelling. Indeed the next step would be cross lake comparison projects including biogeochemical processes simulation using a similar open source community biogeochemical model such as the Framework for Aquatic Biogeochemical Models (FABM: Bruggeman and Bolding, 2014) and/or AED (Hipsey et al., 2013). The establishment of well-defined standards for modelling techniques (set up, output analysis), and a diversity of lakes and scientists provides enormous opportunity for further advances by aquatic ecosystem modellers. The significance of the MLCP resides in a common and collaborative approach to answering globally relevant lake science questions, and providing a benchmark for model performance and an associated parameter set that future applications can refer to.

Acknowledgements and contributions

GLM development and funding support for LCB, BDB, CB and MRH was provided by the Australian Research Council (ARC) (grants DP130104078 & LP130100756). Additional contributions from individuals and organisations as well as sources of data, provided from a variety of organisations are summarised in Appendix D. This study was made possible through the sharing of ideas, data and models across the AEMON and GLEON networks as well as discussions and working groups held during AEMON workshops and GLEON meetings.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.envsoft.2017.11.016.

References

Adams, J., 2012. Collaborations: the rise of research networks. Nature 490, 335–336.
Adrian, R., O’Reilly, C.M., Zagarasec, H., Baines, S.B., Hessen, D.O., Kellerf, W., Livingstone, D.M., Sommarug, R., Straile, D., Van Donk, E., Weyhenmeyer, G.A., Winder, M., 2009. Lakes as sentinels of climate change. Limnol. Oceanogr. 54, 2283–2297.
Anneville, O., Molinero, J.C., Rouiller, S., Gerdeaux, D., 2010. Seasonal and interannual variability of cladoceran communities in two per-alpine lakes: uncoupled response to the 2003 heat wave. J. Plankton Res. 32, 913–925.
Anneville, O., Domaizon, I., Kerimoglu, O., Ritzen, F., Jacques, S., 2015. Blue-green algae in a “greenhouse century”? New insights from field data on climate change impacts on cyanobacteria abundance. Ecosystems 18, 441–458.
