Modeling reservoir surface temperatures for regional and global climate models: a multi-model study on the inflow and level variation effects

Manuel Almeida¹, Yurii Shevchuk², Georgiy Kirillin³, Pedro M. M. Soares⁴, Rita M. Cardoso⁵, José Matos⁶, Ricardo Rebelo⁷, António Rodrigues⁸, Pedro Coelho⁹

¹, 7, 8, 9 Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisboa, 2825 - 516, Portugal
² MX Automotive GmbH, Berlin, 13355, Germany
³ Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin, 12587, Germany
⁴, 5 Instituto Dom Luís (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, 1749 - 016, Portugal
⁶ Stucky SA, Rue du Lac 33, 1020 Renens, Switzerland

Correspondence to: Manuel Almeida (mcvta@fct.unl.pt)

Abstract. The complexity of the state-of-the-art climate models requires high computational resources and imposes rather simplified parameterization of inland waters. The effect of lakes and reservoirs on the local and regional climate is commonly parameterized in regional or global climate modeling as a function of surface water temperature estimated by atmosphere-coupled one-dimensional lake models. The latter typically neglect one of the major transport mechanisms specific to artificial reservoirs: heat and mass advection due to in- and outflows. Incorporation of these essentially two-dimensional processes into lake parameterizations requires a trade-off between computational efficiency and physical soundness, which is addressed in this study. We evaluated the performance of the two most used lake parameterization schemes and a machine learning approach on high-resolution historical water temperature records from 24 reservoirs. Simulations were also performed at both variable and constant water level to explore the thermal structure differences between lakes and reservoirs. Our results highlight that surface water temperatures in reservoirs differ significantly from those found in lakes, reinforcing the need to include anthropogenic inflow and outflow controls in regional and global climate models. Our findings also highlight the efficiency of the machine learning approach, which may overperform process-based physical models both in accuracy and in computational requirements, if applied to reservoirs with long-term observations available. A relationship between mean water retention times and the importance of inflows and outflows is established: reservoirs with the retention time shorter than ~100 days, if simulated without in- and outflow effects, tend to exhibit a statistically significant deviation in the computed surface temperatures regardless of their morphological characteristics.

1 Introduccion

Numerical weather prediction (NWP) and climate modeling are essential tools in research and applied science applications (e.g., Bauer et al., 2015; Forster, 2017; Jacob et al., 2020). Motivated by the need to increase the reliability of climate and
weather projections, the core numerical models undergo continuous improvements aiming at the best compromise between model representativity and computational efficiency (Flato et al., 2013). Air-lake heat and moisture fluxes affect the near surface atmospheric layers and are essential to accurate estimation of the future climate or weather forecast. Therefore, parameterization of inland waterbodies in atmospheric modeling has quickly evolved to increase the accuracy of the land-atmosphere boundary layers (Bennington, 2014; Xue et al., 2017; Wang et al., 2019a).

According to previous studies, the presence of waterbodies affects significantly the turbulent heat exchange with the atmosphere (Philips, 1972; Bates et al., 1993; Niziol et al., 1995; Lofgren, 2006; Notaro et al., 2013; Wright et al., 2013). In northern latitudes, surface waters tend to absorb heat in summer and release it in autumn, damping the temperature fluctuations in their vicinity and creating both a lag in diurnal and annual cycles of the air temperature, as well as increased precipitation (Dutra et al., 2010; Nordbo et al., 2011; Samuelsson et al., 2010; Subin et al., 2012). Overall, missing the lake and reservoir effects has been shown to deteriorate the simulation results of regional and global climate simulations (Ljungemyr et al., 1996; Long et al., 2007; Deng et al., 2013; Dutra et al., 2010; Samuelsson et al., 2010; Subin et al., 2012; Le Moigne et al., 2016; Irambona et al., 2018).

Waterbodies display larger thermal inertia than the surrounding land areas due to the high specific heat capacity of water and the vertical turbulent heat transport from the water surface to its deeper layers. Furthermore, they absorb a higher fraction of solar radiation than land due to a lower albedo and a higher transparency. The heat storage and thermal characteristics of inland waterbodies, acting primarily but not only through water column stability, are influenced by bathymetry, surface area, turbidity, and ice conditions (Schertzer, 1997; Rouse et al., 2003, Oswald and Rouse, 2004; Magee and Wu, 2017). Surface heat fluxes, in particular the evaporation rate, are also affected by advection due to inflows and outflows and by water level (WL) fluctuations (Rimmer et al., 2011; Friedrich et al., 2018). These fluctuations are usually much more pronounced in reservoirs than in natural lakes. Herewith, neglecting of the aforementioned water budget variations may lead to errors in surface heat flux predictions, especially in reservoirs.

The progressive increase of the spatial resolution of general circulation models (GCM) and regional climate models (RCM) resulted in wide implementation of coupled one-dimensional (1-D) models simulating surface energy fluxes in waterbodies, neglecting however the variation of in-, outflows, and WL. The coupled lake and reservoir models differ among each other mainly by the vertical mixing parameterization, classified into three major categories: eddy diffusion models, turbulence models, and bulk mixed layer models. In eddy diffusion models, vertical turbulent mixing is defined by eddy diffusion, parameterized as a function of velocity and stratification strength in form of the gradient Richardson number (e.g., HOSTETLER model, Hostetler and Bartlein, 1990; SEEMOD, Zamboni et al., 1992; LIMNMOD, Karagounis et al., 1993; MINLAKE, Fang and Stefan, 1996; CLM, Oleson et al., 2004; CLM4-LISSS, Subin et al., 2012; WRF-Lake, Gu et al., 2015).

More complex approaches, based on k-ε turbulence model, parameterize eddy diffusion based on the Kolmogorov-Prandtl relationship (Svensson, 1978; Burchard et al., 1999; Goudsmitt et al., 2002; Stepanenko and Lykossov, 2005). Bulk mixed layer models rely on the self-similarity concept for the temperature-depth profile in the stratified layer and integral budgets for the mixed and bottom layers (Kraus and Turner 1967, Mironov et al., 2010). The performance of some of these models has
already been evaluated in modeling intercomparison studies (e.g., Perroud, 2009; Stepanenko et al., 2010; Stepanenko et al., 2013; Thiery et al., 2016; Huang et al., 2019, Wang et al., 2019). Generally, these intercomparison studies evaluated the model performance in application to one to three lakes, usually with very particular morphological characteristics, (e.g., very deep or very shallow), over a limited time period. Overall, the results of these studies had an important impact in the further development of the models. In particular, they highlighted the need for intercomparison research projects that include a larger number of waterbodies and a longer modeling simulation.

Data-driven models such as artificial neural networks (ANN) have not yet been considered for the parameterization of lakes in climate models. Nevertheless, they have been successfully used to estimate mean daily and hourly water temperatures in rivers (e.g., Chenard and Caissie, 2008, Hebert et al., 2014) and in lakes (Sharma et al., 2008, Samadianfard et al., 2016, Read et al., 2019). The approach is particularly advantageous when the modeled processes are complex and nonlinear (Sharma et al., 2008), as in the case of surface water temperatures (SWT). In view of the trade-off between results quality and computational efficiency, data-driven models have potential advantages in estimating the effect of lake inflows/outflows on SWT, that motivates their inclusion into model intercomparison studies.

Currently, the major challenge in the parameterization of lakes and reservoirs in climate models is the need to ensure that the models’ response is consistent and accurate considering the wide range of morphological characteristics and the high variability of the meteorological forcing. While incorporation of in- and outflows may crucially improve the quality of model predictions, the increased complexity can restrain extension of process-based models and require alternative data-based approaches.

In this study, we evaluate the importance of the energy transfers due to water inflows and outflows when modeling surface water energy fluxes in artificial reservoirs and elaborate a methodology to improve this essential aspect of RCM and GCM. For this purpose, we (i) model 24 Portuguese reservoirs by using four models: a 2-D model to define a calibrated and validated baseline scenario, two 1-D models without the parameterization of inflows/outflows and an ANN, (ii) assess the modeling error in SWT of lakes and reservoirs, potentially associated with atmosphere-lake interactions, and (iii) compare the performance and computational requirements of different approaches to predict the evolution of SWT in lakes and reservoirs.

2 Study area

Portugal is located in southern Europe and has a typical Mediterranean climate. Daily air temperatures vary from maximum daily means ranging from 13 °C in the central highlands to 25 °C in the southeast region, and minimum daily means that vary from 5 °C in the northern and central regions to 18 °C in the south (Soares et al., 2012a). Complex topography and costal processes define the spatial and temporal heterogeneity of precipitation, which differs from a relatively wet annual maximum above 2 800 mm/yr in the mountainous northwest to a much drier 400 mm/yr in the tendentially flat southeast (Soares et al., 2012b; Cardoso et al., 2013).

The 24 reservoirs selected for this study are almost entirely located in mainland Portugal, apart from Alto Lindoso (R19) and Alqueva (R24) reservoirs, which are shared with neighboring Spain (Fig.1).
Figure 1: Location of the simulated waterbodies (ordered according to the simulated mean volume from smallest to largest)

The reservoirs were selected for the study based on their water residence time (WRT) and morphological characteristics (volume, depth, surface area) (Table 1). Most of reservoirs are classified as warm monomictic, with a stratified period during the warmer months (May – September), and one mixing period each year, during the colder part of the year, from October to April. As exceptions, Cercosa (R1) and Torrão (R15) are weakly stratified, while Penide (R3), Penha Garcia (R4), Enxoé (R7) and Crestuma-Lever (R16) (a run-of-the-river hydropower scheme, located in the north coastal region) are well mixed during the entire year.
Table 1 Morphometric details of the reservoirs and main water uses

<table>
<thead>
<tr>
<th>Reservoir</th>
<th>Simulation period</th>
<th>Mean volume (hm³)</th>
<th>Full supply volume (FSV) (hm³)</th>
<th>Maximum depth (m)</th>
<th>Mean depth (m)²</th>
<th>Water surface area at the FSV (ha)</th>
<th>Watershed area (km²)</th>
<th>Mean inflow (m³/s)</th>
<th>Mean water residence time (days)</th>
<th>Main uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 Cerçosa</td>
<td>1994-2008</td>
<td>0.05</td>
<td>0.06</td>
<td>16.00</td>
<td>3.00</td>
<td>2.00</td>
<td>59.89</td>
<td>0.72</td>
<td>0.79</td>
<td>P</td>
</tr>
<tr>
<td>R2 Vouga Covo</td>
<td>1994-2008</td>
<td>0.10</td>
<td>0.20</td>
<td>14.00</td>
<td>1.67</td>
<td>12.00</td>
<td>53.41</td>
<td>0.00</td>
<td>2093.86</td>
<td>W</td>
</tr>
<tr>
<td>R3 Penide</td>
<td>1989-2008</td>
<td>0.11</td>
<td>0.50</td>
<td>9.00</td>
<td>0.72</td>
<td>69.00</td>
<td>3.73</td>
<td>33.63</td>
<td>0.04</td>
<td>P</td>
</tr>
<tr>
<td>R4 Penha Garcia</td>
<td>1989-2008</td>
<td>0.38</td>
<td>1.10</td>
<td>10.00</td>
<td>5.39</td>
<td>20.40</td>
<td>14.73</td>
<td>0.05</td>
<td>91.50</td>
<td>W, I</td>
</tr>
<tr>
<td>R5 Alto Cévado</td>
<td>1989-2008</td>
<td>1.17</td>
<td>3.30</td>
<td>21.00</td>
<td>6.60</td>
<td>50.00</td>
<td>101.23</td>
<td>4.73</td>
<td>2.85</td>
<td>P</td>
</tr>
<tr>
<td>R6 São Domingos</td>
<td>1994-2008</td>
<td>4.61</td>
<td>7.90</td>
<td>34.00</td>
<td>8.23</td>
<td>96.00</td>
<td>42.04</td>
<td>0.07</td>
<td>728.27</td>
<td>W, I</td>
</tr>
<tr>
<td>R7 Enxoé</td>
<td>1998-2008</td>
<td>6.48</td>
<td>10.40</td>
<td>12.00</td>
<td>5.07</td>
<td>205.00</td>
<td>60.54</td>
<td>0.27</td>
<td>275.67</td>
<td>W</td>
</tr>
<tr>
<td>R8 Monte Novo</td>
<td>1989-2008</td>
<td>10.75</td>
<td>15.27</td>
<td>19.85</td>
<td>5.50</td>
<td>277.40</td>
<td>260.75</td>
<td>0.15</td>
<td>830.80</td>
<td>W, I</td>
</tr>
<tr>
<td>R9 Funchão</td>
<td>1994-2008</td>
<td>24.37</td>
<td>47.72</td>
<td>36.00</td>
<td>13.26</td>
<td>360.00</td>
<td>211.58</td>
<td>3.33</td>
<td>84.57</td>
<td>I</td>
</tr>
<tr>
<td>R10 Bouçã</td>
<td>1989-2008</td>
<td>27.69</td>
<td>48.40</td>
<td>62.00</td>
<td>9.68</td>
<td>500.00</td>
<td>2601.71</td>
<td>44.34</td>
<td>7.23</td>
<td>P</td>
</tr>
<tr>
<td>R11 Fronhas</td>
<td>1989-2008</td>
<td>28.08</td>
<td>62.10</td>
<td>48.00</td>
<td>11.61</td>
<td>535.00</td>
<td>630.46</td>
<td>15.05</td>
<td>21.59</td>
<td>P</td>
</tr>
<tr>
<td>R12 Odeleite</td>
<td>1997-2008</td>
<td>42.58</td>
<td>130.00</td>
<td>47.00</td>
<td>18.06</td>
<td>720.00</td>
<td>347.27</td>
<td>2.84</td>
<td>173.51</td>
<td>W, I</td>
</tr>
<tr>
<td>R13 Azibo</td>
<td>1989-2008</td>
<td>45.62</td>
<td>54.50</td>
<td>43.00</td>
<td>13.29</td>
<td>410.00</td>
<td>92.56</td>
<td>0.79</td>
<td>670.93</td>
<td>W, I</td>
</tr>
<tr>
<td>R14 Pedrópio</td>
<td>2005-2008</td>
<td>86.16</td>
<td>106.00</td>
<td>25.00</td>
<td>9.60</td>
<td>1104.00</td>
<td>59160.00</td>
<td>39.94</td>
<td>24.97</td>
<td>P, I</td>
</tr>
<tr>
<td>R15 Torrão</td>
<td>1989-2008</td>
<td>91.57</td>
<td>124.00</td>
<td>56.00</td>
<td>19.08</td>
<td>650.00</td>
<td>3268.28</td>
<td>76.44</td>
<td>13.86</td>
<td>P</td>
</tr>
<tr>
<td>R16 Crestuma-Lever</td>
<td>1989-2008</td>
<td>101.08</td>
<td>110.00</td>
<td>13.00</td>
<td>8.47</td>
<td>1298.00</td>
<td>96932.81</td>
<td>423.97</td>
<td>2.76</td>
<td>P, W</td>
</tr>
<tr>
<td>R17 Caia</td>
<td>1989-2008</td>
<td>112.35</td>
<td>203.00</td>
<td>44.00</td>
<td>10.30</td>
<td>1970.00</td>
<td>563.26</td>
<td>2.45</td>
<td>530.77</td>
<td>W, I</td>
</tr>
<tr>
<td>R18 Santa Clara</td>
<td>1989-2008</td>
<td>205.74</td>
<td>485.00</td>
<td>72.00</td>
<td>24.42</td>
<td>1986.00</td>
<td>519.69</td>
<td>2.18</td>
<td>1091.19</td>
<td>P, F, W, I</td>
</tr>
<tr>
<td>R19 Alto Lindoso</td>
<td>1992-2008</td>
<td>274.57</td>
<td>390.00</td>
<td>92.00</td>
<td>36.38</td>
<td>1072.00</td>
<td>1510.93</td>
<td>39.44</td>
<td>80.58</td>
<td>P</td>
</tr>
<tr>
<td>R20 Alto Rabagão</td>
<td>1989-2008</td>
<td>317.82</td>
<td>569.00</td>
<td>84.00</td>
<td>25.72</td>
<td>2212.00</td>
<td>106.97</td>
<td>9.41</td>
<td>390.92</td>
<td>P</td>
</tr>
<tr>
<td>R21 Aqueira</td>
<td>1989-2008</td>
<td>335.59</td>
<td>423.00</td>
<td>76.00</td>
<td>21.15</td>
<td>2000.00</td>
<td>3063.29</td>
<td>87.45</td>
<td>44.42</td>
<td>P, F, W, I</td>
</tr>
<tr>
<td>R22 Cabril</td>
<td>1989-2008</td>
<td>344.79</td>
<td>719.00</td>
<td>120.00</td>
<td>35.54</td>
<td>2023.00</td>
<td>2416.32</td>
<td>38.69</td>
<td>103.14</td>
<td>P</td>
</tr>
<tr>
<td>R23 Castelo do Bode</td>
<td>1989-2008</td>
<td>859.46</td>
<td>1095.00</td>
<td>96.00</td>
<td>33.27</td>
<td>3219.00</td>
<td>3964.09</td>
<td>64.94</td>
<td>153.18</td>
<td>P, W, F</td>
</tr>
<tr>
<td>R24 Alqueva</td>
<td>2005-2008</td>
<td>2974.66</td>
<td>4150.00</td>
<td>76.00</td>
<td>16.60</td>
<td>25000.00</td>
<td>55289.00</td>
<td>38.25</td>
<td>900.02</td>
<td>I, W, P</td>
</tr>
</tbody>
</table>

P – Power generation; W – Water supply; I – Irrigation; F – Flood prevention

(1) The mean depth results from the division of the mean volume and the mean water surface area.
(2) The mean water residence time is the ratio between the mean volume of the reservoir and the mean inflow.

3 Models and data

3.1 Forcing and calibration data

The baseline scenario was forced by monthly records of inflow and discharge for the period 1989-2008. To characterize inflow daily temperatures of 70 reservoir tributaries, a total of 31 air and water temperature linear regressions were additionally computed from 8 492 pairs of values ($\bar{x} = 274$; SD ± 565). The mean $R^2$ considering all regressions varied from 0.75 to 0.90 ($\bar{x} = 0.82$; SD ± 0.03). The calibration of the baseline scenario was performed on 677 water temperature profiles ($\bar{x} = 53$ per
reservoir) and 3 738 surface observed values (\( \bar{x} = 163 \) per reservoir). The hydrometric and water quality data was collected by the Portuguese Environmental Agency, Energies of Portugal, and the Alqueva Development and Infrastructure Company and is available from: www.snirh.ambiente.pt.

### 3.1.1 Meteorology

The hourly meteorological datasets of air temperature, relative humidity and wind velocity used as forcing of reservoir models were produced by a high-resolution (9 km horizontal grid spacing) simulation with the Weather Research and Forecasting model (WRF; Skamarock et al., 2008), forced by 20 years of ERA-Interim reanalysis (1989-2008), nested in a domain with a 27 km x 27 km cell size. A more detailed description of the model set-up and simulation results are provided by Soares et al., (2012a) and Cardoso et al., (2013). The WRF hindcast simulation results were extensively validated for inland surface variables, namely: temperatures and precipitation in Portugal (Soares et al., 2012a), Iberian precipitation (Cardoso et al., 2013), and wind (Soares et al., 2014; Rijo et al., 2018; Nogueira et al., 2019). Cloud cover datasets were derived from mean monthly values described in the climatological normal of Portugal (1951-1980), while solar shortwave radiation was computed with an algorithm based on the EPA method (Thackston and Parker, 1971). Cloud cover reduction of shortwave radiation uses the approach defined by Wunderlich (1972). The daily meteorological datasets, also used to force the models, correspond to the daily mean values obtained from the hourly meteorological datasets.

### 3.2 Models/scenarios

To evaluate the importance of inflow and outflow in SWT simulations, a 2-D numeric model and two 1-D models were applied. Since the model validation was limited by the scarcity of temperature profile measurements and observed time series of SWT, a major challenge of this study consisted in development of realistic baseline scenarios (forcing data and targets) having the necessary continuity and heterogeneity to evaluate the performance of different models. To overcome this limitation, a well-established 2-D model, CE-QUAL-W2 version 3.6 (Cole and Wells, 2008) was validated with observed data and used to create the baseline scenario, forced with daily and hourly meteorological datasets (Table 2) covering a period of 20 years, from 1989 to 2008 (with the exceptions described in Table 1). The 2-D model, forced with daily meteorology and monthly inflows and outflows, was calibrated by minimizing the Mean Absolute Error (MAE) between simulated water temperature profiles, and measurements spanning the period from 1989 to 2008 made in each reservoir, in all cases near the dam.

An alternative baseline scenario was produced by forcing the model with hourly meteorology (daily values were used for the first one), enabling evaluation of the sub-daily convection effects on the overall results. Both daily and hourly baseline scenarios were designated “W2 Reservoir”. To assess the importance of heat transfer and mixing within the waterbodies, the two “W2 Reservoir” scenarios were modified and simulated in a steady-state “constant mass budget” excluding precipitation, inflows or outflows. These steady-state scenarios were designated “W2 Lake”. Apparently, “W2 Lake” simulations maintain a constant water level, corresponding to the Full Supply Level (FSL). SWT time series were compared using statistic error measures (see Sect. 3.3 for more details)
which allowed the assessment of the relation between reservoir WRT and the error that results when the advection due to inflows and outflows is neglected (as mentioned in the introduction, a common feature of contemporary GCMs and RCMs).

While generally accurate, the use of calibrated 2-D models in the scope of complex GCMs and RCMs is restricted by high computational costs. Therefore, the next step of the analysis aimed at evaluation of more computationally effective 1-D models, typically used to parametrize waterbodies within GCMs and RCMs. The reservoirs were simulated with a 1-D eddy diffusion model based in the approach considered by Hostetler and Bartlein (1990) and a 1-D bulk mixed layer model (FLake), both forced with hourly and daily meteorological data (Table 2).

Table 2 Meteorologic datasets considered in the modeling process

<table>
<thead>
<tr>
<th>Model</th>
<th>Air temperature (°C)</th>
<th>Relative humidity (%)</th>
<th>Vapor pressure (mbar)</th>
<th>Wind velocity (m/s)</th>
<th>Wind direction (rad)</th>
<th>Cloud cover</th>
</tr>
</thead>
<tbody>
<tr>
<td>CE-QUAL-W2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1-D Hostetler</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1-D FLake</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>ANN</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The eddy diffusion model considers the vertical variation of both eddy diffusion and cross-sectional area. Simulations were undertaken using the maximum depth. In turn, FLake operates with volume-integrated equations. Accordingly, its simulations were performed based on the mean reservoir depth.

Results obtained with the 1-D models, without any reservoir-specific calibration, were compared with the baseline scenarios obtained with the 2-D model (W2 Reservoir and W2 lake scenarios). In addition to the 1-D models, SWT in all the reservoirs was modeled with an artificial neural network (ANN) trained using the momentum gradient-based optimization algorithm (Qian, 1999). SWT from both daily and hourly 2-D baseline scenarios (“W2 Reservoir”), covering the period from 1989 to 2004 and the predictor variables described in Table 3, were used to improve the input data dimension (Fig. 2).

Figure 2. Schematic and simplified representation of the ANN preparation concept.
Two different temporal sampling of the input meteorological data, daily and hourly, were used to train and validate the ANNs. 80% of the data was used for finding optimal network weights (of which 70% were directly applied in the training and 30% were employed in validation). These 80% covered 16 years from 1989 to 2004. The daily discretization resulted in a dataset with N = 5,843 entries, while the hourly discretization produced N = 140,232. The remaining 20% of data had no intervention in the search for optimal network weights and covered the period from 2005 to 2008. This period, considered for the ANN forecast included three dry years, 2005, 2007, 2008 and one wet year, 2006. All the years were warm except for the cold year of 2008. The daily discretization resulted in a dataset with N = 1,461 entries, while the hourly discretization produced N = 35,064. When the reservoir total simulation period (see Table 1) was shorter than 20 years, the dimension of the test dataset was preserved, and the training and validation data sets were reduced. The raw input data used to train the networks included: the SWT obtained from the baseline scenarios prepared with the 2-D model, air temperature (Tair), inflow water temperature (Tbr), dew point temperature (Tdew), relative humidity (HR), and wind velocity (u2). In order to improve model performance additional time series were included as input. They were defined to provide implicit information about seasonal changes (Table 3).

### Table 3 Predictor variables considered for the training and validation of the ANN

<table>
<thead>
<tr>
<th>Temporal sampling</th>
<th>Predictor variables</th>
<th>Total number of predictor variables</th>
</tr>
</thead>
</table>
| **Daily**         | Meteorological variables = Tair; Tbr; Tdew; HR; u2;  
|                   | Day of year (1 to 365 or 366); week index number (1 to 52 or 53); month index number (1 to 12);  
|                   | Cosine (2π*(day index number/365)); Sine (2π*(day index number/365));  
|                   | Cosine (2π*(week index number/52)); Sine (2π*(month index number/52));  
|                   | Cosine (2π*(week index number/26)); Sine (2π*(month index number/26));  
|                   | Cosine (2π*(month index number/12)); Sine (2π*(month index number/12));  
|                   | Moving average of the meteorological variables with a window of 31 days;  
|                   | Moving variance of the initial meteorological variables with a window of 31 days; | 26 |
| **Hourly**        | Meteorological variables = Tair; Tbr; Tdew; HR; u2;  
|                   | Hour index number, day index number; week index number; month index number  
|                   | Cosine (2π*(hour index number/24)); Sine (2π*(hour index number/24));  
|                   | Cosine (2π*(day index number/365)); Sine (2π*(day index number/365));  
|                   | Cosine (2π*(week index number/52)); Sine (2π*(month index number/52));  
|                   | Cosine (2π*(week index number/26)); Sine (2π*(month index number/26));  
|                   | Cosine (2π*(month index number/12)); Sine (2π*(month index number/12));  
|                   | Moving average of the initial meteorological variables with a window of 31 days  
|                   | Moving variance of the initial meteorological variables with a window of 31 days  
|                   | Moving average of the initial meteorological variables with a window of 744 days  
|                   | Moving variance of the initial meteorological variables with a window of 744 days | 39 |

* Cosine and sine series reproduce cyclical yearly variations that can be recognized by the ANN in the inputs without any breaks (as is the case if linear series such as the day of year are used: large jump from 365 to 1 in the beginning of a new year).
The input time series were subsequently standardized through removal of the mean and scaling to unit variance. After the initial tests, which included different network architectures, backpropagation algorithms, regularization strategies, learning rate rules, activation functions, parameter initialization, and the extraction and transformation of features from the input meteorological data, the algorithm was selected, whose results were in the best agreement with SWT from the baseline scenario simulated with the 2-D model (see Sect. 3.2.4 for more details).

Beside accuracy, the computation time can also be a critical factor in the suitability of models to be used within GCMs or RCMs. The simplified 1-D models considered in this study have a clear advantage regarding the computation time when compared with more complex 1-D and 2-D approaches – a condition that was at the core of their development and that is directly linked with the neglect of inflows and outflows. Recognizing the importance of computational efficiency, the analysis included the quantification of the overall computation times for the 1-D models and for the ANN. This evaluation was produced with a 2.21 GHz Quad-Core Intel Core i7 (memory: 16 GB 1600 MHz DDR3), by repeating 20 times each simulation.

3.2.1 2-D Water quality and hydrodynamic modeling – CE-QUAL-W2

Due to the lateral and layer averaging of the governing equations, the 2-D hydrodynamic and water-quality model CE-QUAL-W2, version 3.6 (Cole and Wells, 2008) is particularly suitable for modeling relatively long and narrow waterbodies, where transverse variations in velocities, temperatures and constituents are negligible. The model has been widely applied to stratified water surface systems such as lakes and reservoirs around the world, including Portugal (e.g., Diogo et al., 2008, Almeida et al., 2015). In order to illustrate the performance of CE-QUAL W2 in reservoir thermal simulations, Cole and Wells (2008), describe the calibration results obtained for 70 reservoirs. In their study, the MAE obtained for all reservoirs was smaller than 1.0 ºC, and for many of them much smaller. The result can be considered outstanding, especially considering that errors were partially related to the quality of the boundary conditions and forcing meteorological data. The model utilizes implicit solution for the transport equation of vertical momentum, with the W2 algorithm for the form of the vertical turbulence closure and with the Ultimate algorithm (Cole and Wells, 2008). Surface heat exchange was computed with the term-by-term algorithm described by Cole and Wells (2008).

3.2.2 Eddy diffusion model – Hostetler model (HLM)

The governing equation for the 1-D eddy-diffusion model is based on Hostetler and Bartlein (1990):

\[
\frac{\partial T}{\partial t} = \frac{1}{A(z)} \frac{\partial}{\partial z} \left( A(z) \left[ k_m + K(z, t) \frac{\partial T}{\partial z} \right] \right) + \frac{1}{A(z) C_w} \frac{\partial \Phi A(z)}{\partial z}
\]

where \( T, t, z, A, k_m, K, C_w, \) and \( \Phi \) are: water temperature (ºC), time (s), depth (m), area (m²), molecular diffusion (1.39·10⁻⁷ m²·s⁻¹), eddy diffusion (m²·s⁻¹), the volumetric heat capacity of water (J·m⁻³·ºC⁻¹) and a heat source term (W·m⁻²), respectively.
Within the model, eddy diffusion is computed at each depth with the analytical representation developed by Henderson-Sellers (1985) as a function of the 2 m wind velocity ($u_2$), and a latitude-dependent parameter of the Ekman profile.

The surface boundary condition is described by the following equation:

$$\rho C_w[k_m + K(z, t)] \frac{\partial T}{\partial z}|_{z=0} = q_n$$  \hspace{1cm} (2)

The net surface heat flux ($q_n$) (W·m$^{-2}$), which is the algebraic sum of solar radiation, atmospheric radiation, latent and sensible heat fluxes and back radiation, was computed with the equilibrium temperature approach defined by Edinger et al., (1968), while latent and sensible heat fluxes were computed explicitly from surface water temperature with the same expressions defined in Cole and Wells (2008). In this study the heat transferred from the sediments to the water column has been neglected. Accordingly, the bottom boundary condition takes the following form:

$$\rho C_w[k_m + K(z, t)] \frac{\partial T}{\partial z}|_{z=\text{max depth}} = 0$$  \hspace{1cm} (3)

The solution of the heat diffusion equation was obtained resorting to the implicit numeric Crank-Nicholson scheme with centered differences in space and time. Convective mixing is conceptualized by a full-depth mixing scheme that detects buoyancy-induced instabilities and mixes all layers from the surface down to the unstable layer while preserving the available energy. HLM has predicted accurately water temperature profiles of several lakes located in the United States (e.g., Hostetler and Bartlein, 1990; Hostetler and Giorgi, 1995) and a modified version of the model is currently used in the Community land model that is coupled with the International Centre for Theoretical Physics (ICTP) Regional Climate Model, version 4 (RegCM4) (Bennington et al., 2014). The model governing equation and the parameterization of eddy diffusion is also the base of the 1-D lake model included in the Weather Research and Forecasting (WRF) model (LISSS) (Xiao et al., 2016).

### 3.2.3 FLake model

The FLake model was developed for use in NWP and is currently implemented in several NWP models, for example: the Consortium for Small-scale Modeling (COSMO) from the German Weather service (Mironov et al., 2010); the High Resolution Limited Area Model (HIRLAM), from the Finnish Meteorological Institute; the Icosahedral Nonhydrostatic (ICON), from the German Weather service; or the Integrated Forecast System (IFS), from the European Centre for Medium-Range Weather Forecasts. The model has also been used to evaluate the effects of lakes in the climate system (Gula and Peltier, 2012; Le Moigne et al., 2016) and in the future scenarios for lake water temperature and mixing regimes (Kirillin, 2010; Shatwell et al., 2019). Conceptually, the FLake belongs to the family of “bulk” mixed layer models (Kraus and Turner 1967), widely used in lake studies (e.g. DYRESM: Magee and Wu, 2017; GLM: Hipsey et al., 2019; CSLM: MacKay, 2019). A distinguishing feature of the FLake consists in the extension of the “bulk” approach on the stratified part of the lake water.
column from the base of the mixed layer down to the lake bottom. The extension relies on the concept of the thermocline self-
similarity (Kitaigorodskii and Miropolsky, 1970), i.e. preserved shape of the temperature profile in the stratified part of the
water column. In FLake, a waterbody can be represented as a two-layered system, where the vertical profile of water

\[
T = \begin{cases}
    T_s & \text{at } 0 \leq z \leq h \\
    T_s - (T_s - T_b) \cdot \Phi_T(z) & \text{at } h \leq z \leq D
\end{cases}
\]

where \( z \) is the vertical coordinate, \( h \) is the surface mixed layer depth, \( D \) is the lake depth, \( T_s \) is the mixed layer temperature
and, \( T_b \) the temperature at the water-sediment interface in the bottom.

### 3.2.4 Artificial Neural-Network

The prototyping and building of the ANN was implemented with the python library NeuPy (Shevchuk, 2015). NeuPy uses
Tensorflow (an open source platform for machine learning) as a computational backend for deep learning models (Abadi et
al., 2015). The momentum algorithm used in the selected ANN is an iterative first order optimization method that uses gradient
calculated from the average loss of a neural network (usually the mean squared error). The “momentum” applies to information
about past gradients during the training in the way that promotes a gradual transition in the balance between stability and rate
of change (Qian, 1999).

In addition to the input and output layers, the chosen network has one hidden layer with 24 nodes. Each of these used Rectified
Linear Activation Functions (ReLu). Training data was shuffled before training, weights were randomly initiated, and the loss
function included the MSE (see further below) to measure the accuracy of the results. Additionally, it used L2 regularization
(the adopted regularization constant was 0.002). The step decay algorithm was used to regularize the learning rate (initial value
= 0.05, reduction frequency = 750).

### 3.3 Evaluation metrics

Model assessment was undertaken relying primarily on the mean bias (Bias), the mean absolute error (MAE), the root mean
square root error (RMSE), the centered root mean square error (RMSEC) and the coefficient of determination (R²). The metrics
were computed with the following equations, where \( m_i \) and \( o_i \) are the modeled and observed values, and \( \bar{m} \) and \( \bar{o} \) are their
means:

\[
Bias = \bar{m} - \bar{o}
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |m_i - o_i|
\]
When assessing differences between the models, \( m_i \) and \( o_i \) are the values obtained for reservoir and for lake simulations, respectively.

4 Results

4.1 Models calibration/validation

The wind sheltering coefficient reducing the wind effect on the surface fluxes was found to be the most relevant calibration parameter for the 2-D model. The overall mean value of the wind sheltering coefficient was of 0.6, with a minimum value of 0.1 in Bouçã (R10) and a maximum of 1.0 in Fronhas (R11), Pedrogão (R14), Agueira (R21) and Alqueva (R24) reservoirs.

The light extinction coefficient was also adjusted during calibration with its value varying from 0.25 to 1.0 (\( \bar{x} = 0.38; \text{SD} \pm 0.22 \)). Other coefficients, involved in the water temperature calibration, had a significantly smaller effect and were kept with their default values: 1 m\(^2\)·s\(^{-1}\) for longitudinal eddy viscosity and diffusivity; 70 m\(^2\)·s\(^{-1}\) for Chézy coefficient, and 0.45 for solar radiation percentage absorbed in the surface layer (\( \beta \)). The water temperature profiles and surface temperature time series obtained at the downstream edge of the reservoirs (near the dams) suggested that the reservoirs were reasonably well simulated by the 2-D model after the calibration forced with daily meteorology. When comparing the model results with a total of 3 608 observed surface temperature values (Fig. 3a), the MAE varied from 0.87 °C to 3.54 °C (\( \bar{x} = 1.89 \) °C; \( \text{SD} \pm 0.40 \) °C) and the RMSE varied from 1.49 °C to 4.58 °C (\( \bar{x} = 2.41 \) °C SD \( \pm 0.50 \) °C). The three major RMSE values were obtained for reservoirs with short WRT, suggesting that the major source of inaccuracy was attributed to the inflow temperatures (R11: 4.58 °C, WRT: 21.6 days; R1: 3.44 °C, WRT: 0.79 days; R4: 3.44 °C, 91.50 days). For the 677 observed water temperature profiles (Fig. 3b), the MAE varied from 1.64 °C to 2.62 °C (\( \bar{x} = 2.14 \) °C; \( \text{SD} \pm 1.35 \) °C) (Fig. 3c) and the RMSE varied from 1.77 °C to 3.52 °C (\( \bar{x} = 2.46 \) °C; \( \text{SD} \pm 1.49 \) °C) (Fig. 3d).

Additionally, daily and hourly SWT results were compared with the observed SWT values in order to assess the performance of the different models and the influence of the model time resolution. Simulations with the daily time step had a similar accuracy in all models (Table 4), with the Hostetler model results being slightly closer to the observed time series. Daily metrics were obtained by comparing SWT values observed at a specific hour in the reservoirs with the daily averages obtained with the model. Therefore, they tend to level the metrics results for each model, in particular the bias (Table 4). In simulations...
with the hourly time step, the 2-D Reservoir model performed expectedly the best among the process-based models, highlighting the robustness of the baseline scenario. FLake had a worse performance than HLM, considering the hourly results, which can be attributed to differences in the conceptualization of diurnal variations of SWT. Complete mixing within the mixed layer of FLake model reduced the diurnal temperature variations (Martynov et al., 2010). The differences in the diurnal SWT variability were observed across all reservoirs.

The ANN performed best in terms of similarity to observations. The results obtained for each dataset show that the RMSE obtained with the 2-D model and with the ANN had less variations across all reservoirs than the results obtained with the 1-D models (Fig. 4). This result can be attributed to the wind forcing treatment by 1-D models. The latter do not consider the wind sheltering effect, which was the most relevant parameter for calibration of the 2-D model, reducing the wind velocity by around 34%. The response to wind stress of elongated reservoirs depends strongly on whether the dominant wind is directed across or along the reservoir main axis (Mackay, 2019). Therefore, wind direction can significantly affect SWT predictions by influencing surface mass and heat fluxes, which effect is evaluated in more detail in Sect. 4.3. Additionally, the comparison of W2 Reservoir and W2 Lake scenarios suggests that the SWT of reservoirs R3, R10 and R22 were particularly affected by in- and outflows and/or water level variations. RMSE values reached 2.7 °C, 1.2 °C and 0.9 °C, respectively (Fig. 4).

The ensemble analysis of the results obtained with the 1-D models for the period 2005-2008 (Fig. 4, lower right Panel) shows that the models had a similar performance. Overall, results highlight the large interannual variability of reservoir SWT and emphasize the difficulties that arise when modeling these systems.
Table 4. MAE, RMSE and bias (with standard deviation) between observed SWT values and SWT values obtained with all models, forced with daily and hourly meteorology for the 24 waterbodies

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Model</th>
<th>MAE, °C (mean)</th>
<th>RMSE, °C (mean)</th>
<th>Bias, °C (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
</tr>
<tr>
<td>1989-2008</td>
<td>W2 Reservoir</td>
<td>1.89 (± 0.40)</td>
<td>1.85 (± 0.46)</td>
<td>2.45 (± 0.50)</td>
</tr>
<tr>
<td></td>
<td>W2 Lake</td>
<td>2.13 (± 0.69)</td>
<td>2.16 (± 0.71)</td>
<td>2.71 (± 0.75)</td>
</tr>
<tr>
<td></td>
<td>Hostetler</td>
<td>1.72 (± 0.62)</td>
<td>1.93 (± 0.64)</td>
<td>2.27 (± 0.62)</td>
</tr>
<tr>
<td></td>
<td>FLake</td>
<td>1.75 (± 0.56)</td>
<td>2.67 (± 0.72)</td>
<td>2.32 (± 0.56)</td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2005-2008</td>
<td>W2 Reservoir</td>
<td>1.89 (± 0.40)</td>
<td>1.81 (± 0.35)</td>
<td>2.45 (± 0.50)</td>
</tr>
<tr>
<td></td>
<td>W2 Lake</td>
<td>2.08 (± 0.74)</td>
<td>2.14 (± 0.64)</td>
<td>2.62 (± 1.02)</td>
</tr>
<tr>
<td></td>
<td>Hostetler</td>
<td>1.75 (± 0.69)</td>
<td>1.94 (± 0.66)</td>
<td>2.25 (± 0.85)</td>
</tr>
<tr>
<td></td>
<td>FLake</td>
<td>1.66 (± 0.53)</td>
<td>2.63 (± 0.67)</td>
<td>2.22 (± 0.86)</td>
</tr>
<tr>
<td></td>
<td>ANN</td>
<td>1.77 (± 0.48)</td>
<td>1.78 (± 0.44)</td>
<td>2.28 (± 0.71)</td>
</tr>
</tbody>
</table>
4.2 Model intercomparison: 1-D models and the ANN

4.2.1 Model accuracy

In order to evaluate the consistency and accuracy of the models, the SWT time series were compared with the baseline scenario W2 Reservoir (Table 5). When forced with hourly meteorological data, the ANN reduced significantly the error in SWT predictions for the period 2005-2008 (Fig. 5). This fact emphasizes both the potential of data-driven models to simulate the
SWT and the importance of the temporal resolution of the training data sets. Overall, the ANN results remained consistent across both dry and wet seasons, reducing the annual RMSE to 0.86 °C (± 0.31; daily meteorology) and to 0.71 °C (± 0.21; hourly meteorology), as well as the interannual variability of RMSE. Results of both 1-D models were similar to each other (Fig. 5), with both models reproducing well the seasonal variation of SWT and exhibiting a significant variation between the simulations performed with daily and hourly meteorological forcing and during the wet and dry seasons. Nevertheless, FLake and HLM demonstrated a reasonable performance (Table 5), similar to that reported in previous studies (Stepanenko et al., 2010, Stepanenko et al., 2013, Thiery et al., 2016).

The two 1-D models revealed a contradictory behavior with respect to the temporal resolution. In contrast to the HLM, the FLake model had a slightly better performance with the daily than with the hourly meteorological input, which can also be attributed to differences in the conceptualization of diurnal variations of SWT. Therefore, with daily simulations, these differences between models are much less pronounced.

Considering the bias values obtained for each reservoir (Fig. 5), FLake and the HLM underestimated the SWT in 83% and in 54% of cases, respectively. The negative SWT bias can be primarily ascribed to the overestimation by 1-D models of the wind stress effect on the surface heat flux due to ignoring the wind direction variability over wind-sheltered elongated reservoirs. The lower bias in the HLM than in FLake is more consistent with the 34% wind velocity reduction obtained in the 2-D model calibration, suggesting the FLake performance was affected by other factors, such as the diurnal SWT variability.

The analysis of the mean annual RMSE obtained with the 1-D models and with the 2-D Lake results considering the hourly meteorology indicate that Penide reservoir (R3), with a WRT of approximately 0.04 days, had the highest mean RMSE, clearly highlighting the relevance of inflows and outflows in SFTs. HLM had a worse performance for reservoirs R3, R11, R14, R1 and for the six deepest reservoirs, R19, R20, R21, R23, R22 and R24, which indicates that the vertical heat diffusion was not optimally computed (Fig. 5b). Specifically, the explicit approximation of convective mixing in the HLM model by convective adjustment of unstable temperature profiles is apparently too rough, to simulate convective mixing in deep lakes (Bennington et al., 2014).

The analysis of the ensemble of RMSE results obtained with all models (Fig. 5e) reveals a high variability among SWT predictions by different models. In general, the performance of 1-D models suggests that their simplified nature and the neglect of inflows/outflows can impose high uncertainties in SWT predictions (Table 5).
Table 5 Evaluation of model performances. 2-D baseline scenario (W2 Reservoir) versus simulated SWTs with the exclusion of inflows and outflows (W2 Lake), HLM, FLake and ANN. Models forced with daily and hourly meteorology for the 24 waterbodies (RMSE, bias with the standard deviation)

<table>
<thead>
<tr>
<th>Model</th>
<th>Period</th>
<th>RMSE, ºC (mean)</th>
<th>Bias, ºC (mean)</th>
<th>RMSE, ºC (max)</th>
<th>Bias, ºC (max)</th>
<th>RMSE, ºC (min)</th>
<th>Bias, ºC (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td>W2 Lake</td>
<td>1989-2008</td>
<td>1.23 (± 1.13)</td>
<td>1.20 (± 1.03)</td>
<td>5.03</td>
<td>4.54</td>
<td>0.17</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>1.24 (± 1.09)</td>
<td>1.22 (± 1.00)</td>
<td>5.02</td>
<td>4.57</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>Hosteller</td>
<td>1989-2008</td>
<td>2.04 (± 0.80)</td>
<td>1.93 (± 0.72)</td>
<td>5.07</td>
<td>4.39</td>
<td>0.83</td>
<td>1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>2.08 (± 0.78)</td>
<td>1.98 (± 0.71)</td>
<td>4.97</td>
<td>4.35</td>
<td>0.83</td>
<td>1.17</td>
</tr>
<tr>
<td>FLake</td>
<td>1989-2008</td>
<td>1.85 (± 0.58)</td>
<td>2.31 (± 0.73)</td>
<td>3.80</td>
<td>3.93</td>
<td>0.93</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>1.83 (± 0.61)</td>
<td>2.19 (± 0.81)</td>
<td>3.78</td>
<td>3.91</td>
<td>0.92</td>
<td>0.83</td>
</tr>
<tr>
<td>ANN</td>
<td>1989-2008</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>0.86 (± 0.31)</td>
<td>0.71 (± 0.21)</td>
<td>1.61</td>
<td>1.25</td>
<td>0.52</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
</tbody>
</table>

Wet season | Dry season

<table>
<thead>
<tr>
<th>Model</th>
<th>Period</th>
<th>RMSE, ºC (mean)</th>
<th>Bias, ºC (mean)</th>
<th>RMSE, ºC (max)</th>
<th>Bias, ºC (max)</th>
<th>RMSE, ºC (min)</th>
<th>Bias, ºC (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td>W2 Lake</td>
<td>1989-2008</td>
<td>1.21 (± 0.86)</td>
<td>1.20 (± 0.86)</td>
<td>0.19</td>
<td>0.34</td>
<td>1.18</td>
<td>1.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>1.23 (± 0.82)</td>
<td>1.22 (± 0.83)</td>
<td>0.32</td>
<td>0.34</td>
<td>1.18</td>
<td>1.16</td>
</tr>
<tr>
<td>Hosteller</td>
<td>1989-2008</td>
<td>1.92 (± 0.67)</td>
<td>1.77 (± 0.57)</td>
<td>-1.16</td>
<td>0.19</td>
<td>2.05</td>
<td>1.99</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>1.97 (± 0.69)</td>
<td>1.85 (± 0.63)</td>
<td>-1.03</td>
<td>0.25</td>
<td>2.08</td>
<td>2.02</td>
</tr>
<tr>
<td>FLake</td>
<td>1989-2008</td>
<td>1.46 (± 0.62)</td>
<td>2.00 (± 0.84)</td>
<td>-0.23</td>
<td>1.40</td>
<td>2.09</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>1.47 (± 0.60)</td>
<td>1.95 (± 0.81)</td>
<td>-0.23</td>
<td>1.38</td>
<td>1.98</td>
<td>2.32</td>
</tr>
<tr>
<td>ANN</td>
<td>1989-2008</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
<tr>
<td></td>
<td>2005-2008</td>
<td>0.81 (± 0.28)</td>
<td>0.69 (± 0.23)</td>
<td>-0.11</td>
<td>-0.11</td>
<td>0.88</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
<td>Daily</td>
<td>Hourly</td>
</tr>
</tbody>
</table>

https://doi.org/10.5194/gmd-2021-64
Preprint. Discussion started: 2 June 2021
© Author(s) 2021. CC BY 4.0 License.
Figure 5. Evaluation of simulation bias and RMSE. 2-D baseline scenario (W2 Reservoir) simulated SWT for: a) the exclusion of inflows and outflows (W2 Lake); b) HLM; c) FLake; and d) ANN. Models forced with hourly meteorology for the 24 waterbodies (2005-2008). The ensemble graphic, e), shows a box-plot of RMSE values (maximum, 75th percentile, median, 25th percentile, and minimum) considering the 1-D models results (2005-2008). In f) the wind sheltering coefficient considered during the calibration of the W2 Reservoir scenario is presented.
Overall, the statistical comparison by Taylor diagrams (Fig. 6) suggests that FLake had a slightly better performance than HLM in simulating SWT. Noteworthy, the standard deviation of the simulations forced with hourly meteorology was consistently closer to the standard deviation of the baseline scenario (W2 Reservoir) (Figs. 6c and d), showing the importance of meteorological data temporal resolution. ANN results were closer to the baseline scenario than the 2-D model (W2 Lake) regardless of the meteorological data discretization.

Figure 6. Taylor diagrams showing standard deviation (ºC), RMSEc (ºC), and correlation of SWT for the baseline scenario (W2 Reservoir) and for each model (a) 1989 – 2008 (daily met.); b) 2005 – 2008 (daily met.); c) 1989 – 2008 (hourly met.); and d) 2005 – 2008 (hourly met.). Statistics are calculated over all 24 reservoirs/lakes for the 1989-2008 period, and over 22 reservoirs/lakes for the 2005-2008 period (Alqueva, R24, and Pedrógão, R14, reservoirs were not modeled with the ANN).

4.2.2 Modeling computation time

The analysis of computation times was conducted through the comparison of the mean CPU time per time-step in the case of the 1-D models, with the mean CPU time per prediction sample obtained with the ANN, across all reservoirs. The results (Tables 6) show, considering the hourly simulations that the prediction phase of the ANN is approximately 26 times faster than FLake, the fastest process-based 1-D model optimized for coupling with climate models. The Hostetler model code written
in Python is approximately 45 times slower than FLake, nevertheless it is important to mention that a Fortran implementation of Hostetler can be much faster as described by Thiery et al., (2016). In their work the Hostetler model was only 3.6 times slower than FLake during the modeling of a deep lake (60 m deep).

It is important to mention that the performance of the models depends on the software implementation, therefore, the computation time values can vary significantly from the ones presented in this research.

Table 6. Computation time for 1-D models and for the ANNs prediction phase

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of layers</th>
<th>Number of time-steps</th>
<th>CPU time (s)/time-step</th>
<th>Total CPU time (s)</th>
<th>Programming language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostetler</td>
<td>30.0 ± 14.1</td>
<td>6 209 ± 1 786</td>
<td>0.8 ± 1.7 x 10⁻²</td>
<td>9.5 ± 5.0</td>
<td>Python 3.7 (Numpy 1.19.1)</td>
</tr>
<tr>
<td>Flake</td>
<td>2</td>
<td>6 209 ± 1 786</td>
<td>1.8 x 10⁻² ± 0.2 x 10⁻²</td>
<td>0.2 ± 3.7 x 10⁻²</td>
<td>Fortran 77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of training samples</th>
<th>Number of Predictions</th>
<th>CPU time (s)/number of prediction samples</th>
<th>Total CPU time (s)</th>
<th>Programming language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostetler</td>
<td>3 577 ± 1185</td>
<td>1 192 ± 395</td>
<td>0.7 x 10⁻³</td>
<td>6.9 x 10⁻³ ± 8.0 x 10⁻³</td>
<td>C++ and Python 3.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of layers</th>
<th>Number of time-steps</th>
<th>CPU time (s)/time-step</th>
<th>Total CPU time (s)</th>
<th>Programming language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostetler</td>
<td>30.0 ± 14.1</td>
<td>149 016 ± 42 866</td>
<td>0.8 ± 1.7 x 10⁻²</td>
<td>163.0 ± 89.2</td>
<td>Python 3.7 (Numpy 1.19.1)</td>
</tr>
<tr>
<td>Flake</td>
<td>2</td>
<td>149 016 ± 42 866</td>
<td>1.9 x 10⁻² ± 0.2 x 10⁻²</td>
<td>3.0 ± 0.9</td>
<td>Fortran 77</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Number of training samples</th>
<th>Number of Predictions</th>
<th>CPU time (s)/number of prediction samples</th>
<th>Total CPU time (s)</th>
<th>Programming language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hostetler</td>
<td>117 924 ± 36 785</td>
<td>29 481 ± 9 196</td>
<td>0.7 x 10⁻³</td>
<td>1.8 x 10⁻² ± 3.2 x 10⁻³</td>
<td>C++ and Python 3.7</td>
</tr>
</tbody>
</table>

4.3 The influence of reservoir inflows and level variations on SWT predictions

Additionally, to fully evaluate the influence of the inflows and level variations on SWT predictions in the reservoirs, and as a result, on surface latent and sensible heat fluxes, we considered the mean annual SWT results obtained with all models for six reservoirs with SWT most sensitive to the exclusion of inflows/outflows. The reservoirs were chosen based on the six highest maximum RMSE values obtained between 2-D baseline scenario (W2 Reservoir) SWT time series and SWT time series simulated with the exclusion of inflows and outflows (W2 Lake) (Fig. 7). Mean annual wind velocity, surface latent and sensible heat fluxes in these reservoirs are presented in (Figs. 8, 9 and 10), respectively. The results for the small Penide reservoir (R3), with a maximum depth of 9 m and an average volume of 0.11 hm³, while revealing large errors in all model runs, also show that these errors were significantly improved by the ANN. The HLM overperformed the FLake model in four of the six waterbodies (R1, R16, R5 and R22). Nevertheless, both 1-D models had an overall comparable performance. The ANN reduced significantly the annual maximum RMSE obtained for all reservoirs with all the models (Fig. 7). The aggregated analysis of results presented in Figs. 8, 9 and 10, allows estimating the combined effect of the wind forcing and the influence of inflows and level variations in the surface heat fluxes. Separation of the wind effects from the mass budget...
variability is possible because the differences between W2 Reservoir and W2 Lake scenarios describe only the combined influence of inflows and level variations on SWT, whereas the results obtained with the 1-D models describe the joint influence of the wind forcing and the influence of inflows and level variations (Figs. 9 and 10). The results obtained for reservoirs R1, R3, R5 and R10 show an appreciable effect on the surface heat fluxes caused by the neglect of inflows. As expected, the mean annual surface heat fluxes increased and decreased during the dry and wet seasons, respectively (Figs. 9 and 10). However, results obtained with the 1-D models, reveal a strong effect of the wind forcing across all reservoirs except reservoir R16. The differences in surface heat fluxes were as expected less pronounced in reservoir R16, due to the smaller difference between the wind forcing of the models (15%) (Fig. 8). Generally, the 1-D models overestimated the latent heat fluxes, in particular HLM, because FLake model results demonstrated a significant underestimation of SWT for reservoirs R1, R5, R10 and R22 as described by the corresponding maximum RMSE (Fig. 7). Accordingly, the mean annual sensible heat fluxes had a larger daily variability due to the need to balance the differences between air and water temperatures reaching, 21.09; SD ± 4.12 W/m² (Figure 10). The ANN reduced significantly the annual bias obtained for the surface heat fluxes for all reservoirs with all the models (Figs. 9 and 10). The only exception were the results obtained for R16, a run-of-the-river hydropower scheme, for which the 2-D modeling results were strongly affected by computation instability due to large inflow values. The training of the ANN partially reflected this instability into the final ANN structure causing a small overestimation of the surface heat fluxes during the dry season (Figs. 9 and 10).
Figure 7. Mean annual SWT values obtained with W2 Reservoir scenarios (W2R), W2 Lake scenarios (W2L), HLM, FLake and ANN considering hourly meteorology (2005-2008). Annual Maximum RMSE between W2 Reservoir (W2R) and the other models SWT results ( Graphics are ordered from the highest to the lowest RMSE values obtained for W2 Lake scenarios (W2L)).
Figure 8. Mean annual wind velocity values obtained with W2 Reservoir scenarios (W2R), W2 Lake scenarios (W2L), HLM, FLake and ANN considering hourly meteorology (2005-2008). Bias between W2 Reservoir (W2R) and the other models SWT results.
Figure 9. Mean annual latent heat values obtained with W2 Reservoir scenarios (W2R), W2 Lake scenarios (W2L), HLM and FLake considering hourly meteorology (2005-2008). Bias between W2 Reservoir (W2R) and the other models SWT results.
Figure 10. Mean annual sensible heat values obtained with W2 Reservoir scenarios (W2R), W2 Lake scenarios (W2L), HLM and FLake, considering hourly meteorology (2005-2008). Bias between W2 Reservoir (W2R) and the other models SWT results describe quite well the combined influence of inflows and level variations in SWT evolution, which can be parametrized using the WRT. The results obtained for both scenarios reveal a significant logarithmic correlation (Eq. 10) between the RMSE of SWT from the two scenarios and WRT (Figure 11a).
\[ \text{RMSE} = -0.36 \ln(WRT) + 2.73, \quad R^2 = 0.88; \text{MAE} = 0.27 \degree C \]  

with RMSE and WRT expressed in °C and in days, respectively.

The results additionally show that the computed SWT values in reservoirs with a residence time shorter than 100 days may have large errors if simulated without inflows/outflows (Figure 11).

![Figure 11. RMSE as a function of WRT. Between the 2-D baseline scenario (W2 Reservoir) and simulated SWT with: a) the exclusion of inflows and outflows (W2 Lake, 1989-2008); b) HLM (1989-2008); c) FLake model (1989-2008); and d) ANN (2005-2008), driven with hourly meteorology.](image-url)

5 Discussion and conclusions

The thermodynamics of natural and artificial lakes are similar. Nevertheless, the evolution of SWT in lakes and reservoirs differs substantially as a result of heat advection by inflows, outflows and, to a lesser extent, due to water level variations. Evaluation of differences between thermal regimes of lakes and reservoirs from observational data is limited by the availability of comparable waterbodies. The model-based approach used in the present study provides an effective alternative, complementary to the studies evaluating the thermal structure differences between lakes and reservoirs across the latitudinal gradient (e.g. Doubek and Carey, 2017; Hayes et al., 2017). We show that, for the same morphometry and climatic conditions, the SWT in reservoirs (approximately 46 %) is higher than the SWT in lakes. The results also suggest that the SWT predictions can be significantly affected by the water surface level variations. Nevertheless, in the present study only the combined effect of advection/level variation was evaluated, and the individual effect of level variations was not correlated with the SWT simulation errors. Therefore, the partial contribution of this variable to SWT was not fully explored and requires a future in-depth analysis.

One of the main novel aspects of the study lies in the fact that computationally efficient models (1-D and ANN) are compared against a baseline target instead of among themselves. Additionally, the study relies on the analysis of a large number of waterbodies and simulations conducted over a several decades long period. The methodological approach exposed the strengths and weaknesses associated with the simulation of the SWT of reservoirs by both process-based physical and data-driven
models. We demonstrated that inflows and outflows have a relevant effect on the evolution of SWT, with broader implications in the quality of GCMs and RCMs used in numerical weather prediction and climate modeling. It was also shown that there are other factors besides inflows and outflows that affect SWT. Examples are the wind forcing, the temporal sampling of the meteorological forcing data and the simplification of processes for quantifying turbulent energy flows. The low computational costs of 1-D process-based models in particular of the FLake model is the decisive factor for their integration in numerous GCMs and RCMs. Indeed, 1-D models such as FLake and HLM present a particularly good alternative to model reservoirs with missing field data and external parameters. Overall, Hostetler and FLake models demonstrated a reasonable performance, the latter being slightly better in modeling SWTs. Nevertheless, the results highlight that their SWT predictions can diverge significantly from observed values unless advective heat transport by in- and outflows and water level variations are integrated in the models. As an alternative to process-based models, an improvement can be achieved both in accuracy and computational requirements by using data-driven models. The ANN approach demonstrated a remarkably good performance by reducing the average value of RMSE of hourly simulations by at least 64% and running 26 times faster than FLake model. Nevertheless, there are two important limitations to the implementation of ANNs in GCM or RCM contexts. The first is the need for sufficient amount of accurate observational data to train the model; the second is the availability of river inflow temperatures. Both are still scarce, but their availability is rapidly increasing due to recent developments in remote sensing.

The present results suggest that reservoirs with a WRT shorter than 100 days, if simulated without representation of inflows and outflows, tend to exhibit an important deviation in the computed SWT values regardless of their morphological characteristics. Neglecting inflows and outflows while modeling these waterbodies may cause an overestimation of the turbulent energy fluxes, which can produce spurious local instabilities if surface water temperatures are higher than average mean air temperatures.

Incorporation of inflows and outflows in 1-D models for regional and global climate simulations will decrease computation efficiency and add an additional layer of uncertainty in the modeling of systems whose real nature is three-dimensional. The data-driven model considered in this study outperformed process-based physical models in computation time and in accuracy, being capable of accounting for the influence of inflows and outflows. In the context of waterbody simulation within numerical weather prediction and climate models, the use of data-driven approaches to complement their process-based counterparts may be highly efficient when data necessary to train the models is available. Given the growing capabilities and increasingly common use of remote sensing data acquisition techniques, the possibility of improving the performance of GCMs and RCMs through the enhanced modeling of waterbody-atmosphere turbulent heat exchanges is promising.

**Code availability**

The exact version of the models source code is archived on Zenodo at [http://doi.org/10.5281/zenodo.4803480](http://doi.org/10.5281/zenodo.4803480) (Almeida, 2021a). The current version of the open-source CE-QUAL-W2 model (version 3.6) used in this study, is also available from the project website ([http://www.ce.pdx.edu/w2/](http://www.ce.pdx.edu/w2/)). FLake (version 1.0) is freely available under the terms of the GNU Lesser...
General Public License (http://www.gnu.org/licenses/lgpl.html). The model source code, a windows executable, as well as a comprehensive model description are freely available from the official FLake website (http://www.lakemodel.net). For completeness, the windows pre-compiled version of FLake as used in the present calculations is also archived on Zenodo (Almeida, 2021a). The open-source Hostetler model source code is also available from the repository. The Python library used to construct the ANN, NeuPy version 0.8.2, is available from the NeuPy website (http://neupy.com/pages/home.html) under the terms of the MIT License and the ANN source code and scripts used to train the model are archived on Zenodo (Almeida, 2021a).

Data availability

Input files needed to run the models and the hydrometric, water quality and meteorological data sets used to force and validate each model, are freely available and are archived on Zenodo at http://doi.org/10.5281/zenodo.4756312 (Almeida, 2021b).

Author contribution

MA conceived the study, produced the code of Hostetler model and performed the simulations. YS developed and optimized the ANN, while GK provided support to FLake model simulations. PS and RC provided the meteorological data sets. JP supported code development. PC, AC and RR contributed to the study design and to the results analysis. All authors contributed to the discussion and manuscript revision.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

The authors thank the Portuguese Environmental Agency for providing the hydrometric and water quality data sets that were used in this study. Sincere thanks are given to one anonymous reviewer for the invaluable and constructive comments and suggestions for improving the paper quality.

References


