Environmental Modelling and Software 139 (2021) 105017

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Multi-scale evaluation of a 3D lake model forced by an atmospheric model against standard monitoring data

Marina Amadori [a,b,*], Lorenzo Giovannini [a], Marco Toffolon [a], Sebastiano Piccolroaz [c,d], Dino Zardi [a], Mariano Bresciani [b], Claudia Giardino [b], Giulia Luciani [b], Michael Kliphuis [c], Hans van Haren [e], Henk A. Dijkstra [c]

[a] Department of Civil, Environmental and Mechanical Engineering (DICAM), University of Trento, Italy
[b] Institute for Electromagnetic Sensing of the Environment, National Research Council, Milan, Italy
[c] Institute for Marine and Atmospheric research Utrecht (IMAU), Department of Physics, Utrecht University, the Netherlands
[d] Physics of Aquatic Systems Laboratory, School of Architecture, Civil and Environmental Engineering, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland
[e] Royal Netherlands Institute for Sea Research, the Netherlands

ARTICLE INFO

Keywords:
Numerical simulation
WRF
Delft3D
Water temperature
Wind
Remote sensing

ABSTRACT

Evaluating a three-dimensional lake model requires large datasets of many variables, including velocity fields, that are seldom available. Here we discuss how to assess the performance of a model at multiple scales (in time and space) with data from standard monitoring systems, i.e., mostly limited to water temperature. The modeling chain consists of a lake hydrodynamic model (Delft3D-Flow) forced by an atmospheric model (WRF, Weather Research and Forecasting). The two models are tested on the case study of Lake Garda (Italy), where a comprehensive dataset of atmospheric and water temperature observations is available. Results show that a consistent picture of the inherent dynamics can be reproduced from a heterogeneous set of water temperature data, by distilling information across diverse spatial and temporal scales. The choice of the performance metrics and their limitations are discussed, with a focus on the procedures adopted to manage and homogenize data, visualize results and identify sources of error.

1. Introduction

When modeling thermo-hydrodynamic processes in lakes, two main factors guide the choice of the appropriate numerical model and of the calibration/validation procedure: one is the process to be investigated, and the other is the available data. Once the model has been chosen and the data gathered, the model is set up, calibrated and validated. It is usually assumed that, when the model reproduces reasonably well the observed data, it can be accepted as reliable, and used to address specific research questions. However, in principle, an exact and complete model verification can never be fully achieved (Oreskes et al., 1994), since the verification of a model implies the demonstration of its truth, which is only possible for closed problems. For the case of lake modeling, as of any natural system, the problem to be solved is inherently not closed and the numerical solution is always non-unique. Available measurements are often limited to water temperature data, and current velocity measurements are missing. A given water temperature profile can hence be compatible with different velocity fields, and one-dimensional (1D) models have been successfully used to reproduce lake thermal dynamics. In a three-dimensional (3D) model, water temperature observations need to be supplemented with velocity measurements at several locations and with sufficient temporal resolution, to guarantee that both temperature and flow fields are consistently reproduced. In most real cases, however, such comprehensive datasets are not available. Thus, a viable option is using the largest number of (incomplete) data from standard monitoring systems (often not specifically designed for modeling applications), and carefully defining the modeling goals, so that the available observations are adequate for the specific scientific question.

Usually, modeling studies approach environmental systems investigating their behaviour on specific time scales. For instance, lake modeling applications may be divided into three categories: (i) long-term simulations and projections of future states under climate change; (ii) seasonal to daily dynamics of circulation and transport in

* Corresponding author. Department of Civil, Environmental and Mechanical Engineering (DICAM), University of Trento, Via Mesiano 77, Trento, Italy.
E-mail address: marina.amadori@unitn.it (M. Amadori).

https://doi.org/10.1016/j.envsoft.2021.105017
Accepted 22 February 2021
Available online 5 March 2021

the whole lake system; (iii) short-term, non-stationary processes, such as internal waves, involving multiple scales from hours to several days.

In case (i), large-scale temperature variations and long-term trends of water temperature (warming trends, most frequently) are investigated. Due to the substantial computational costs of long-term simulations and to the relatively low resolution, both in time and space, of climatic model output, such projections are commonly addressed by means of 1D models forced by climate change scenarios (e.g. Fang and Stefan, 2009; Perroud and Goyette, 2010; Kirillin, 2010; Piccolroaz and Toffolon, 2013, 2018; Butcher et al., 2015; Valerio et al., 2015; Schwefel et al., 2016; Schlabing et al., 2014, among others). Instead, 3D model simulations in this field are still rare (Beletsky et al., 2006; Yamashiki et al., 2010; Wahl and Peeters, 2014; Xue et al., 2017), and are normally evaluated on a monthly basis.

At interannual to seasonal, or even daily time scales, as in case (ii) above, the range of the relevant physical processes to be resolved widens. This is the case of wind-driven circulations and up/downwelling effects, surface gyres and riverine intrusions, and their interplay with the spatial heterogeneity of the atmospheric forcing and topographic constraints. In the literature, several recent examples are available concerning 3D short-term modeling studies of lake circulation and heat transport (e.g. Razmi et al., 2013; Bouffard et al., 2018; Piccolroaz et al., 2019). In such a variety of investigations, even methods for model evaluation may be widely diverse. In general, a multi-site validation is performed, but the time resolution may range from monthly to hourly, or even sub-hourly time series (Pilotti et al., 2013; Råman Vinnå et al., 2017).

When approaching daily and shorter time scales (case (iii) above), lake temperature is directly affected by surface air-water heat and momentum exchanges. Also internal waves may be excited, which in turn modify the basin-scale circulation and promote temperature fluctuations. Internal waves dynamics require high-resolution observations as temperature fluctuations may span periods ranging from several days to few minutes. Most of internal wave studies in lakes (e.g. Hodges et al., 2000; Pan et al., 2002; Rueda and Schladow, 2003; Laval et al., 2003; Pilotti et al., 2013; Soulignac et al., 2017; Valerio et al., 2017; Dissanayake et al., 2019) evaluated the models with hourly or sub-hourly temperature data sampled by moored thermistors. Frequently, the model performance in these applications is assessed with sophisticated tools, such as signal decomposition on basis functions (e.g., Fourier series and wavelet functions).

Most of the above studies aimed at investigating targeted processes, for which the model was calibrated and validated. Few of them from comprehensive datasets, including water velocity measurements, whereas most of them relied on water temperature data only. In the present paper, we face the challenge of evaluating the performance of a 3D lake model at all the time scales discussed above, by relying only on water temperature observations. The underlying idea is that coherent lake dynamics can be satisfactorily reproduced by distilling information from water temperature data at different spatial and temporal scales.

A multi-scale evaluation also allows interpreting the capabilities of the model. For instance, to understand weaknesses and strengths of a lake model, specific attention has to be paid to the atmospheric conditions forcing it. Indeed, weather forcing plays a key role in the development of the lake thermal structure, as it directly regulates heat fluxes, lake mixing and circulation. In many lake models, atmospheric forcing variables are assumed as spatially uniform, or are obtained from interpolating observations at a few points (e.g. Strub and Powell, 1986; Laval et al., 2005, who both tested different wind fields and stressed the importance of temporal and spatial variability). More recently, spatially resolved atmospheric forcing from atmospheric models have been used (e.g. Pan et al., 2002; Valerio et al., 2015; Amadori et al., 2018), including output from numerical weather prediction models (e.g. Kelley et al., 1998; Wahl and Peeters, 2014; Baracchini et al., 2020b), and also suitable atmosphere-lake coupling procedures have been successfully tested (e.g. Song et al., 2004; Leon et al., 2007; Xue et al., 2017).

In the present paper, we critically discuss the steps of setup, calibration and performance assessment of a modeling chain composed of a hydrodynamic (Delft3D-Flow) model forced by an atmospheric model (WRF). The aim of our investigation is twofold: on one hand, we aim at demonstrating that the evaluation of a 3D model performance can be based on water temperature at multiple spatial and temporal scales. On the other hand, we show how the performance of the externally simulated atmospheric forcing affects the results of the hydrodynamic model, again at multiple scales. To achieve these goals, we adopt two models, that are well known and widely used within the atmospheric and limnological communities, respectively, and suggest a practical approach for the evaluation of their performance by using data from standard monitoring schemes.

The case study is Lake Garda in Italy, for which an operational model has not been developed. Indeed, we propose a paradigm to assess the performance of candidate lake models to be possibly adopted as operational standards (e.g. Wang et al., 2019; Baracchini et al., 2020b).

The paper is organized as follows: once the case study is introduced (Sect. 2) and the available observations are described (Sect. 3), we present the modeling chain (Sect. 4), together with the quality measurements used for testing the models performance (Sect. 4.3) and the calibration of the hydrodynamic model (Sect. 4.5). In the results section (Sect. 5), we first present an overview of performance of the two models against the available dataset (Sect. 5.1) of the lake model, then we specifically focus on different time scales moving from interannual (Sect. 5.2) to seasonal (Sect. 5.3) to sub-daily time scales (Sect. 5.4). Next, we test the spatial variability patterns of the temperature field at the lake surface (Sect. 5.5). Finally, we discuss our results (Sect. 6) and provide final conclusions (Sect. 7).

2. Case study

Lake Garda (see Fig. 1) is a large lake at the feet of the Alps in northern Italy, displaying a surface area of 368 km^2 and a maximum depth of 346 m. The altitude of the water surface is 65 m above sea level, and its shape combines a long, narrow and deep trunk (length 30 km, width 3 km, maximum depth 346 m) in the northern part with a wider shallow sub-basin (maximum length 13 km, width 20 km and depth 80 m) in the southern part.

Over the last decades, Lake Garda has been deeply investigated and several data have been collected, although the protocols and monitoring actions adopted were not optimized for hydrodynamic studies. Investigations based on remote sensing have been conducted since the '90s (Zilioli et al., 1994), using data gathered from multi-spectral Landsat and Sentinel-2 sensors (e.g. Bresciani et al., 2018), ocean color radiometers as MERIS and MODIS (e.g. Bresciani et al., 2011a), and imaging spectrometry from both satellite (Giardino et al., 2007) and airborne platforms (Bresciani et al., 2012). Lake Garda was extensively used as a case study for estimating indicators of water quality, such as chlorophyll-a and lake surface temperature (Bresciani et al., 2011b; Lenstra et al., 2014; Nicolantonio et al., 2015). Moreover, remote sensing techniques were used to map shallow water properties in the lake (Giardino et al., 2014), and algorithms for atmospheric correction of the signal were developed and tested (e.g. Guanter et al., 2010; Odermatt et al., 2010). As a LTER[1] site, Lake Garda was the subject of long-term studies investigating the ecological state (see Salmaso et al., 2017). Moreover, the lake was adopted as a case study for an experiment on homogenization of daily lake surface water temperature (LSWT) by merging data from 13 satellites from 1986 to 2015 (Pareeth et al., 2016).

The typical meteorological processes of the region surrounding the lake have also been well investigated. In particular, the development of peculiar local wind circulations, induced by the thermal contrasts between the lake and the surrounding orography, have been documented

[1] https://lternet.edu/.

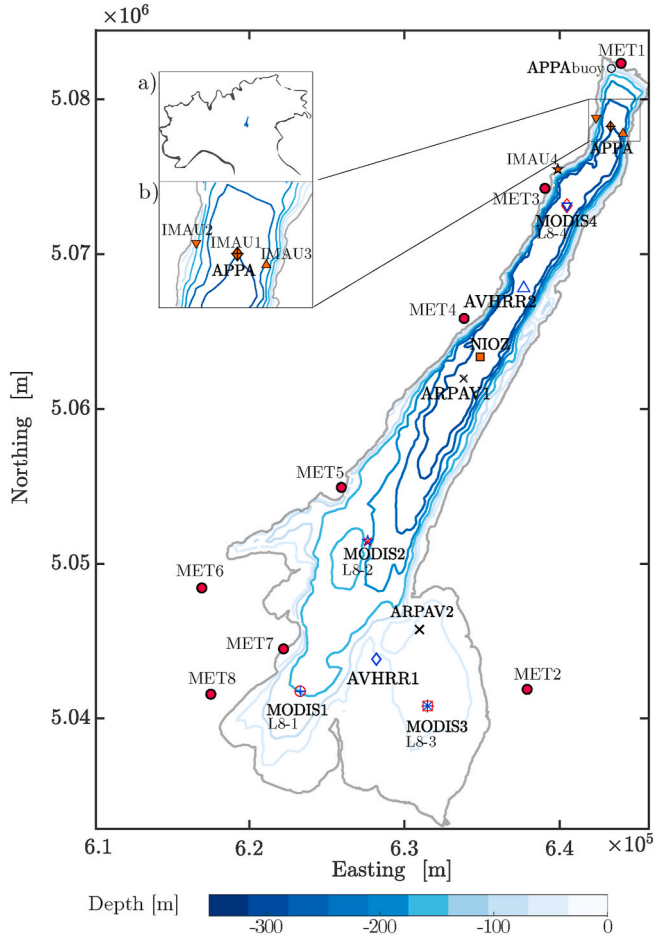

Fig. 1. Map of Lake Garda's monitoring stations and bathymetry, Easting and Northing refer to the UTM coordinate system. Insets: a) location of Lake Garda in Northern Italy; b) zoom on IMAU stations along a cross-section having the APPA point as mid-point. Note: the APPA-IMAU1 overlap. See Table 1 for more detail on the station coding.

by means of observations and numerical modeling studies (Laiti et al., 2013a, b, 2014; Giovannini et al., 2015). Moreover, a Wind Atlas for Trentino, including the area of Lake Garda, was produced on the basis of 10-year long atmospheric simulations (2004–2013) with the Weather Research and Forecasting (WRF) model (Giovannini et al., 2014a).

A series of preliminary experiments using three-dimensional hydrodynamic numerical models were also undertaken in the past (Casulli and Pecenik, 1994; Lovato and Pecenik, 2012), although limited to rather idealized conditions, such as constant and uniform wind field and/or constant eddy viscosities. More recently, numerical experiments (Amadori et al., 2018) have been performed by forcing the Delft3D model (Lesser et al., 2004) with wind fields resulting from WRF model (Skamarock et al., 2008). These preliminary modeling experiments allowed a detailed description of the typical transport processes and circulations occurring in the lake under different seasonal conditions (i.e., different thermal stratification and typical wind field). However, these modeling efforts were hampered by the lack of *in situ* observations for model calibration and validation. A combined observation-modeling approach was adopted to reproduce episodic wind events (Piccolroaz et al., 2019) The authors showed that strong and long-lasting northerly winds can promote lateral flows influenced by the Coriolis force, and hence significantly affect the deep ventilation of the lake. Recently, the same hydrodynamic model setup was also used in two studies tracing the surface transport dynamics: Amadori et al. (2020) used the modeling results to test lake users knowledge on surface currents, and Ghirardi et al. (2020) integrated modeling with remote sensing to quantify a turbidity discharge event in Lake Garda.

In the present study, the same modeling approach previously adopted by Amadori et al. (2018, 2020), Piccolroaz et al. (2019) and Ghirardi et al. (2020), consisting in Delft3D-Flow simulations forced by WRF, is extensively tested against a comprehensive dataset of water temperature observations and weather stations measurements. An overview of the collected observations is provided in the following section.

3. Observations

For the purpose of the present work, water temperature data are gathered by merging: i) historical long-term series of data from *in situ* observations of the water column at different locations, ii) multi-site high-resolution profiles measured between 2017 and 2018 during a dedicated monitoring program, and iii) remote sensing data from various satellites. The dataset covers the period from 1990 to present, and merges data acquired with different techniques, under different projects, and by different institutions. At the same time, data from several weather stations are collected and compared with the results of the atmospheric model. Table 1 provides a summary of all the data used in this study, while the locations of the observation points are shown in Fig. 1. In particular, Table 1 provides details of the monitoring activities on the lake, including the time coverage of data, the data provider, the location of the stations, the sampling frequency, the relevant metadata, and the number of available measurements. The criterion adopted for the station codes (as reported in Fig. 1) is the following: water temperature monitoring stations are named according to the operating institution (e.g., APPA, ARPAV1 and ARPAV2, IMAU and so on); remote sensing control points are named according to either the sensor (MODIS, AVHRR) or the satellite (LANDSAT8). In order to avoid confusion between water temperature stations and meteorological stations, the latter are numbered after the prefix MET according to the order they are listed in Table 1. In the following paragraphs, each dataset is described in detail.

3.1. EPAs routine in situ measurements

The current database of *in situ* observations provided by the Environmental Protection Agencies (EPAs) of the Autonomous Province of Trento (APPA) and of the Veneto Region (ARPAV) represents the most complete time series of water temperature available for Lake Garda. Water temperature profiles are routinely collected on a monthly/bi-monthly basis. Water temperature profiles are measured through a shipborne Conductivity Temperature Depth (CTD) profiler. Data from APPA are available from 1990 up to the present. Measurements are taken in the 'APPA' monitoring point, located about 4 km south of northern shore of the lake (Fig. 1). In addition, APPA performs continuous temperature measurements, recorded every hour, at 10 m below the water surface, through a buoy located close to the north-western edge of the lake. These data were also used in the present study and will be hereafter referred to as 'APPA buoy'. ARPAV water temperature profiles are taken at several locations, but the longest series (2000 to present) are available at two points, namely Brenzone (the deepest point of the lake), indicated in the map as 'ARPAV1', and Bardolino, in the more shallow and wider sub-basin, indicated as 'ARPAV2'.

3.2. The 2017–2018 field campaign

High-resolution water temperature profiles were collected from 2017 to 2018 during a joint field campaign of the Institute for Marine and Atmospheric research Utrecht (IMAU, Utrecht University) and of the University of Trento (Toffolon et al., 2017; Piccolroaz et al., 2019). Data were acquired using a turbulence microprofiler (MicroCTD, distributed by Rockland Scientific International, Canada) operated at depths down to 100 m at four stations (see Fig. 1). IMAU1 point coincides with the

Table 1
Summary of all *in situ* and remotely sensed data used in this study.

In situ water temperature data								
Time availability	**Institution**	**ID**	**East**	**North**	**Frequency**	**Type of data**	**Depth**	**n**
1990–2018	EPA Province of Trento	APPA	643286	5078237	monthly	profiles	0–70 m	151
2008–2018		APPA buoy	643308	5082367	hourly	time series	10 m	78848
2000–2018	EPA Veneto Region	ARPAV1	634031	5062154	monthly	profiles	0–330 m	137
2000–2018		ARPAV2	630893	5045631		profiles	0–70 m	148
2017–2018	IMAU-UniTrento	IMAU1	643430	5078401	monthly	profiles	2–100 m	20
		IMAU2	642544	5078596				18
		IMAU3	644214	5077921				18
		IMAU4	640555	5075483				14
2017–2018	NIOZ	NIOZ	635637	5063948	5 s	profiles	187–337 m	5906

Satellite products									
Time availability	**Satellite**	**Sensor**	**ID**	**East**	**North**	**Frequency**	**Pixel size**	**Acquisition time**	**n**
1989–2013	NOAA	AVHRR	AVHRR1	628235	5044098	daily	1 km	see Riffler et al. (2015)	5117
			AVHRR1	637973	5067770				5117
2004–2010	Aqua	MODIS	MODIS1	623394	5042025	daily	1 km	12:45 UTC	1302
			MODIS2	626743	5051353				1300
			MODIS3	631143	5040947				1302
			MODIS4	640711	5072895				1079
2013–2018	Landsat-8	TIRS	L8-1	623394	5042025	bi-weekly	100 m	09:50 UTC	25
			L8-2	626743	5051353				29
			L8-3	631143	5040947				57
			L8-4	640711	5072895				53
			L8 full maps	1440 pixels					18

Meteorological data							
Time availability	**Institution**	**ID**	**East**	**North**	**Frequency**	**Altitude***	**Wind meas.****	**n**
2004–2018	Edmund Mach Foundation	MET1	643975	5082174	hourly	63 m asl	5 m agl	354415
2004–2018	EPA Veneto Region	MET2	637944	5041943	hourly	142 m asl	10 m agl	376358
2012–2018	EPA Lombardia Region	MET3	639088	5074143	hourly	74 m asl	16 m agl	162889
2013–2018		MET4	633895	5065781		420 m asl	10 m agl	127213
2012–2018		MET5	626016	5054923		67 m asl	21 m agl	161487
2013–2018		MET6	617051	5048460		284 m asl	10 m agl	129299
2017–2018		MET7	622300	5044545		66 m asl	6 m agl	18510
2016–2018		MET8	617629	5041631		182 m asl	10 m agl	46563

*m above sea level.
**m above ground level.

APPA monitoring station. IMAU2 and IMAU3 are located on opposite sides with respect to IMAU1, respectively close to the western and eastern shores. Together, these three stations define a cross-section perpendicular to the lake main axis, which is oriented in a northeast-southwest direction. The fourth station, IMAU4, is located further south, near Limone.

A total of 652 profiles are available from the period from March 3, 2017 to June 11, 2018. The sampling time is always between mid morning and early afternoon. The upper 2 meters of each profile have been removed in the post-processing step due to non-homogeneous profiling speed of the instrument.

3.3. Deep mooring

Concurrent with the IMAU-UniTrento field campaign, temperature data were measured by a taut-wire submerged mooring station including a single point current meter and a thermistor chain. The mooring was deployed by NIOZ (Royal Netherlands Institute for Sea Research) a few km north of the ARPAV1 monitoring point (see Fig. 1). The local depth at the point where the instrument was installed is 344 m, with a bottom slope of about 3.6° (1 km scale). Temperature measurements were taken between 187 and 337 m below the water surface by 100 self-contained high-resolution temperature sensors ('NIOZ4') deployed at 1.5 m vertical intervals, at a sampling rate of 0.5 Hz. Data were corrected for the electronic drift, and are available with a precision as small as 0.5 mK and a noise level of less than 0.1 mK (van Haren, 2018).

The mooring started to operate on May 24, 2017 (10 UTC) and was recovered on May 31, 2018 (09 UTC). However, in the present study, temperature data until March 26, 2018 (23 UTC) only are used, since after then half of the temperature sensors failed, due to insufficient batteries, and therefore interpolation is not viable. Hence, we use a reduced set of hourly sampled data.

3.4. Remote sensing

Lake Surface Water Temperature (LSWT) data have been collected from different satellite sensors. The longest time series is taken from the dataset produced by Riffler et al. (2015) for European Alpine lakes (1989–2013). This dataset is based on the images from AVHRR (Advanced Very High Resolution Radiometer) onboard various NOAA satellite platforms. Within this database,[2] LSWT is available on a daily basis as an average over a 3 × 3 pixel array (pixel size: 1 km) at two specific locations in the lake (we refer to Table 1 for the exact coordinates).

In addition, six-year long multi-temporal records from MODIS products (MYD11A, derived from MODIS sensors onboard Aqua) are used in four stations in Lake Garda. The MODIS sensor from the Aqua satellite acquires surface temperature at 9:45 UTC. Both AVHRR and MODIS sensors have a pixel size of 1 km. We combine the daily output from those sensors with the spatially more resolved Landsat-8 satellite imagery. Landsat-8 images provide fine scale mapping of LSWT at 100-m pixel size (TIRS sensor onboard), with a biweekly revisit time. In this

[2] Data can be downloaded from https://doi.org/10.1594/PANGAEA.830988 and https://doi.org/10.1594/PANGAEA.830988 for upper/lower Lake Garda respectively.

Table 2
Delft3D model setup and calibration parameters.

Model settings			
Simulation time step	30 s	Wall boundary conditions	*free slip*
Grid size	64 × 224 x 100	Heat fluxes model	*Ocean*
Horizontal grid spacing	100-400 m	Turbulence model	*k-ε*
Vertical grid spacing	1-25 m	Secchi depth	*monthly measured values*
Wind drag coefficient Cd*	4.40×10^{-3}	Bottom roughness (Chzy)	60 $m^{1/2}$ s^{-1}
	1.00×10^{-3}	Horizontal eddy viscosity	2×10^{-1} m^2s^{-1}
	2.00×10^{-3}	Horizontal eddy diffusivity	2×10^{-1} m^2s^{-1}
Calibration parameters	**Value**	**Tested range of variation**	
Stanton number	6.50×10^{-4}	1.77×10^{-4} — 2.00×10^{-3}	
Dalton number	1.30×10^{-3}	3.95×10^{-4} — 2.30×10^{-3}	
Free convection coefficient	1.00×10^{-1}	5.00×10^{-2} — 5.00×10^{-1}	
Background vertical eddy viscosity [m^2s^{-1}]	0	0 — 1.00×10^{-3}	
Background vertical eddy diffusivity [m^2s^{-1}]	0	0 — 1.00×10^{-3}	

* Breakpoints for the piecewise linear function of wind speed (1–5–10 m s^{-1}) at 10 m above lake surface.

work, Landsat-8 products from 2013 to 2018 are obtained by transforming TIRS radiances at satellite level in surface water temperature (i.e., skin temperature) according to Barsi et al. (2005). For a similar use of these maps we refer to Prats et al. (2018), who used Landsat to estimate lake skin surface temperature for more than 400 French water bodies. The full Landsat-8 maps are used when all pixels covering the lake surface are available (18 dates in 5 years). In addition, LSWT values at the same 1 km target areas of the MODIS Aqua products are extracted from the Landsat-8 maps, allowing a comparison of model output with data from different sensors.

3.5. Meteorological data

Several weather stations are spread all around Lake Garda, providing regular measurements of the main atmospheric variables. Ground weather stations are operated by various agencies, as shown in Table 1. In particular, the northernmost station (MET1) is operated by the Edmund Mach Foundation (FEM), the station on the east coast (MET2) belongs to the Enviromental Protection Agency of the Veneto Region (ARPAV), while the other ones (MET3 to MET8) to the Enviromental Protection Agency of the Lombardia Region (ARPAL). The height above the ground at which wind speed and direction are measured is not the same for all the stations (see Table 1). Air temperature and relative humidity are taken at 2 m above ground in all the weather stations.

4. Model setup

4.1. Atmospheric model

The atmospheric simulations used in this work are performed with the Weather Research and Forecasting model (Skamarock et al., 2008). WRF was successfully used in previous studies to reproduce thermally-driven winds in the Alpine valleys surrounding Lake Garda (Giovannini et al., 2014b). Part of the simulations (period 2004–2013) are derived from the Trentino Wind Atlas (Giovannini et al., 2014a). The remaining part (period 2014–2018) has been performed *ad hoc*, adopting the same modeling setup as for the above Atlas. The computational domain is composed of four nested domains with decreasing horizontal grid spacing of 54, 18, 6 and 2 km respectively, each having 28 vertical levels. Initial and boundary conditions are supplied by the 6-hourly National Center for Environmental Prediction (NCEP) Final Operational Global Analysis data on 1-degree grids. The simulations are initialized at 00 UTC of the last day of each month and finish at the end of the following month; the first day, which is mostly affected by the initialization, is not considered. The model output is written every hour. The Noah scheme (Chen and Dudhia, 2001) is used as land surface model, while the Yonsei State University scheme (Hong et al., 2006) is adopted as planetary boundary layer parameterization. The long-wave radiation is parameterized with the Rapid Radiative Transfer Model (Mlawer et al., 1997), while the Dudhia scheme (Dudhia, 1989) is used for the short-wave radiation, including the effects associated with slope inclination and topographic shading. The microphysics scheme adopted is the WRF single-moment 3-class simple ice scheme (Hong et al., 2004); the Kain-Fritsch cumulus scheme (Kain and Fritsch, 1993) is used in the three outer domains, while no cumulus parameterization is adopted in the inner domain. In the two inner domains we adopt the correction for wind speed over complex terrain proposed by Jiménez and Dudhia (2012), which considers the effects of unresolved topographic features. The original resolution of the Digital Elevation Model adopted in the inner domain is 30" (0.9 km). In the present work, we use the hourly model outputs only from the inner domain, covering Lake Garda.

4.2. Hydrodynamic model

Lake thermo-hydrodynamics are simulated using the open-source modeling suite Delft3D (Lesser et al., 2004) over the period 2004 to 2018, using the WRF output as forcing. Table 2 summarizes the main model settings.

The domain is discretized by a non-uniform, locally-orthogonal grid, staggered along the horizontal and vertical dimension, with 64 × 224 cells and 100 vertical layers. The horizontal grid spacing is of 200 m, while the thickness of the vertical layers increases from 1 m at the surface to 25 m at the bottom. The simulation time step is set to 30 s, as required for numerical stability. Initial conditions for Delft3D are set as water at rest, constant water level, and horizontally uniform water temperature profile, equal to the profile measured by the EPA of the Veneto Region in the ARPAV1 point in December 2003. Secchi depth is provided as a monthly mean value according to the measurement available at ARPAV1 point.

Vertical eddy diffusivity and viscosity are calculated with a *k*-ε turbulence model. In the parameterization implemented in Delft3D, a background value for both coefficients is required in order to guarantee a minimum value taking into account sub-grid eddies. The background turbulent value is treated as a calibration parameter, then the molecular viscosity and diffusivity is added to the turbulent part. In our simulations, the best performance is achieved by setting the background value of both viscosity and diffusivity to zero, such that the molecular value is used as a physically meaningful minimum and artificial mixing is avoided. Horizontal eddy viscosity and diffusivity are taken uniform and constant in the whole computational domain. Their value is essentially dependent on the spatial scale (Okubo, 1971), hence on the grid size. Preliminary tests showed that the simulated temperature is not sensitive to the choice of horizontal eddy viscosity and diffusivity: for this reason, the default value of 0.2 m^2s^{-1} is chosen. The bottom roughness

coefficient (Chezy formulation) is set to 60 $m^{1/2}s^{-1}$, as tests performed with different values did not show much differences. The wind drag coefficient is defined as a piecewise linear function of wind velocity (at 10 m above ground level). After testing different relations for this parameter (e.g., Wahl and Peeters, 2014), a clear picture of its effect did not emerge from the simulated temperature profiles, so breakpoints values are chosen following Wüest and Lorke (2003).

The surface boundary conditions (i.e., atmospheric forcing) are prescribed by the WRF model output every hour on a region covering the lake surface. The fields of wind velocity, air temperature, air pressure, relative humidity, shortwave incoming solar radiation and cloudiness are automatically interpolated by Delft3D on the lake model computational grid and time step. For the simulation of air-water heat exchanges, the *'Ocean'* model is used, which is implemented in Delft3D in the version originally formulated for the North Sea, based on Gill (1982), and then extended to large lakes. The model parameterizes the total heat flux through the lake surface by computing each term of the heat balance equation in each horizontal grid cell. In order to take into account the shadowing effect of the mountains, the net incident solar radiation (short wave) is directly prescribed by the atmospheric model. The remaining terms of the heat budget equation are computed by the *'Ocean'* model from the meteorological variables provided by WRF. These terms are the net incident atmospheric radiation (long wave), which takes into account the cloud cover, the long wave back radiation (from the water to the atmosphere), the evaporative heat flux and the convective heat flux. The latter two terms are split into a contribution by forced and free convection. The forced convection of latent and sensible heat is related to the wind forcing and is proportional to wind velocity, multiplied by a calibration constant (Dalton for evaporative flux, Stanton for convective flux). The free convection part includes the heat exchanges due to inverse temperature gradients between the 10 m air layer above the lake surface and the lake surface itself, when buoyancy fluxes develop in the atmospheric boundary layer due to unstable conditions (Gill, 1982). The relative importance of the free convection contribution to both latent and sensible heat flux can be regulated by tuning the free convection coefficient. In Table 2 the value chosen for the mentioned parameters is reported.

The outputs of the hydrodynamic simulations are recorded every day at 10 UTC in the whole computational domain. At observation points, coinciding with the monitoring stations, instantaneous numerical outputs are recorded every hour.

4.3. Assessment of the performance of the models

The performance of both atmospheric and hydrodynamic models is assessed by evaluating the metrics summarized in Table 3. We select those metrics that have been most extensively used in environmental modeling applications (see e.g., Hipsey et al., 2020).

Mean Absolute Error (*MAE*), Mean Error (*ME*, i.e., bias), Root Mean Square Deviation (*RMSD*), Standard Deviation (*SD*) and Correlation (*CORR*) are computed comparing simulated M_i and observed values O_i, where $i = 1, \ldots, N$ indicates time index and N the total number of observations. In Table 3, the bar symbol $\overline{M}(\overline{O})$ refers to the average over time of simulated (observed) quantities. An exception is made when comparing the water temperature from the hydrodynamic model and from Landsat-8 satellite maps: here i denotes a number of pixels, N the total number of pixels (as discussed later), $\overline{M}(\overline{O})$ the spatial mean over the lake surface.

In order to account for the site-specific variability on the evaluation of the performance metrics, the Standard Deviation of the observed data (SD_O) is used as normalization factor for computing the Normalized centered Root Mean Square Deviation *NRMSD* (according to Taylor, 2001), and the Normalized Standard Deviation of model results *NSD*. In addition, the Nash-Sutcliffe efficiency (*NSE*) is computed to test the robustness of the hydrodynamic model results only.

For both models, data from stations (point measurements) are compared with model's output (gridded fields). Such a comparison is unavoidably affected by the different meaning of point-wise vs spatially distributed data. In fact, local observations may not be representative for a large surrounding area, and the computational grid of the model output may be too coarse to capture the local conditions observed at the monitoring stations. In this regard, errors may be introduced also by the discretization of the computational domain, either along the horizontal or along the vertical direction. An example of this effect can be found in both models: in the atmospheric model, when areas with steep slopes are smoothed and underrepresented, or when specific microclimatic conditions at the transition between land and water are not appropriately captured; in the hydrodynamic model, when the vertical variation along the water column is not properly reproduced because of too thick layers. To overcome these issues, the metrics listed in Table 3 are computed by adopting some specific corrections, which will be presented in the following two subsections.

Table 3
Performance metrics.

Mean Absolute Error $MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert M_i - O_i\rvert$	Mean Error (bias) $ME = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)$
Root Mean Square Deviation $RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}$	Normalized centred Root Mean Square Deviation $NRMSD = \frac{\sqrt{\sum_{i=1}^{N}[(M_i - \overline{M}) - (O_i - \overline{O})]^2}}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}$
Standard Deviation $SD_O = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - \overline{O})^2}$ $SD_M = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - \overline{M})^2}$	Normalized Standard Deviation $NSD = \frac{\sqrt{\sum_{i=1}^{N}(M_i - \overline{M})^2}}{\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}$
Correlation $CORR_{MO} = \frac{\sum_{i=1}^{N}(M_i - \overline{M})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N}(M_i - \overline{M})^2}\sqrt{\sum_{i=1}^{N}(O_i - \overline{O})^2}}$	Nash Sutcliffe Efficiency (or R^2) $NSE = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$

4.3.1. Atmospheric model

The atmospheric model performance is evaluated by testing the simulated wind speed at 10 m above ground level, air temperature and relative humidity at 2 m above ground level, against measurements. In fact, these are the variables that mainly affect the results of the hydrodynamic model. For a fair evaluation of the atmospheric model performance, we correct observations from those stations located on the northwestern side of the lake, where a significant mismatch exists between the real and the computational orography. In these areas, the elevation of each model grid cell can be significantly higher than the elevation of the corresponding weather station, thus systematically affecting the air temperature, relative humidity and wind speed.

1. Air temperature. Air temperature is corrected considering the altitude difference between the model grid cell and the actual height of the weather station. We assume a constant lapse rate and use the mean standard tropospheric value $\Gamma = -dT/dz = 6.5\text{K km}^{-1}$.
2. Relative humidity. Relative humidity is corrected according to the air temperature above, without changes in the water vapour content simulated by the model.
3. Wind speed. For consistency between measurements and model results, wind observations taken at different heights are extrapolated to the standard height of 10 m above ground level, assuming a logarithmic profile for a neutrally stratified atmosphere. Deviations from the logarithmic profile are not taken into account, since data to estimate atmospheric stability are not available from the meteorological monitoring network. We use a roughness length of 0.001 m for MET1 and MET7, as these stations are located on a dock, and of 0.01 m for MET3 and MET5, located over grass (Oke, 1987).

4.3.2. Hydrodynamic model

When testing the performance of the hydrodynamic model, the structural inhomogeneity of observation data is an issue to be addressed. Since data are available in several formats (e.g., vertical profiles, pixels of different size from remote sensing maps), the variables M and O considered in the performance indices of Table 3 have to be computed according to the data format, for the comparison to be consistent. Below, we specify how we handle each source of data, using X to indicate the generic variable, either observed (O) or simulated (M), obtained by spatially averaging the single realization x.

1. Water temperature profiles (APPA, ARPAV, IMAU, NIOZ). Observed and simulated values are averaged within layers of pre-defined thickness. The mean $X_{i,j}$ value per each layer j at the i-th time step is computed as follows:

$$X_{i,j} = \frac{1}{N_j}\sum_{n=1}^{N_j} x_{i,n}\ , \tag{1}$$

where $x_{i,n}$, $n = 1, \ldots, N_j$ are the values within the j-th layer, at time i.

2. Satellite time series at single pixels (AVHRR, MODIS, LANDSAT8). Observations are referred to a pre-defined area A generally larger than the single computational grid cell. Hence simulated values are computed by evaluating a horizontal average among the computational cells laying within A:

$$X_i = \frac{1}{N}\sum_{n=1}^{N} x_{i,n}\ , \tag{2}$$

where $x_{i,n}$, $n = 1, \ldots, N$ are the values within A and i is the time step.

3. Satellite maps (LANDSAT-8 maps).The number of pixels where observations are available depends on the sensor resolution (100 m), the cloudiness conditions and the acquisition time. Satellite and model LSWTs are upscaled to a regular grid of 500 m grid spacing, which is comparable with the largest hydrodynamic grid cell. On these maps, the synthetic indices in Table 3 indicate space instead of time: hence the index i identifies a particular pixel, such that $i = 1 \ldots N$, with N the total number of up-scaled pixels where LSWT from satellites is available at each acquisition time. Although a bias might exist between the temperature retrieved from satellite imagery (skin) and the in situ water temperature (bulk) (e.g., Prats et al., 2018), in this study the skin-to-bulk correction is not applied. Such bias can reach up to few °C, it is larger during daylight hours, and is strictly dependent on the meteorological conditions (Wilson et al., 2013). Therefore, in order to overcome the skin-to-bulk issue, a centered (i.e., unbiased[3]) anomaly $\Delta_c T$ is computed for comparing model results with Landsat LWST full maps. We determine this index in each i-th pixel of the maps as follows:

$$\Delta_c T_i = \left(M_i - \overline{M}\right) - \left(O_i - \overline{O}\right)\ , \tag{3}$$

where $\overline{M}$ ($\overline{O}$) is the spatial mean of simulated (observed) temperature over the lake surface in one single scene, and M_i (O_i) is the single pixel of the scene. The computation of this index allows to filter out the inherent difference between remotely sensed skin temperature and the simulated lake surface temperature, assuming that this is associated with the respective biases.

When moving down from monthly to daily and sub-daily scales, the range of temperature variations greatly reduces. In our dataset, temperature data at a scale shorter than one month are only available from the APPA buoy (see Table 1 for the temporal resolution and Fig. 1 for the location of the point) and from the NIOZ thermistor chain. To analyze the performance of the model in these two points, we use the wavelet time frequency analysis, which has already been applied to the processing of non-stationary temperature signals in other lakes (e.g. Antenucci et al., 2000; Boegman et al., 2005; Guyennon et al., 2014; Woolway et al., 2014). We perform this analysis on the time series of water temperature and wind. Both signals are decomposed by using the Morlet continuous wavelet (Grossmann and Morlet, 1984), in order to derive the variation in time of the power spectrum of dominant frequencies (WPS, Wavelet Power Spectrum). By integrating in time the WPS, the Global Wavelet Spectrum (GWS) is also computed. This allows for an overall comparison between simulated and observed time series, based on a consistent estimation of the most energetic levels within their power spectrum.

4.4. Calibration of the hydrodynamic model

The hydrodynamic model is calibrated comparing the simulated water temperature profile in the APPA point with the corresponding measured temperature profiles in the biennium 2004–2005. The calibration period is chosen so as to correctly reproduce one year of meromixis (2004) and one year of complete overturning (2005). As anticipated in Sect. 4.2, the Delft3D model allows tuning a limited number of parameters, some related to the heat flux model (Stanton, Dalton and free convection coefficients), some to the turbulence model (background eddy viscosity and diffusivity, both horizontal and vertical) and to the boundary conditions at the surface and at the bottom (wind and bottom drag coefficients). In this work, we calibrated the Stanton, Dalton and free convection coefficients, and the background vertical turbulent terms. Table 2 lists the calibrated parameters, the adopted values, and the range of values tested, while in Table 4 the statistics from the calibration simulation are displayed in terms of the indices presented in Table 3.

The indices of performance are computed on the temperature

[3] We refer to Taylor (2001) for the use of the term 'centered' in the sense of 'unbiased'.

Table 4
Performance metrics of the hydrodynamic model in the calibration phase (APPA point, 2004–2005 run).

ID	Depth	MAE [°C]	ME [°C]	RMSD [°C]	NRMSD [-]	NSD [-]	CORR [-]	NSE [-]
APPA	0–2 m	1.08	−0.40	1.49	0.26	1.03	0.97	0.93
	2–20 m	0.98	−0.25	1.32	0.29	1.08	0.96	0.91
	20–100 m	0.33	−0.02	0.39	0.41	1.21	0.95	0.83
	100–270 m	0.21	−0.21	0.23	0.33	0.91	0.95	0.21

profiles in the APPA point considering 4 layers, defined in terms of depths as 0–2 m (surface), 2–20 m (epilimnion), 20–100 m (metalimnion), 100-bottom m (hypolimnion). In APPA point, bottom depth is 270 m. The performance of the model is quite satisfactory at all depths: the maximum MAE is obtained at the surface and in the epilimnion (1.08 °C and 0.98 °C respectively), corresponding to a RMSD of 1.49 °C and 1.32 °C. This error is below 30% of the standard deviation of temperature in both surface and epilimnion (NRMSD of 0.26 and 0.29). In these layers NSD, CORR and NSE indices are very close to 1, showing that the model is very well correlated with observations, even in case of extreme temperature values. However, smaller MAE (0.33 °C) and RMSD (0.39 °C) in the interior layer are associated with a normalized error of the order of 40%, as the standard deviation of temperature in the metalimnion is smaller than in the surface layer. Nevertheless, the good performance of the model is demonstrated by the other indices, which are still very close to 1. In the bottom layer, the order of magnitude of MAE and RMSD is the same as ME. CORR and NSD are close to 1 (0.95 and 0.91, respectively), but NSE is below the "good" threshold value (0.5 according to Moriasi et al., 2007), suggesting that the error is small but systematic, and larger than the very limited variability of observations. As a general trend, the model bias (ME) is negative in all layers, indicating that the simulated temperatures are on average slightly colder (less than 0.5 °C along the whole water column) than the observed ones.

Table 5
Performance metrics of the WRF model.

ID	MAE [m s^{-1}]	ME [m s^{-1}]	RMSD [m s^{-1}]	NRMSD [-]	NSD [-]
Wind speed					
MET1	1.78	0.19	2.49	0.14	0.77
MET2	1.49	−0.24	2.05	0.13	0.93
MET3	1.89	1.16	2.5	0.21	1.32
MET4	1.84	1.49	2.46	0.23	1.92
MET5	1.96	1.09	2.91	0.21	1.82
MET6	1.33	0.96	1.84	0.22	1.9
MET7	1.64	0.03	2.21	0.15	1.17
MET8	1.42	1.03	1.97	0.27	1.86
Air Temperature					
ID	MAE [°C]	ME [°C]	RMSD [°C]	NRMSD [-]	NSD [-]
MET1	2.52	−1.27	3.23	0.08	0.96
MET2	2.19	−1.11	2.95	0.06	1.01
MET3	2.50	−1.39	3.10	0.08	0.95
MET4	2.67	−1.79	3.22	0.07	0.90
MET5	2.33	−0.68	2.91	0.07	0.92
MET6	1.97	−0.81	2.50	0.06	0.98
MET7	2.00	−0.68	2.50	0.08	1.45
MET8	1.90	0.03	2.41	0.05	0.98
Relative Humidity					
ID	MAE [%]	ME [%]	RMSD [%]	NRMSD [-]	NSD [-]
MET1	15.01	−8.37	19.17	0.19	0.80
MET2	13.05	0.02	17.06	0.19	0.91
MET3	13.30	−7.14	17.02	0.18	0.82
MET4	15.30	−5.88	19.44	0.21	0.71
MET5	13.34	−5.44	17.16	0.19	0.82
MET6	14.82	−8.80	19.36	0.21	0.86
MET7	13.79	−8.84	17.84	0.22	1.18
MET8	15.51	−10.64	20.15	0.23	0.79

5. Results

In this section, the models are tested against the whole dataset. The results of the long-term simulation on the period 2004–2018 are compared with observations in all the monitoring locations described in Fig. 1. The general performance of the atmospheric and hydrodynamic models is presented in Section 5.1, as a fundamental step ensuring the consistency of the subsequent analyses. Then, the analysis at different temporal and spatial scales follows in sections 5.2-5.5.

5.1. Overview of the model performance against the available dataset

5.1.1. Atmospheric model

Table 5 summarizes the performance metrics of the WRF model against wind speed, air temperature and relative humidity observations at hourly time resolution.

The model tends to overestimate wind speed at all weather stations except MET2, with ME of the order of ~ 1 m s^{-1}. A similar overestimation of the wind speed, in particular in low wind speed conditions, is also found in other studies using WRF (e.g., Horvath et al., 2012). The simulated air temperature leads to negative ME at most weather stations, with the largest values at MET3 (−1.39 °C) and MET4 (−1.79 °C), while the highest MAE and RMSD are found at MET1 (2.50 °C and 3.42 °C) and MET5 (2.67 °C and 3.22 °C). NSD values are close to 1 at all stations, showing that the model is able to capture the actual temperature variability. Finally, the relative humidity predicted by the model is generally lower than the observed value, with negative ME up to −10.64% (MET8), with the exception of MET2 (0.02%). In particular, air humidity is underestimated at stations located in mixed water-soil cells, where the model hardly captures the microclimatic conditions. In fact, the stations close to the shores are affected by higher humidity associated with evaporation from the lake surface.

As a general pattern, the largest errors are found at the stations located along the northwestern shores of the lake (MET1, MET3, MET4 and MET5). This fact may be related to the poor representation of the steep topography characterizing this area, due to the 2 km resolution of the model grid.

Fig. 2 shows the ME of WRF for wind speed and temperature at MET1 and MET2, evaluated on a monthly basis. MET1 (Fig. 2a) shows a clear positive correlation between temperature and wind speed MEs. In fact, overestimation (underestimation) of air temperature leads to larger (smaller) land-lake thermal contrasts, and, as a consequence, to stronger (weaker) lake breezes, especially during spring and summer. Although the atmospheric model is generally able to capture the diurnal cycle of lake-land breezes, errors can be present in the exact timing and strength of the circulations, depending also on the correct simulation of lake-land temperature contrasts. Moreover, errors present a seasonal dependence, with positive MEs in winter and negative in summer. Also at MET2 (Fig. 2b), air temperature tends to be overestimated during wintertime, but no evident correlation with wind speed errors is found. In fact, MET2 is located some kilometers away from the lake shore and thus it is less affected (than MET1) by lake breezes.

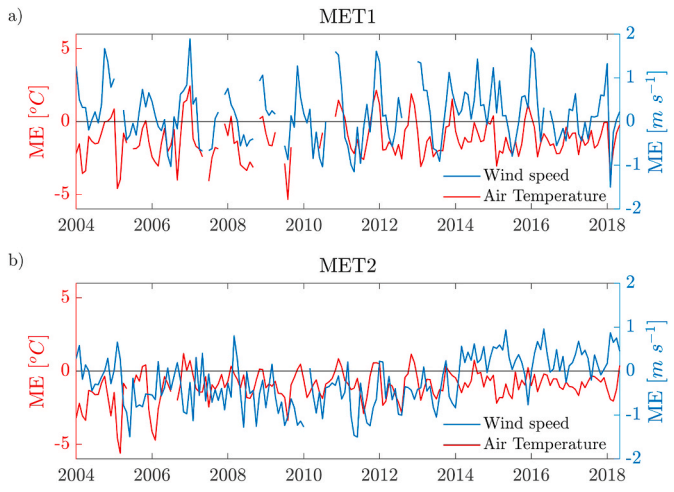

Fig. 2. Temporal evolution of monthly averaged bias (ME) of wind speed (right axis) and air temperature (left axis) at a) MET1 and b) MET2. Note: gaps correspond to missing data in the time series of air temperature and wind speed measurements.

5.1.2. Hydrodynamic model

The performance of the hydrodynamic model is evaluated by testing the simulated water temperature against the whole dataset (i.e., at all observation points and for the whole simulated period). The performance metrics are summarized in Table 3. Results are provided in Table 6 for each dataset, and are summarized in Fig. 3.

For data available along the water column (temperature profiles), the simulated and observed values are averaged following Equation (1) in four depth intervals (0–2 m, 2–20 m, 20–100, 100-bottom m, as in Sect. 4.4), where the performance metrics are computed. We include in the surface layer (i.e., 0–2 m, Fig. 3a) also the comparison with the time series of satellite data: in this case the indices are derived according to Equation (2).

In Fig. 3, the four Taylor diagrams (Taylor, 2001) display the NSD and the NRMSD of the simulations versus the measurements for each layer. Differently from the standard use of Taylor diagrams, here we use the numerical outputs as reference (red dots), and compare them with the measurements in the different locations (different markers). Since each observation point has its SD, the normalization of the statistical indices is necessary in view of the comparison. In the four diagrams, all the points are in the bottom right part of the plot and close to the reference point. This indicates that the model is well correlated with all datasets and well captures the temperature variability in each location along the water column.

In the surface layer (Fig. 3a), the NRSMD is between 0.25 and 0.5 almost everywhere, with the lowest values in AVHRR points. Among all the time series from remote sensing, AVHRR1 and AVHRR2 show the best agreement with model results, with MAE less than 1.2 °C, while MAE is 2.5 °C in MODIS, and between 1.3 and 2 °C in Landsat-8 control points (see also Table 6 for the numerical values of the performance metrics). The worst performance is found in MODIS-4, L8-4 and APPA *in situ* points. All these points lie in the middle of the northern narrow trunk of the lake, and are very close to each other. Interestingly, the NSD is larger than 1 in APPA (1.20) and L8-4 (1.25), but smaller than 1 in MODIS4 (0.88). Similarly, the ME in this point switches from positive in APPA (0.65), APPA buoy (1.47) and L8-4 (1.32), to negative in MODIS4 (−0.34).

In the epilimnion (Fig. 3b), all IMAU points show larger NSD in comparison with APPA point (IMAU ∼ 1.30, APPA ∼ 1.18). On the other hand, the APPA buoy point, providing data at 10 m below water surface, shows an unsatisfactory performance. In this location, the largest ME and RMSD are achieved (1.47 °C and 2.85 °C), but the correlation is extremely high (0.90). Both in surface and epilimnion layers the NSE is above 0.8 everywhere except for APPA buoy points, where it is still >0.5.

Table 6
Performance metrics of the hydrodynamic model (all points, 2004–2018 run).

ID	Depth	MAE [°C]	ME [°C]	RMSD [°C]	NRMSD [-]	NSD [-]	CORR [-]	NSE [-]
APPA	0–2 m	1.44	0.65	2.17	0.42	1.20	0.94	0.81
	2–20 m	1.21	0.48	1.70	0.39	1.18	0.95	0.84
	20–100 m	0.53	−0.28	0.70	0.54	1.04	0.86	0.65
	100–270 m	0.25	−0.21	0.32	0.63	1.27	0.87	0.29
APPA buoy	10 m	1.94	1.47	2.85	0.55	1.25	0.90	0.58
ARPAV1	0–2 m	1.12	−0.22	1.94	0.38	1.13	0.94	0.85
	2–20 m	1.03	0.25	1.49	0.32	1.14	0.96	0.89
	20–100 m	0.43	−0.26	0.54	0.44	1.00	0.90	0.75
	100–330 m	0.22	−0.20	0.30	0.49	1.14	0.90	0.55
ARPAV2	0–2 m	1.04	−0.57	1.73	0.30	1.09	0.96	0.90
	2–20 m	1.04	0.34	1.57	0.31	1.14	0.96	0.90
	20–70 m	0.52	−0.19	0.74	0.51	1.03	0.87	0.72
IMAU1	2–20 m	0.76	0.52	1.40	0.35	1.30	0.99	0.86
	20–100 m	0.34	0.05	0.45	0.35	1.16	0.96	0.88
IMAU2	2–20 m	0.72	0.47	1.37	0.34	1.29	0.99	0.87
	20–100 m	0.32	0.05	0.41	0.31	1.23	0.98	0.90
IMAU3	2–20 m	0.84	0.45	1.52	0.38	1.31	0.98	0.84
	20–100 m	0.35	−0.05	0.45	0.35	1.11	0.95	0.88
IMAU4	2–20 m	0.68	0.49	1.44	0.36	1.30	0.98	0.85
	20–100 m	0.36	−0.10	0.47	0.33	1.05	0.95	0.88
NIOZ	100–330m	0.04	−0.04	0.04	1.37	2.24	0.92	−8.18
AVHRR1	surface	1.18	−0.23	1.59	0.25	1.07	0.97	0.93
AVHRR2	surface	1.13	0.18	1.52	0.27	1.09	0.97	0.93
MODIS1	surface	1.54	0.14	2.01	0.32	1.09	0.96	0.90
MODIS2	surface	1.65	−0.02	2.17	0.35	1.05	0.94	0.88
MODIS3	surface	1.64	−0.19	2.16	0.34	1.05	0.95	0.88
MODIS4	surface	2.51	−0.34	3.14	0.46	0.88	0.89	0.78
L8-1	surface	1.45	−0.09	1.70	0.32	1.12	0.96	0.90
L8-2	surface	1.37	−0.07	1.86	0.33	1.08	0.95	0.89
L8-3	surface	1.98	0.47	2.57	0.45	1.12	0.92	0.79
L8-4	surface	1.88	1.32	2.53	0.46	1.25	0.94	0.71

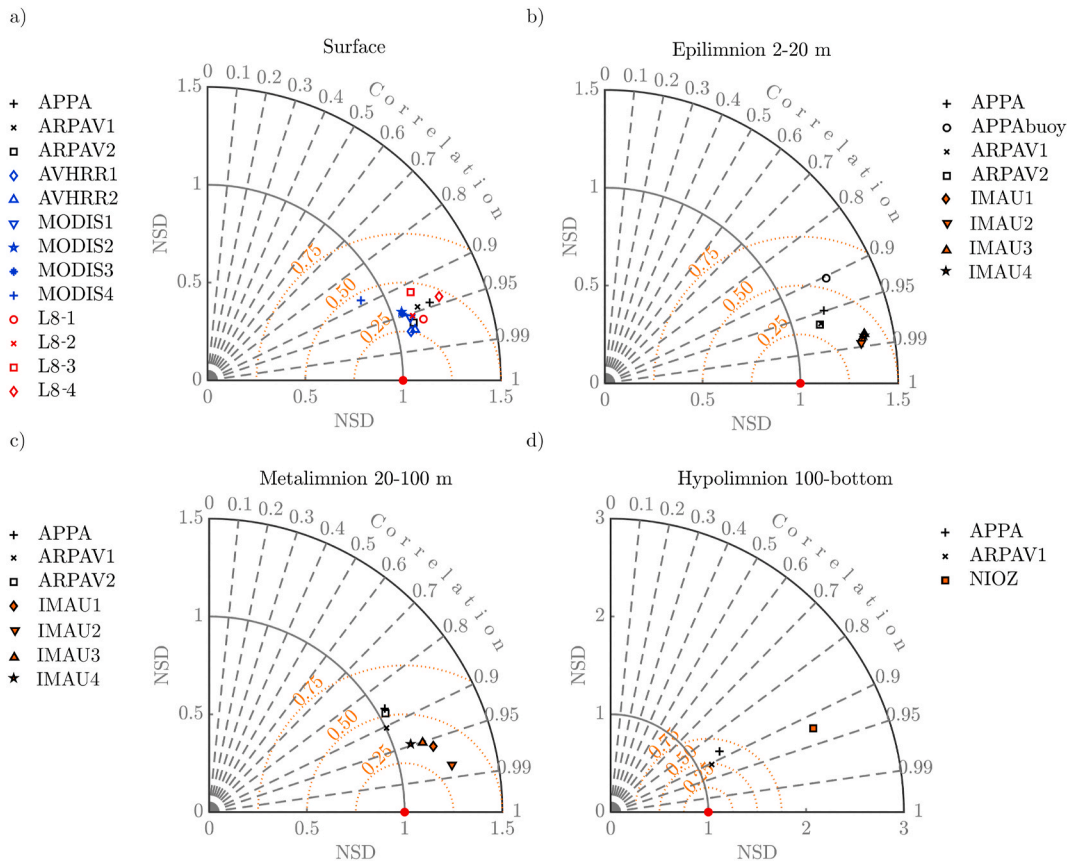

Fig. 3. Taylor diagrams comparing the Delft3D model performance against all data available in different layers: a) surface, b) epilimnion, c) metalimnion; d) hypolimnion. Orange dotted curves describe the range of NRSD where each model-observation comparison resides. NRSD, NSD and CORR are computed as defined in Table 3. The numerical values of the points are provided in Table 6.

In the metalimnion (Fig. 3c), the model errors are smaller in absolute values, but the normalized performance indices are larger due to the smaller range spanned by temperature observations: in this layer, the largest NRMSD is at the APPA point (0.54). Correlation is everywhere larger than 0.90 with the exception of APPA (0.86). The NSE is smaller due to the limited variance of data, but the values are still acceptable, and the worse value is again at the APPA point (0.65, see Table 6).

In the hypolimnion (Fig. 3d), the comparison is limited to a rather small number of observation points. The smallest RMSD in this layer (0.04 °C) at the NIOZ point is linked to the largest NRMSD (1.37) and NSD (2.24). In APPA and ARPAV1 points, the normalized indices present better values (NRSD = 0.63, 0.49, NSD = 1.27, 1.14, respectively), but the RMSD is larger (0.32, 0.30, respectively). In general, the normalized indices (NRSD, NSD) provide worse results for deeper layers and shorter time series, due to the smaller standard deviation of the observations, which is the normalization factor. In fact, the EPAs datasets span a longer period, and the bottom temperature varied up to 1 °C in the 14 simulated years, while the NIOZ dataset contains measurements for just one year (2017–2018), during which the maximum temperature variation is about ~0.3 °C. Hence, even if the RMSD is 10 times smaller in the NIOZ point (0.04 °C against 0.30 °C in APPA and ARPAV1), the normalized values and the Taylor diagram (Fig. 3d) suggest that the model performance is poor at that point (NRMSD = 1.37 against 0.63 and 0.49 in APPA and ARPAV1 points, respectively). On the other hand, the correlation (CORR) is very good in all datasets. The same reasoning applies to NSE, which is positive in APPA and ARPAV1 (being even larger than 0.5 in the latter point), with the exception of the NIOZ point, where the negative NSE is again due to the small range of variability of the time series. In all deep points, however, the RMSD is on the same order of magnitude as the ME, and the error can be attributed to a negative bias of the model near the bottom. In this regard, ME is generally negative everywhere in the surface and in the bottom, except for the surface layer in the northern sub-basin (APPA, APPA buoy, AVHRR2, L8-4, IMAU points). At the same time, the metalimnetic layer shows a positive bias in all points.

5.2. Interannual variability

In this section, we analyze the long-term dynamics of the temperature changes, showing also how they may be affected by specific events, such as a complete overturning, that produce a long-lasting signature on deep water temperature. To this end, we focus on observed and simulated temperature changes over 14 years in two observation points: ARPAV1 and ARPAV2 (see Fig. 1). This analysis aims at evaluating the model performance in reproducing the interannual variability of the thermal regime in two parts of the lake displaying very different morphological characteristics. In fact, ARPAV1 is located in the deepest part of the northern trunk and reaches the depth of 337.5 m, while ARPAV2 lies in the southern wider basin and its depth is 69 m. In Fig. 4 (top plots), the simulated daily temperatures at ARPAV1 (a) and ARPAV2 (b) are compared with the temperature monthly measured. For both points, the temperature is plotted at three levels. Simulated and measured values are calculated as the mean value over the upper 10 m (hereafter referred as “surface”), the deepest 10 m (“bottom”), and in a 10 m-thick layer around the depth of 30 m (“interior”), according to Equation (1). The figure shows that in both points the model captures the interannual trend of temperature and the thermal gradients between the surface and the deeper layers. In ARPAV1 (Fig. 4a), the bottom temperature does not change significantly during single years, but the warming of 1 °C observed from 2004 to 2018 is correctly reproduced by

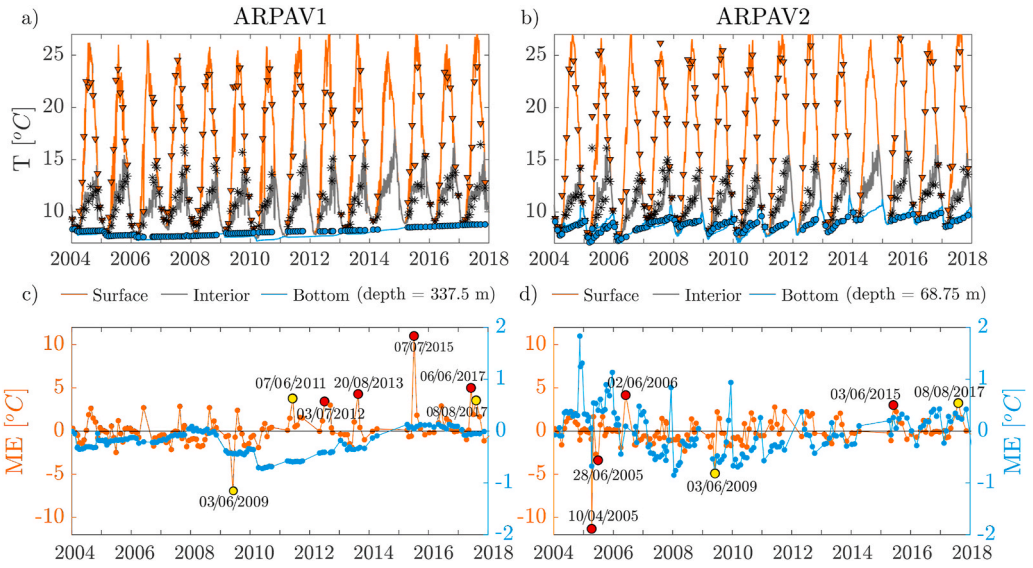

Fig. 4. a-b) Evolution of temperature at the surface (mean over the upper 10 m), interior (mean between 25 and 35 m depth) and bottom (mean over the deepest 10 m) from model results (continuous lines) and observations (markers) in ARPAV1 (a) and ARPAV2 (b) points. c-d) Evolution of mean error at the surface (orange dotted line) and at the bottom (light blue dotted line) in the two points. Red dots indicate days (date format dd/mm/yyyy) when surface $|ME| > 3\ ^{\circ}C$. Yellow dots highlight the dates for which a more detailed analysis of the atmospheric forcing is provided in Fig. 5.

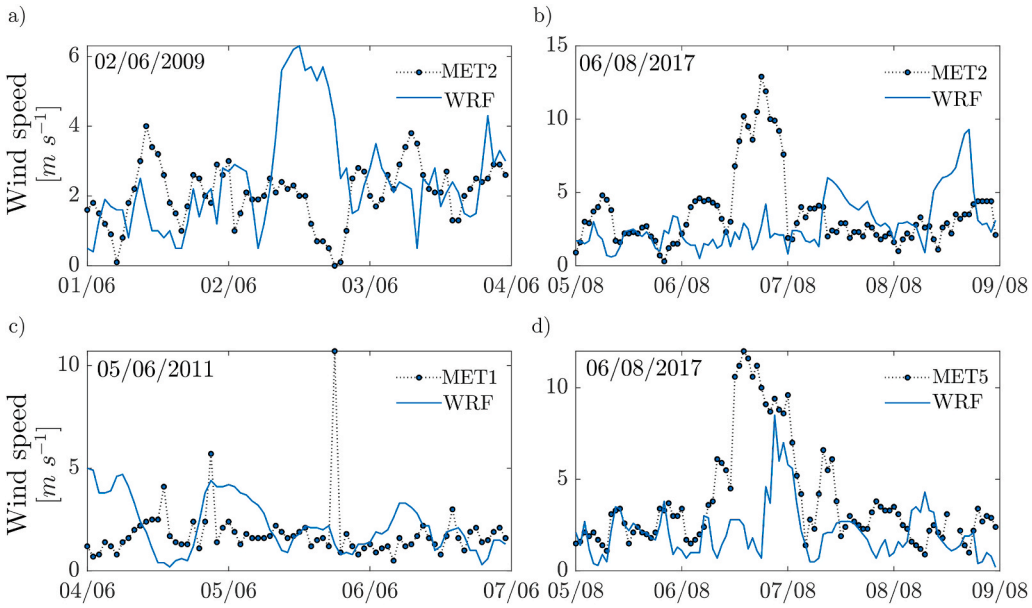

Fig. 5. Time series of observed and simulated wind speed at a) MET2 on June 02, 2009, b) MET2 and d) MET5 on August 06, 2017, c) MET1 on June 05, 2011 (date format dd/mm/yyyy).

the model. The model correctly captures the deep mixing events in 2005 and 2006 (the last events Lake Garda experienced, according to Salmaso et al., 2017). Likewise, the model reproduces the partial mixing observed in 2017, that Piccolroaz et al. (2019) attributed to a wind-driven deep mixing. In ARPAV2 (Fig. 4b), the model correctly shows that the southern part of the lake is warmer than the northern part, with maximum summer temperatures generally higher than at ARPAV1 (Fig. 4a). At the bottom, the model correctly reproduces a larger variability of temperature (up to 3 °C in 2005) and the complete mixing of the whole water column occurring every year.

The evolution of the mean model error is shown in Fig. 4 (bottom plots, panels c and d). For an easier reading of the plot, different y-axes have been used for surface (left, orange) and bottom (right, light blue) errors. The figures show that in both points the model error does not increase in time either at the surface or at the bottom. The largest surface errors occur during summer, and are often connected with the large errors at the bottom. In the deep ARPAV1 point (Fig. 4c), the error at the bottom is maximum (in absolute terms) between 2009 and 2013, when the simulated value is colder than observations. This is associated with an incorrect simulation of two deep mixing events, in 2009 and 2010, which are not observed and cause two consecutive drops in the bottom temperature. However, the error decreases from 2011 onwards, as the model does not produce false deep mixing events anymore and the bottom temperature rises until reaching more realistic values in 2013. In the shallower ARPAV2 point (Fig. 4d), the model error is larger at the bottom, compared to that at the ARPAV1 point (Fig. 4c): here, the maximum depth is shallower (69 m instead of 344 m), hence the error at the bottom has the order of magnitude of metalimnetic errors (see in Table 6). In Fig. 4c and d, red and yellow dots indicate days in which significant inaccuracies affect the simulated surface temperature ($|ME| > 3\ ^{\circ}C$). A significant error is found on April 2005 in ARPAV2 point (Fig. 4d), where simulated surface (bottom) temperature is 10 °C (1 °C) colder than observed. Such an error arises from the combination of a large underestimation of air temperature and overestimation of wind

speed in this region of the lake (see Fig. 2b) during the previous months. In fact, the atmospheric model has a mean error on air temperature of $-4.56\ ^{\circ}$C in February and $-5.61\ ^{\circ}$C in March 2005, which drives the fictitious cooling of the water column detected in April. A negative error is found in ARPAV1 (Fig. 4 c) as well, but the absolute value is much smaller than in ARPAV2 ($< 2.5\ ^{\circ}$ C at the surface, $< 0.5\ ^{\circ}$ C at the bottom), due to the larger local depth.

Similarly, the other days displaying large errors coincide with periods of wind forcing significantly over/underestimated by the atmospheric model. In Fig. 5, the comparison between the simulated and observed wind speed at hourly scale is provided for some of these days (indicated as yellow dots in the previous figures). In all reported cases, it can be seen that the atmospheric model provides a good representation of the observed values, except for a limited time slot. More in detail, the WRF model significantly overestimates wind speed at MET2 in the central hours of June 02, 2009 (Fig. 5a), producing an excessive cooling at both ARPAV points on June 03, 2009. On the contrary, an increase of wind speed due to the passage of a cold front is not reproduced at MET2 and MET5 in the evening of August 06, 2017 (Fig. 5b and d, respectively), leading to the underestimation of wind-driven cooling on August 08, 2017. A similar explanation holds for the positive mean error detected in ARPAV1 on June 07, 2011: observations at MET1 on June 05, 2011 show a sudden increase of wind speed at 18 UTC+1, due to a local thunderstorm, which is not captured by the atmospheric model (Fig. 5c).

5.3. Seasonal cycle

The seasonal variability of water temperature in the lake may be characterized in a simple and effective scheme containing most of the relevant information. Fig. 6 shows the temperature variation along the water column at ARPAV1 for each month of the year, as seen from the model and from the observations. The simulated profiles are obtained by computing the statistics of the simulation output over the whole simulated period. The distribution of simulated temperature is indicated with boxplots (excluding outliers, for the sake of clarity). Single water temperature profiles from observations are plotted as grey dots.

The figure shows four characteristic periods. From February to March, the lake is nearly unstratified, with temperatures ranging in an interval of less than $1\ ^{\circ}$C throughout the water column. April, May and June are characterized by the developing stratification, and temperature experiences a great variability (surface temperatures span a range of $10\ ^{\circ}$C in June). The largest surface-bottom difference is reached in the summer months, from July to September, when it ranges between $10\ ^{\circ}$C and $20\ ^{\circ}$C, and surface layers experience a similar range of variability ($15\ ^{\circ}$C). From October to January, the autumn-early winter destratification typically occurs, and the range of surface variability decreases from $5\ ^{\circ}$C (October) to $2\ ^{\circ}$C (January).

The model correctly reproduces the range of variation of the temperature in each month and along the whole water column. Exceptions are found in summer months (June, July and August), when few episodes of very cold temperatures are not captured by the model. The temperatures observed in those episodes lie beyond the normal range of variability, and the dates correspond to extremely windy days. These are not properly reproduced by the atmospheric model (see Fig. 5), inducing subsequent errors in the hydrodynamic model results (see Fig. 4c for the details on the dates).

5.4. From monthly to sub-daily dynamics

In this section, we refer to the data at hourly scale available for the APPA buoy (surface) and from the NIOZ thermistor chain (deep water), and compare observed and simulated temperature in the epilimnion and hypolimnion, using a wavelet time-frequency analysis. At the same time, we also analyze observed and simulated wind speed at the nearest

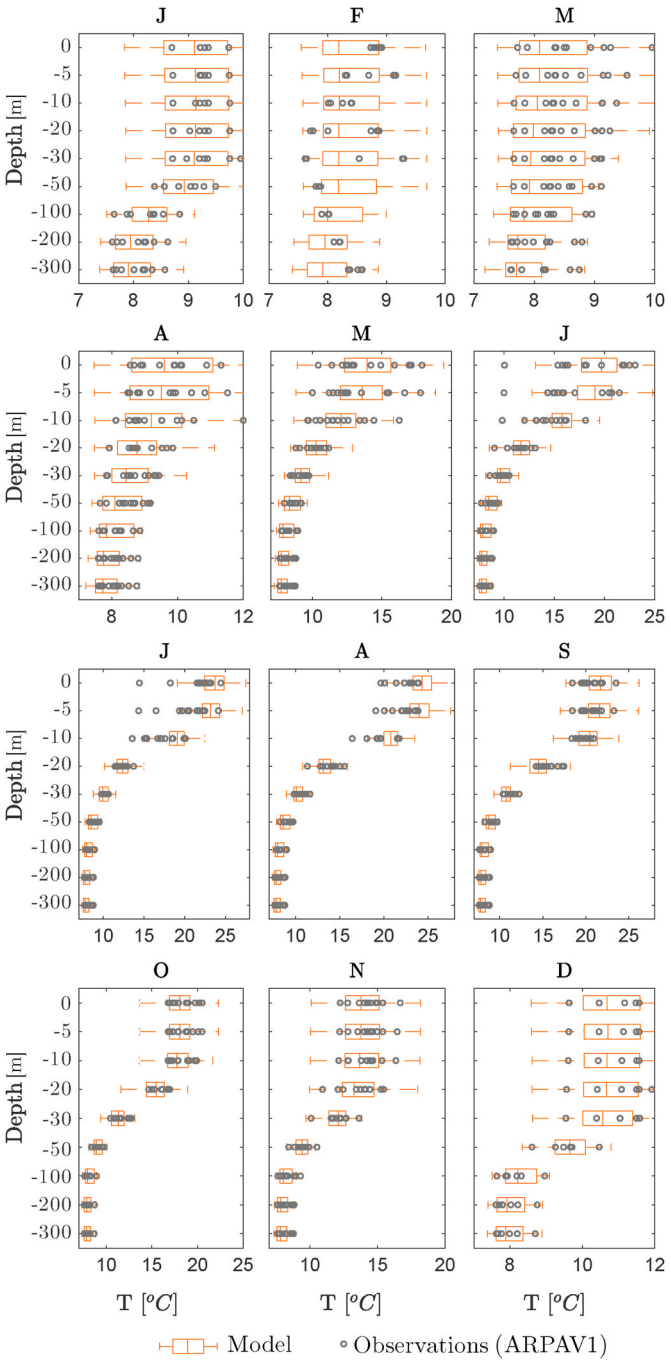

Fig. 6. Comparison between monthly averaged water temperature profiles over 2004–2018 among model results (orange box plots) and observations (grey dots) in the reference point ARPAV1. The boxes depict the median, the 25th and 75th percentiles of the monthly variation of temperature from the model results. The whiskers extend to the minimum and maximum data points not considering outliers, which are not plotted. Observations are plotted as single scattered profiles.

weather stations, using the same approach. Fig. 7 presents the results for the near surface dynamics from 2013 to 2018. The figure compares the wavelet power spectra (WPS) of measured (a) and simulated (b) wind velocity at MET1 station, and the epilimnetic water temperature at 10 m depth at APPA buoy station (d,e). Additionally, the global wavelet spectrum (GWS) for both variables (wind, c, and temperature, f) displays the time-averaged relevant energetic levels from observed and simulated time frequencies.

The figure shows that the WPS of the model results is fully coherent

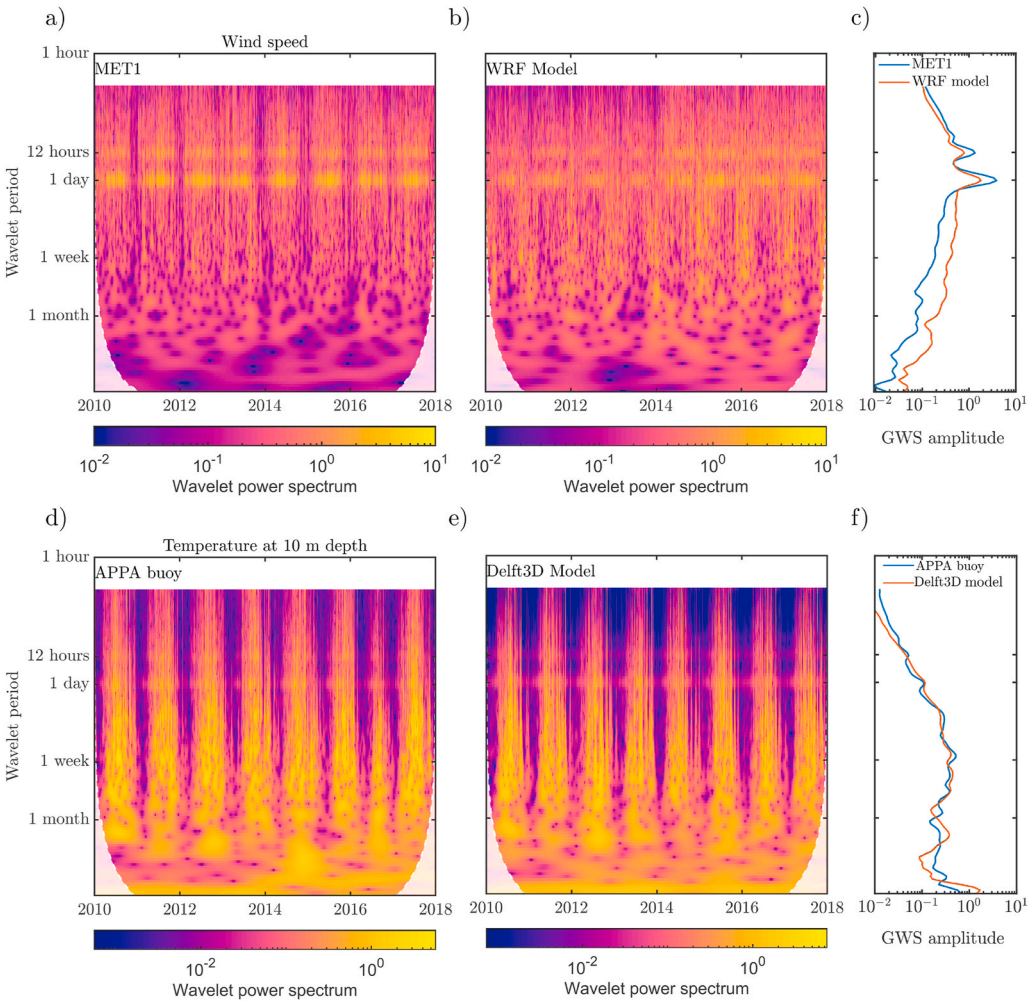

Fig. 7. Wavelet analysis of (top plots) wind velocity and (bottom plots) epilimnetic temperature. Left plots: wavelet power spectrum (WPS) of a) wind observations at MET1 weather station; b) wind simulated by WRF at the corresponding model cell; d) water temperature measured by APPA buoy thermistor at 10 m depth; e) water temperature simulated by Delft3D at APPA buoy point at 10 m depth. Right plots: global wavelet spectrum (GWS) computed from the previous WPS on (c) wind and (f) epilimnetic water temperature signal.

with data from observations, both for the wind and for the water temperature. WPS coherence between observed and simulated wind indicates that the model captures the most important weather events. More energetic areas (orange to yellow) correspond to the summertime period, while less energetic periods (blue to purple) are found in wintertime. The wind signal (Fig. 7a and b) shows that much more energy is associated to semi-diurnal and diurnal breezes in summer (as described by Giovannini et al., 2017). On the contrary, the signal is less intense during wintertime, when synoptic events are responsible for the energy peaks at longer time scales (some days to one week). At a global scale, the main energetic levels are at 12 and 24 h (Fig. 7c). By comparing the WPS of the temperature signal (Fig. 7d and e), we can see how these weather events impact the lake state, involving multiple time scales. The epilimnetic temperature spectrum (Fig. 7d and e) resembles the wind signal at the semi-diurnal to diurnal periodicity in all seasons. During summertime, when the lake is strongly stratified, the energy input from the wind cascades from low-frequency internal waves at basin scale to high-frequency waves (Preusse et al., 2010). Thus wind excites a more heterogeneous wave field, whose dominant periodicities range from hours to months, and can be clearly seen at a global scale in the GWS (Fig. 7f).

In Fig. 8, the dynamics of the last simulated years (2017–2018) are presented at the NIOZ point. The figure compares the WPS of measured (a) and simulated (b) wind velocity at MET5 station, and hypolimnetic water temperature at 187 m depth at NIOZ station (d,e). Again, the GWS of wind (c) and deep water temperature (f) summarize the main energetic levels. In order to better identify the seasonal alternation, the GWS for both variables are represented differently between spring to autumn integrated spectrum (stratified period, from May to October, continuous lines) and winter spectrum (nearly unstratified period from November to April, dashed lines). Measured (a) and simulated (b) wind velocities display dominant energetic levels at semi-diurnal and diurnal frequencies (mainly in the stratified period, see subplot c) and few intense events of daily to weekly periodicity (e.g., November 2017 and March–April 2018). The wavelet analysis confirms that the atmospheric model missed the intense wind event occurred in August 2017 (as already seen in Sect. 5.2, Fig. 5b,d).

The signature of semi-diurnal, daily and synoptic winds can be found in the observed (d) and simulated (e) deep temperature signal as well. The hydrodynamic model reproduces predominant semi-diurnal periodicity during the stratified period (see the GWS in subplot e), while during winter periodicities around 5–6 days are more energetic. These results are consistent with the observations of van Haren et al. (2020). The global power spectra (Fig. 8c,f) highlight that both atmospheric and hydrodynamic model overestimate the energy of wind and deep water temperature signal at all energetic levels, especially at the longer scales in winter.

5.5. Spatial gradients

In the previous sections, the performance of the hydrodynamic model in different points of the lake is presented (sections 5.1.2 and 5.2). While it is locally acceptable in most of the in situ locations (see Table 6), difficulties arise when comparing the model performance (i.e.,

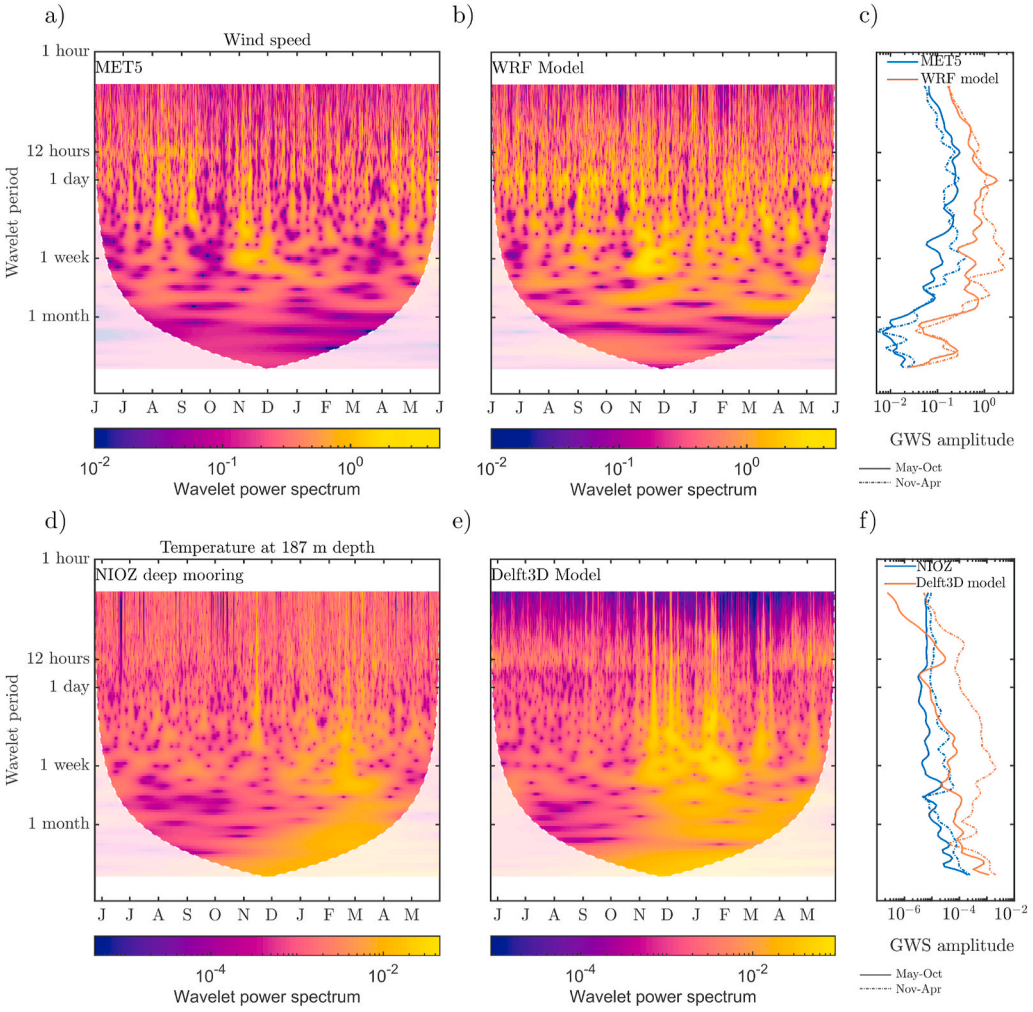

Fig. 8. Wavelet analysis of (top plots) wind velocity and (bottom plots) hypolimnetic temperature. Left plots: wavelet power spectrum (WPS) of a) wind observations at MET5 weather station; b) wind simulated by WRF at the corresponding model cell; d) water temperature measured by NIOZ thermistor at 187 m depth; e) water temperature simulated by Delft3D at NIOZ point at 187 m depth. Right plots: global wavelet spectrum (GWS) computed from the previous WPS on (c) wind and (f) hypolimnetic water temperature signal.

the numerical values of the metrics) in different regions of the lake. In fact, the comparison of statistical indexes is hindered by the different sampling frequencies and lengths of the time series. In order to overcome this issue, in this section we discuss the model performance in reproducing the spatial variation of water temperature on single dates.

In Fig. 9 the water temperature field at different depths is shown for day July 4, 2017 at 10 UTC, when the largest number of (nearly) simultaneous observations is available. At the surface (Fig. 9a), water temperature spans a range of 3 °C, from 23 °C to 26 °C, with increasing temperatures from north-west to south-east. Observations at the surface from Landsat-8 dataset (here reported in the control points listed in Table 1) confirm this gradient, despite a positive bias of approximately 1 °C exists between the simulated and remotely sensed water temperature (see RMSE and ME values reported in Fig. 10l and 11l and the corresponding comment below). At 12 m depth (Fig. 9b), a wider range is found, with temperatures ranging from 20 °C to 25 °C. The west-east gradient observed at the surface is inverted at this depth, possibly due to a tilt of the metalimnion, with colder temperatures being now along the eastern shore. In the southern part, the existence of a gyre structure is evidenced by the presence of cold water at the center of the sub-basin, surrounded by warmer water. In the whole lake the simulated water temperature field shows a good agreement with measurements in APPA, ARPAV1 and ARPAV2 stations. Moving deeper (Fig. 9c,d,e,f), the water temperature sharply decreases and the spatial variability of water temperature reduces to 1 °C at 50 m (c), 0.3 °C at 100 m (d) and less than 0.05 °C below 200 m (e,f). At 310 m (f), water temperature is nearly homogeneous and the simulated value (~ 8.735 °C) slightly overestimates the observations at ARPAV1 and NIOZ (ME 0.035 °C). In all layers, and for the selected date, the model correctly simulates both the horizontal and vertical variation of temperature at all available stations.

In Figs. 10 and 11 we compare the simulated surface water temperature with remote sensing LSWT data from Landsat-8 full maps (see Table 1 for the technical details of the sensor and the time availability of data). The consistency of such a comparison is ensured by the fact that each pixel of the satellite LSWT maps is associated to the same time and to the same sampling method. Fig. 10 shows the absolute value of the centered (unbiased, hence removing the spatial mean) anomaly $\Delta_c T$ between simulated and observed LSWT (see Equation (3), Sect. 4.3). In the top left corner of each subplot, we report the RMSD and ME computed for the single maps according to Table 3, without removing the spatial mean, thus providing the order of magnitude of the error over the whole lake surface. In most of the investigated dates, the RMSD is very similar to the ME. This suggests that most of the error is associated with the bias between skin temperature (retrieved from satellite imagery) and bulk temperature (simulated by the hydrodynamic model), which can reach up to few °C, depending on the weather conditions, especially during daytime (Wilson et al., 2013; Baracchini et al., 2020a).

For this reason, the advantage of introducing the unbiased anomaly $\Delta_c T$ index is easily understood. Indeed, by getting rid of the biases of both model and satellite data, the spatial distribution of the difference between the two can be visualized in an efficient way.

In most maps, $|\Delta_c T| < 1$ °C over the whole lake surface (e.g.,

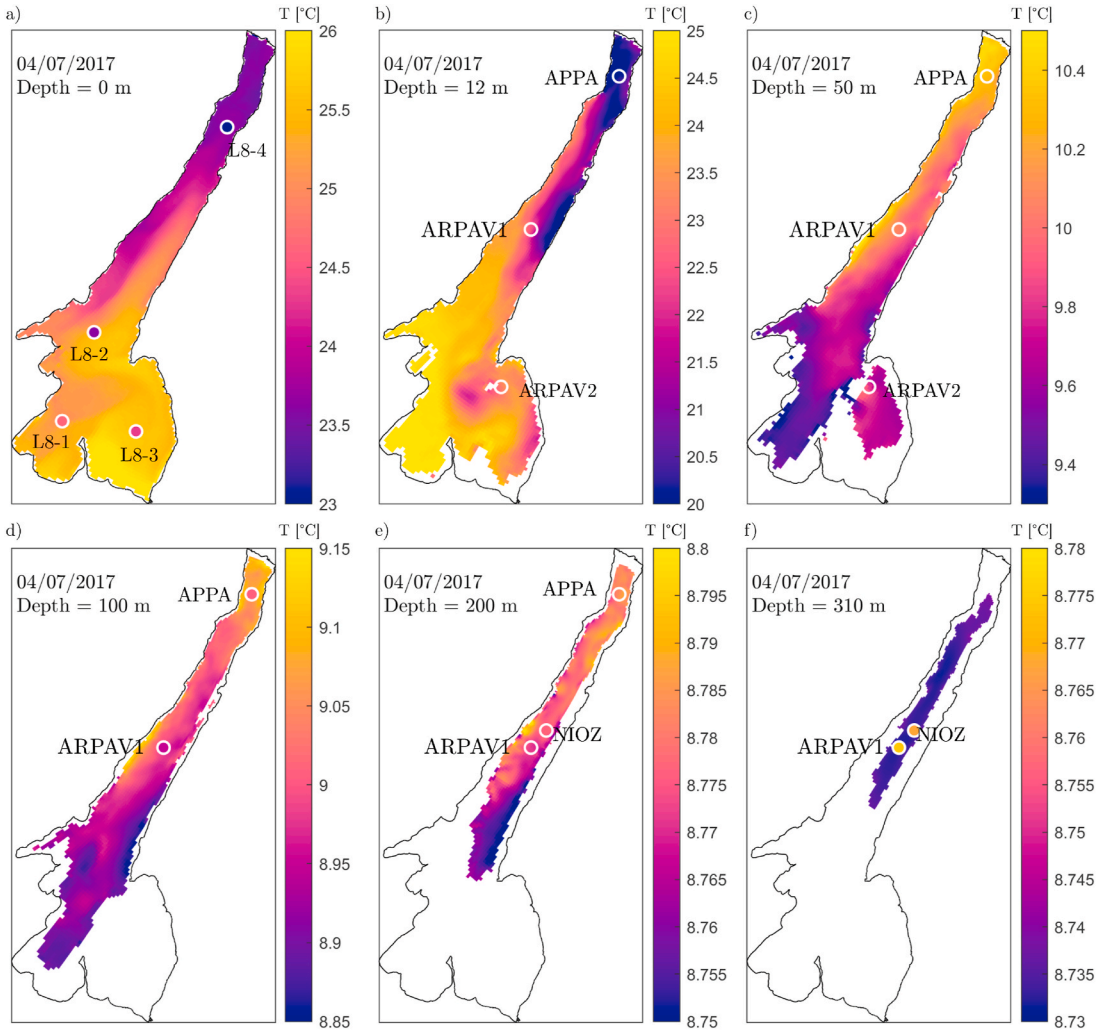

Fig. 9. Spatial variation of simulated (color plot) and observed (filled circles with white stroke) water temperature on July 4, 2017 at 10 UTC. Each subplot reports the simulated temperature field at a fixed depth below water surface and the measurements available at that depth (see Table 1).

Fig. 10a). Differently, the largest $\Delta_c T$ (up to 3 °C) is found in those days when also significant differences between ME and RMSD occur (panels c–d, g, j). The largest differences between model results and satellite data are localized in the northern trunk and during the summer-early autumn period (panels b–d, g, j, m), and often along the eastern shore (panels b–d, g, k, m, o, r). In the southern basin, larger errors can occur again along the eastern shore (panels a–d,f-g, j, m, q), while patterns with smaller errors are frequent in the middle of the sub-basin (panels a, k, m, q).

A similar comparison, but taking into account the sign of the unbiased anomaly, is provided in Fig. 11. By comparing the areas of large anomalies from the previous figure with the sign of the anomalies in the present figure, a clear pattern can be seen. In the northern trunk, large positive anomalies are caused by warmer simulated temperatures (panels b–d, f-h, j-n). In the southern basin, the negative anomalies along the eastern shore are due to colder model temperatures (panels a–d, f-g, j, m, q). In the middle of the sub-basin, the sign of the anomaly oscillates between negative (panels a, k-m) and positive (panels b–c, f, q-r).

These anomalies can be related to transport processes that are not well captured by the model and for which the timing can be crucial. In fact, the two areas showing the largest mismatch patches are found to be affected by specific current patterns in previous studies. Gyre circulations have been observed in the middle of both southern and northern basins by Amadori et al. (2018, 2020). Moreover, in the northern trunk, Piccolroaz et al. (2019) observed and simulated up/downwelling along the shores due to intense and long lasting wind events. The latter processes could actually be responsible for the mismatch between Landsat images (colder) and simulated (warmer) surface temperature along the eastern shore, especially during late morning (Landsat acquisition time is 9:50 UTC). In fact, the model might have missed a cold upwelling caused by a local wind named 'Peler', which regularly blows from northeast starting at late night, and persists until late morning. In this regard, the absence of current measurements or satellite maps at different times does not allow us to examine this bias in detail.

6. Discussion

The main goal of this work is to assess the capabilities of the Delft3D lake model, driven by WRF simulated outputs, to reproduce the lake thermal features over different time and spatial scales, and to evaluate how the atmospheric model results affect the lake model performance. As mentioned in Sect. 4.2, not all the parameters can be effectively tuned, as we calibrate the model on temperature data solely. For these parameters, either the default value is adopted (Chezy coefficient, horizontal turbulence terms) or formulations available from the literature are used (wind drag coefficient).

The performance of the numerical simulations is evaluated with typical statistical metrics (Sect. 5.1), whose values are consistent with

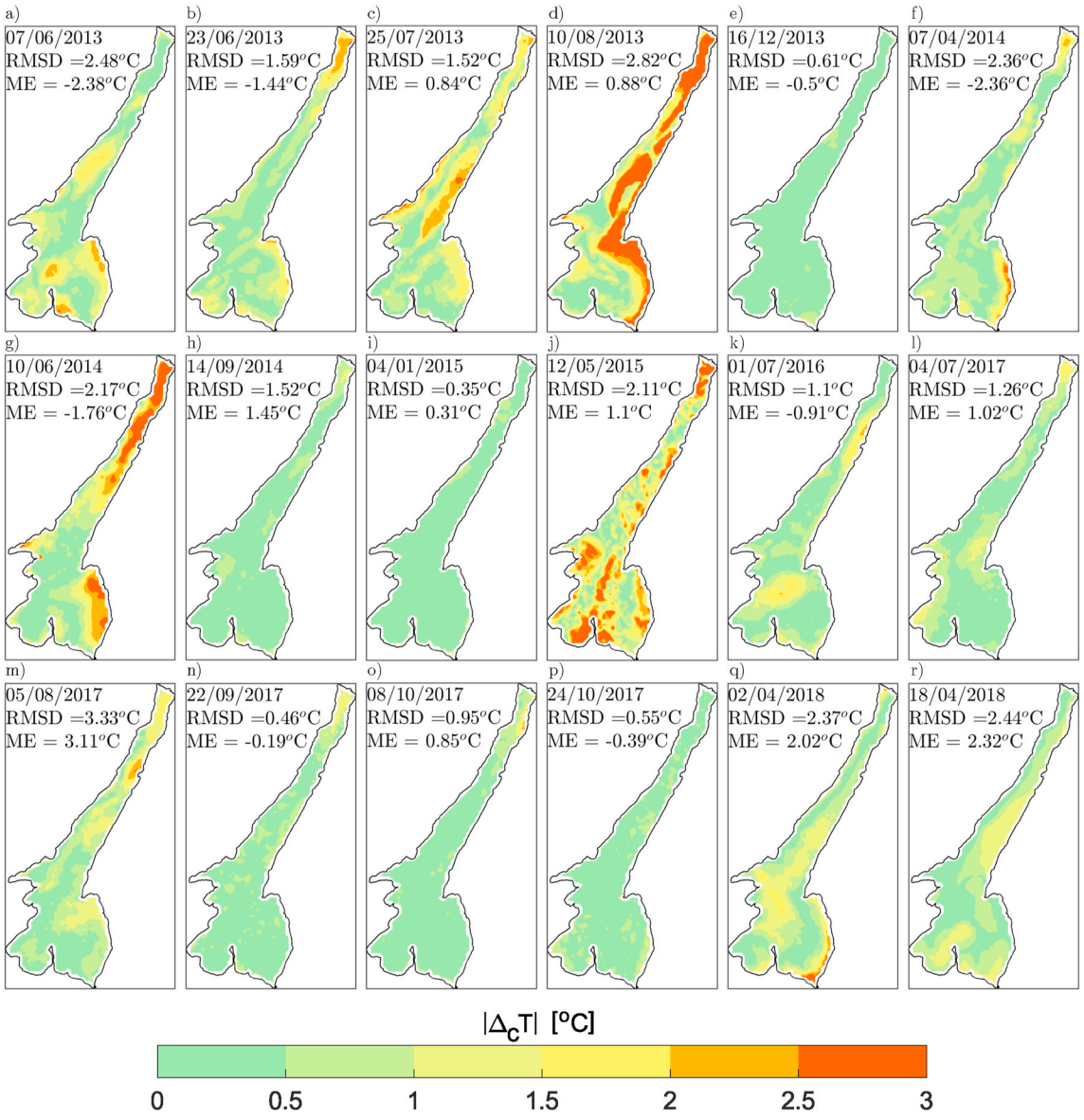

Fig. 10. Absolute unbiased temperature anomaly between simulated and Landsat-8 LSWT computed as in Eq. (3) from all available complete satellite maps. Each subplot reports the date (date format dd/mm/yyyy), the RMSD and the ME (bias). The two indices are computed as in Table 3, with *i* indicating each pixel and *N* the total number of pixels.

previous results available in the literature, both for the model we use, Delft3D (e.g., Wahl and Peeters, 2014; Soulignac et al., 2017; Dissanayake et al., 2019), and for similar models (e.g., Hodges et al., 2000; Rueda and Schladow, 2003; Valerio et al., 2017). Differently from the mentioned experiences, however, we evaluate not only the hydrodynamic model, but also the atmospheric model used to assign the surface boundary conditions. Additionally, both models are tested on different spatial and temporal scales.

The analysis of the interannual variability (Sect. 5.2) highlights that the hydrodynamic model correctly predicts the differences in the thermal regime of the two sub-basins of Lake Garda, i.e., oligomixis in the northern part, and monomixis in the southern part. However, the model reproduces two false complete overturns in the deepest part of the lake. Similar results are also found by previous experiences in large and deep Alpine lakes, and are attributed to excessive bottom mixing, either in 1D (e.g. Schwefel et al., 2016) or 3D models (e.g. Råman Vinnå et al., 2017). For the case of the deep Lake Garda, as in many other oligomictic lakes (Imboden et al., 1987), it is demonstrated that deep mixing is traditionally buoyancy-driven (Salmaso et al., 2017), but that intense and long-lasting wind events can occasionally cause the complete or partial entrainment of the water column, causing deep ventilation (Piccolroaz et al., 2019). Hence, a significant role is played by the atmospheric forcing (both air temperature and wind), which has to be captured properly in order not to miss deep mixing events or reproduce false positives. In our case, the interannual trend of the WRF performance at the northern station MET1 (Fig. 2a) shows a positive ME on wind and air temperature during wintertime. However, in some years an overestimation of wind speed of 1 or 2 m s^{-1} coincides with underestimated air temperatures (e.g., February 2009 and 2010, see Fig. 2), leading to complete overturns in the nearly unstratified period (late winter to early spring, see Fig. 4a). This issue can be overcome through data assimilation (Chen et al., 2019; Baracchini et al., 2020a) or restart procedures (Bouffard et al., 2018), forcing mixing to be well reproduced. In this work, where none of these techniques are applied and a single long simulation is performed, the results achieved in the deep water region are still acceptable, as such errors occurred only in two out of 14 simulated years.

In Sect. 5.3, we show that the lake model satisfactorily captures the

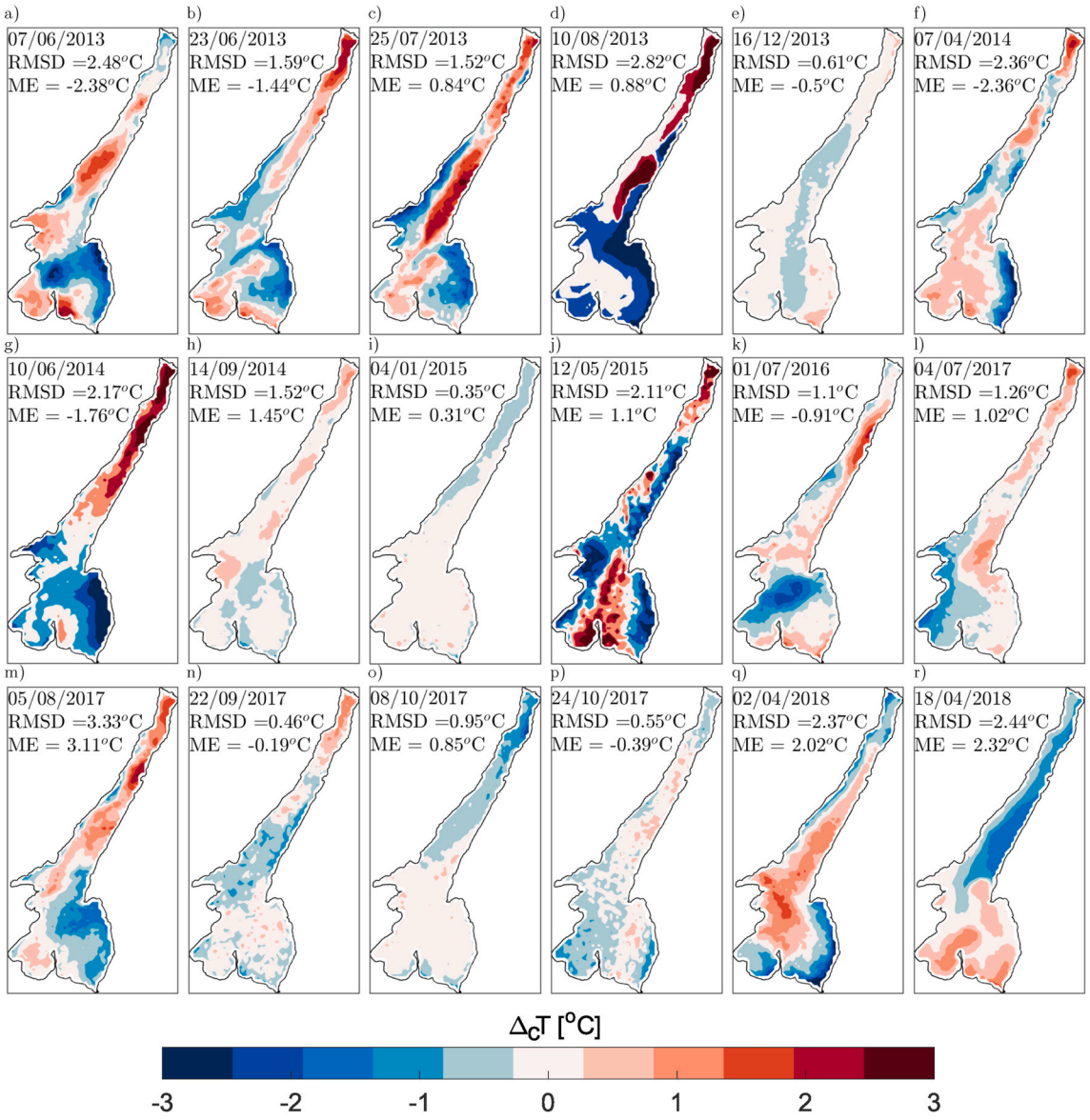

Fig. 11. The same plot as in Fig. 10, but keeping the sign of the unbiased temperature anomaly.

seasonal variations of water temperature along the vertical column. From this analysis, exceptional cooling events are easily identified. During such exceptional events, the hydrodynamic model provides the largest errors, especially in the northern region and on those days when the wind forcing is wrongly predicted by the atmospheric model. In this part of the lake, the resolution of the atmospheric model is not completely adequate to capture local-scale circulations in such a complex topography, since the grid size is 2 km over an average width of 4 km in the northern trunk. Limitations due to grid spacing are not the only source of error in the atmospheric model, since inaccuracies can also come for example from a misrepresentation of large-scale forcing (as in the case of the cold front on August 06, 2017, see Fig. 5). However, for the particular case of Lake Garda, where the climatology is mainly determined by local scale processes, due to the complex orography surrounding the lake, we believe that the misrepresentation of fine-scale details is the main cause of model errors, as highlighted in WRF similar applications over complex terrain (see e.g., Carvalho et al., 2012; Mughal et al., 2017). In this regard, it is well known that in very complex terrain sub-kilometer resolution should be adopted to reproduce appropriately local-scale circulations (Giovannini et al., 2014a; Le Roux et al., 2018; Schmidli et al., 2018). This resolution is not feasible for the present application, due to the high computational costs implied by simulating 14 years. Relevant inaccuracies occasionally occur also in the southern part of the lake. Here, the atmospheric model struggles in reproducing local-scale phenomena (e.g., again the thunderstorm on August 06, 2017, see Fig. 5) and occasionally causes large errors in the simulation of wind speed and/or air temperature over localized areas near the shore. However, it is likely that errors are lower over the lake's surface, where the spatial heterogeneity of the meteorological fields is expected to be less significant. Unfortunately, no long-term meteorological measurements are available over the water. In this regard, in the recent application on Lake Garda reported by Piccolroaz et al. (2019), the authors warned that wind speed above the lake is not well described by the measurements from the nearest ground stations. They found that the wind speed above the lake is on average 1.6 times stronger than land measurements at MET3 and that WRF results are representative of the lake forcing conditions in the investigated periods. It is not redundant to stress that the spatial variability of the wind forcing is crucial for an appropriate representation of the basin-scale circulation and transport patterns, as already demonstrated for Lake Garda by Amadori et al. (2018), and by e.g., Strub and Powell (1986), Lemmin and D'Adamo (1996) and Laval et al. (2005) for other lakes.

Going down to the shorter time scales (Sect. 5.4), the wavelet analysis of the higher time resolution data at MET1, MET5, APPA buoy and NIOZ points gives satisfactory results for both models. The analysis shows that both models capture the relevant time scales associated to wind velocity and temperature fluctuations (at the surface and at the bottom). In the series of epilimnetic water temperature, the predominant semi-diurnal and diurnal periodicity is perfectly reproduced by both atmospheric and hydrodynamic models. In the deep NIOZ point and the corresponding weather station MET5, the two models reproduce the main energetic levels, in good agreement with the observations by van Haren et al. (2020) at NIOZ point. However, both models overestimate the energy associated to each time scale. It is reasonable that the excessive energy predicted by WRF in MET5 is transferred to the hydrodynamic model. However, the error of the latter can also be magnified by the k-ε model implemented in Delft3D, which does not take into account wave breaking. Hence, the turbulence model might miss the dissipation associated with this process, eventually overestimating the energy associated with internal waves.

From Sect. 5.2 to Sect. 5.4, several time scales are considered, but also the spatial distribution of data. We see that the model performance is sensitive not only to the length of the series, but also to the geographic factors characterizing different points. In this regard, the case of the northern trunk of the lake is emblematic, as the worst performance is found in this area. We have already stressed that the resolution of the atmospheric model is too large to capture the sharp topographical variations of that region. However, its complexity also makes not only modeling, but also *in situ* measurements and remote sensing difficult to interpret. In this area residual effects from the adjacent land might affect the retrieval of LSWT from satellites in AVHRR, MODIS and Landsat-8 control points. In this regard, Pareeth et al. (2016) also observed that mountains have a shadowing effect on the remotely sensed LSWT in this part of the lake, which is likely to affect also *in situ* measurements, in a way that depends on the time of the day at which images are taken. The evaluation of the simulated water temperature fields against *in situ* and remotely sensed data (Sect. 5.5) provides useful insights on the model's capability of reproducing the spatial variations on single dates. The comparison of LSWT maps retrieved from satellite and those computed by the model provides the general picture that the model tends to overestimate LSWT in the northern part of the lake and to underestimate it in the south-eastern basin. If the issues related with the atmospheric model uncertainty are put aside for a moment, it is useful to consider that the two sub-basins are extremely different: the northern one is deep and narrow, with steep shores going sharply down to the bottom, while the south-eastern one is wide and shallow with a more uniform bathymetry. The way the lake exchanges heat through its surface is significantly different between the two sub-basins, with different thermal (and mixing) regimes, due to the lake heterogeneous bathymetry, shape and surrounding topography. Hence, the calibration parameters can produce opposite effects: not enough heat loss in the deep part (model too warm at the surface) and too much in the shallow part (model too cold). While currently available hydrodynamic models allow for various types of turbulence parameterization, from assigned eddy coefficients to more or less complex turbulence models (e.g., k-ε as in our case, or more recent approaches like LES, as proposed by Santo et al., 2017), one single parameterization of the heat fluxes across the lake surface is normally adopted for the whole domain. A more sophisticated parameterization of mixing on both sides (atmosphere and water) should be developed for applications on multi-basin lakes, in the same way as time-varying parameters are now obtained through data assimilation. This issue offers a lead for further research efforts in the future. Finally, the spatial distribution of anomalies of LSWT obtained from remote sensing compared to that from numerical modeling is also exploited to infer whether the hydrodynamic model is able to capture the surface transport dynamics in the lake. In such an analysis, we note some discrepancies between the two, but it is not easy to attribute them to malfunctioning of the model or misinterpretation of satellite data. Here, the timing for the comparison is essential: for instance, upwelling phenomena can be episodic and rapidly evanescent and a few hours of delay may make a large difference. If more maps at a higher time resolution are available, the transport patterns could be reconstructed (Steissberg et al., 2005) providing a precious way to verify also the flow field, at least at the lake surface. In this regard, the availability of additional remotely sensed measurements will definitely represent a key factor to improve this kind of analyses.

7. Conclusions

In this work, we evaluate the performance of a 3D lake model (Delft3D) forced by an atmospheric model (WRF) at multiple scales. We compare the results of a single long-term (14 years) simulation, with a number of different datasets of measurements from ground weather stations, buoys, thermistor chains and satellite imagery. While it is obvious that testing the numerical model against observations is the necessary step for any reliable application, in this work we critically analyze the model's performance at multiple scales with diverse types of available data.

Our focus is on the water temperature field, as water flow measurements are not always available in standard monitoring schemes. While a full validation of a 3D model would require to consider also other variables (e.g., water level, flow velocity, and turbulence quantities) here we cope with the typical scarcity of this type of data and extracted all possible information from water temperature data (typically available from routine monitoring of lakes) at different spatial and temporal scales. Even though some parameters are not calibrated, and the hydrodynamic quantities not verified directly, the 3D field of water temperature is satisfactorily reproduced at all investigated spatial and temporal scales. In fact, the model correctly reproduces the interannual variability of temperature along the water column, the main seasonal trends, the daily and subdaily internal variation and the main spatial patterns.

The approach we adopt allows highlighting how different scales interact, eventually affecting the results. On the one hand, we are able to interpret the hydrodynamic model flaws on single days, by observing the performance of the atmospheric model in the preceding months/years (e.g., in 2004 and 2005). On the other hand, we highlight how single events can affect the model's performance for many years afterwards (e.g., the error at the bottom from 2009 to 2013 caused by the wrong prediction of deep mixing event in 2009). When inaccuracies are found, the multi-purpose approach also allows for tracing back to error source either in the atmospheric model (interannual to daily scale) or in the model resolution (in the northern trunk region), or in the flow field uncertainty (largest patches of mismatch between satellite and model maps). In this regard, our results suggest that a better performance might be achieved by improving the resolution of the atmospheric model, ideally assimilating meteorological data from stations located within the lake. Progress in this direction is expected from ongoing research efforts, aiming at improving our capabilities of observing and simulating atmospheric transport and exchange process over complex terrain (Serafin et al., 2018). Our suggestion is to adopt such approach when a model is intended to be applied at an operational level, as it helps identifying the model strengths and weaknesses, and leads the way for further improvements.

Declaration of competing interest

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

Acknowledgments

We thank the EPAs of the Autonomous Province of Trento (APPA),

Veneto (ARPAV) and Lombardia Regions (ARPAL), and Edmund Mach Foundation (FEM), for kindly providing most of the observations used in this study. We are grateful to Circolo Vela Arco for logistic support during the IMAU-UniTrento field campaign, and to the firefigthers nautical rescue team of Trento (Corpo Permanente dei Vigili del Fuoco Trento) and M. van Haren for their assistance in the deployment and recovery of the NIOZ mooring. We also acknowledge Giuliano Morini, Bryan Brouwer and Bouke Biemond who contributed to the data processing, and Menno Genseberger (Deltares) for the initial setup of the Delft3D model of Lake Garda. The Delft3D simulations were carried out on the Cartesius supercomputer at SURFsara (www.surfsara.nl). Part of this work was supported by Utrecht University, by EU Horizon 2020 programme (EOMORES, grant agreement no. 730066; Water-ForCE, grant agreement no. 101004186) and by the WATER-Sat project (CNR DIT.012.115.001).

References

Amadori, M., Piccolroaz, S., Giovannini, L., Zardi, D., Toffolon, M., 2018. Wind variability and Earth's rotation as drivers of transport in a deep, elongated subalpine lake: the case of Lake Garda. J. Limnol. 77 (3) jlimnol.2018.1814.

Amadori, M., Morini, G., Piccolroaz, S., Toffolon, M., 2020. Involving citizens in hydrodynamic research: a combined local knowledge - numerical experiment on Lake Garda, Italy. Sci. Total Environ. 722, 137720. https://doi.org/10.1016/j.scitotenv.2020.137720.

Antenucci, J.P., Imberger, J., Saggio, A., 2000. Seasonal evolution of the basin-scale internal wave field in a large stratified lake. Limnol. Oceanogr. 45 (7), 1621–1638. https://doi.org/10.4319/lo.2000.45.7.1621.

Baracchini, T., Chu, P.Y., Šukys, J., Lieberherr, G., Wunderle, S., Wüest, A., Bouffard, D., 2020a. Data assimilation of in situ and satellite remote sensing data to 3d hydrodynamic lake models: a case study using Delft3D-FLOW v4.03 and OpenDA v2.4. Geosci. Model Dev. (GMD) 13 (3), 1267–1284. https://doi.org/10.5194/gmd-13-1267-2020.

Baracchini, T., Wüest, A., Bouffard, D. Meteolakes, 2020b. An operational online three-dimensional forecasting platform for lake hydrodynamics. Water Res. 172, 115529. https://doi.org/10.1016/j.watres.2020.115529.

Barsi, J.A., Schott, J.R., Palluconi, F.D., Hook, S.J., 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments. In: Butler, J.J. (Ed.), Earth Observing Systems X, ume 5882. International Society for Optics and Photonics, SPIE, pp. 136–142. https://doi.org/10.1117/12.619990.

Beletsky, D., Schwab, D., McCormick, M., 2006. Modeling the 19982003 summer circulation and thermal structure in Lake Michigan. J. Geophys. Res.: Oceans 111 (C10). https://doi.org/10.1029/2005JC003222.

Boegman, L., Ivey, G.N., Imberger, J., 2005. The energetics of large-scale internal wave degeneration in lakes. J. Fluid Mech. 531, 159180. https://doi.org/10.1017/S0022112005003915.

Bouffard, D., Kiefer, I., Wüest, A., Wunderle, S., Odermatt, D., 2018. Are surface temperature and chlorophyll in a large deep lake related? An analysis based on satellite observations in synergy with hydrodynamic modelling and in-situ data. Rem. Sens. Environ. 209, 510–523. https://doi.org/10.1016/j.rse.2018.02.056.

Bresciani, M., Giardino, C., Boschetti, L., 2011b. Multi-temporal assessment of bio-physical parameters in lakes Garda and Trasimeno from MODIS and MERIS. Ital. J. Remote Sens. 43 (3), 49–62. https://doi.org/10.5721/ItJRS20114334.

Bresciani, M., Cazzaniga, I., Austoni, M., Sforzi, T., Buzzi, F., Morabito, G., Giardino, C., 2018. Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. Hydrobiologia 824 (1), 197–214. https://doi.org/10.1007/s10750-017-3462-2.

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., Giardino, C., 2011a. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Sci. Total Environ. 409 (17), 3083–3091. https://doi.org/10.1016/j.scitotenv.2011.05.001.

Bresciani, M., Bolpagni, R., Braga, F., Oggioni, A., Giardino, C., 2012. Retrospective assessment of macrophytic communities in southern Lake Garda (Italy) from in situ and MIVIS (multispectral infrared and visible imaging spectrometer) data. J. Limnol. 71, 180–190. https://doi.org/10.4081/JLIMNOL.2012.E19.

Butcher, J.B., Nover, D., Johnson, T.E., Clark, C.M., Mar 2015. Sensitivity of lake thermal and mixing dynamics to climate change. Climatic Change 129 (1), 295–305. https://doi.org/10.1007/s10584-015-1326-1.

Carvalho, D., Rocha, A., Gmez-Gesteira, M., Santos, C., 2012. A sensitivity study of the WRF model in wind simulation for an area of high wind energy. Environ. Model. Software 33, 23–34. https://doi.org/10.1016/j.envsoft.2012.01.019.

Casulli, V., Pecenik, G., 1994. Modello idrodinamico tridimensionale del lago di Garda [Hydrodynamic model of Lake Garda]. In: Cimeca (Ed.), Scienza e Supercalcolo al Cimeca 67, 255–258.

Chen, C., Huang, J., Chen, Q., Zhang, J., Li, Z., Lin, Y., 2019. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using Ensemble Kalman Filter. Environ. Model. Software 117, 188–199. https://doi.org/10.1016/j.envsoft.2019.03.028.

Chen, F., Dudhia, J., 2001. Coupling an advances land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: model implementation and sensitivity. Mon. Weather Rev. 129, 569–585.

Dissanayake, P., Hofmann, H., Peeters, F., 2019. Comparison of results from two 3D hydrodynamic models with field data: internal seiches and horizontal currents. Inland Waters 9 (2), 239–260. https://doi.org/10.1080/20442041.2019.1580079.

Dudhia, J., 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. J. Atmos. Sci. 46, 3077–3107.

Fang, X., Stefan, H.G., 2009. Simulations of climate effects on water temperature, dissolved oxygen, and ice and snow covers in lakes of the contiguous U.S. under past and future climate scenarios. Limnol. Oceanogr. 54 (6part2), 2359–2370. https://doi.org/10.4319/lo.2009.54.6\-part\-2.2359.

Ghirardi, N., Amadori, M., Free, G., Giovannini, L., Toffolon, M., Giardino, C., Bresciani, M., Oct. 2020. Using remote sensing and numerical modelling to quantify a turbidity discharge event in lake garda. J. Limnol. 80 https://doi.org/10.4081/jlimnol.2020.1981.

Giardino, C., Brando, V.E., Dekker, A.G., Strömbeck, N., Candiani, G., 2007. Assessment of water quality in lake garda (Italy) using hyperion. Rem. Sens. Environ. 109, 183–195. https://doi.org/10.1016/j.rse.2006.12.017.

Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., Brando, V., 2014. Evaluation of multi-resolution satellite sensors for assessing water quality and bottom depth of Lake Garda. Sensors 14 (12), 24116–24131. https://doi.org/10.3390/s141224116.

Gill, A.E., 1982. Atmosphere-ocean Dynamics. Academic Press, New York.

Giovannini, L., Antonacci, G., Zardi, D., Laiti, L., Panziera, L., 2014a. Sensitivity of simulated wind speed to spatial resolution over complex terrain. Energy Procedia 59, 323–329. https://doi.org/10.1016/j.egypro.2014.10.384.

Giovannini, L., Laiti, L., Zardi, D., de Franceschi, M., 2015. Climatological characteristics of the Ora del Garda wind in the Alps. Int.l J. Climatol. 35, 4103–4115. https://doi.org/10.1029/2001JC000954.

Giovannini, L., Laiti, L., Serafin, S., Zardi, D., 2017. The thermally driven diurnal wind system of the Adige Valley in the Italian Alps. Q. J. R. Meteorol. Soc. 143 (707), 2389–2402. https://doi.org/10.1002/qj.3092.

Giovannini, L., Zardi, D., De Franceschi, M., Chen, F., 2014b. Numerical simulations of boundary-layer processes and urban-induced alterations in an Alpine valley. Int. J. Climatol. 34 https://doi.org/10.1002/joc.3750, 03.

Grossmann, A., Morlet, J., 1984. Decomposition of hardy functions into Square integrable wavelets of constant shape. SIAM J. Math. Anal. 15, 723–736. https://doi.org/10.1137/0515056.

Guanter, L., Ruiz-Verdu, A., Odermatt, D., Giardino, C., Simis, S., Estelles, V., Heege, T., Antonio Dominguez-Gomez, J., Moreno, J., 2010. Atmospheric correction of ENVISAT/MERIS data over inland waters: validation for European lakes. Rem. Sens. Environ. 114 (3), 467–480.

Guyennon, N., Valerio, G., Salerno, F., Pilotti, M., Tartari, G., Copetti, D., 2014. Internal wave weather heterogeneity in a deep multi-basin subalpine lake resulting from wavelet transform and numerical analysis. Adv. Water Resour. 71, 149–161. https://doi.org/10.1016/j.advwatres.2014.06.013.

Hipsey, M.R., Gal, G., Arhonditsis, G.B., Carey, C.C., Elliott, J.A., Frassl, M.A., Janse, J. H., de Mora, L., Robson, B.J., 2020. A system of metrics for the assessment and improvement of aquatic ecosystem models. Environ. Model. Software 128, 104697. https://doi.org/10.1016/j.envsoft.2020.104697.

Hodges, B.R., Imberger, J., Saggio, A., Winters, K.B., 2000. Modeling basin-scale internal waves in a stratified lake. Limnol. Oceanogr. 45 (7), 1603–1620. https://doi.org/10.4319/lo.2000.45.7.1603.

Hong, S.-Y., Dudhia, J., Chen, S.-H., 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Mon. Weather Rev. 132, 103–120.

Hong, S.-Y., Noh, Y., Dudhia, J., 2006. A new vertical diffusion package with an explicit treatment of entrainment processes. Mon. Weather Rev. 134, 2318–2341.

Horvath, K., Koracin, D., Vellore, R., Jiang, J., Belu, R., 2012. Subkilometer dynamical downscaling of nearsurface winds in complex terrain using WRF and MM5 mesoscale models. J. Geophys. Res.: Atmosphere 117. https://doi.org/10.1029/2012JD017432.

Imboden, D., Stotz, B., Wuest, A., 1987. Hypolimnic mixing in a deep alpine lake and the role of a storm event: with 5 figures and 1 table in the text. SIL Proceedings 23, 67–73. https://doi.org/10.1080/03680770.1987.11897904, 1922-2010.

Jiménez, P.A., Dudhia, J., 2012. Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. Journal of Applied Meteorology and Climatology 51, 300–316.

Kain, J.S., Fritsch, J.M., 1993. Convective Parameterization for Mesoscale Models: the KainFritcsh Scheme. *The Representation of Cumulus Convection in Numerical Models*.

Kelley, J.G.W., Hobgood, J.S., Bedford, K.W., Schwab, D.J., 09 1998. Generation of three-dimensional lake model forecasts for Lake erie. Weather Forecast. 13 (3), 659–687. https://doi.org/10.1175/1520-0434(1998)013<0659:GOTDLM>2.0.CO;2.

Kirillin, G., 2010. Modeling the impact of global warming on water temperature and seasonal mixing regimes in small temperate lakes. Boreal Environ. Res. 15, 279–293.

Laiti, L., Zardi, D., de Franceschi, M., Rampanelli, G., 2013a. Atmospheric boundary-layer structures associated with the Ora del Garda wind in the Alps as revealed from airborne and surface measurements. Atmos. Res. 132–133, 473–489. https://doi.org/10.1016/j.atmosres.2013.07.006.

Laiti, L., Zardi, D., de Franceschi, M., Rampanelli, G., 2013b. Residual Kriging analysis of airborne measurements: application to the mapping of Atmospheric Boundary-Layer thermal structures in a mountain valley. Atmos. Sci. Lett. 14 (2), 79–85. https://doi.org/10.1002/asl2.420.

Laiti, L., Zardi, D., de Franceschi, M., Rampanelli, G., Giovannini, L., 2014. Analysis of the diurnal development of a lake-valley circulation in the Alps based on airborne and surface measurements. Atmos. Chem. Phys. 14, 9771–9786. https://doi.org/10.5194/acp-14-9771-2014.

Laval, B.E., Imberger, J., Findikakis, A.N., 2005. Dynamics of a large tropical lake: lake Maracaibo. Aquat. Sci. 67, 337–349. https://doi.org/10.1007/s00027-005-0778-1.
Laval, B., Imberger, J., Hodges, B.R., Stocker, R., 2003. Modeling circulation in lakes: spatial and temporal variations. Limnol. Oceanogr. 48 (3), 983–994. https://doi.org/10.4319/lo.2003.48.3.0983.
Le Roux, R., Katurji, M., Zawar-Reza, P., Qunol, H., Sturman, A., 2018. Comparison of statistical and dynamical downscaling results from the WRF model. Environ. Model. Software 100, 67–73. https://doi.org/10.1016/j.envsoft.2017.11.002.
Lemmin, U., D'Adamo, N., 1996. Summertime winds and direct cyclonic circulation: observations from Lake Geneva. Ann. Geophys. 14, 1207–1220. https://doi.org/10.1007/s00585-996-1207-z.
Lenstra, W.K., Hahn-Woernle, L., Matta, E., Bresciani, M., Giardino, C., Salmaso, N., Musanti, M., Fila, G., Uittenbogaard, R., Genseberger, M., van der Woerd, H.J., Dijkstra, H.A., 2014. Diurnal variation of turbulence-related quantities in Lake Garda. Adv. Oceanogr. Limnol. 5 (2), 184–203. https://doi.org/10.1080/19475721.2014.971870.
Leon, L.F., Lamb, D.C.L., Schertzer, W.M., Swayne, D.A., Imberger, J., 2007. Towards coupling a 3D hydrodynamic lake model with the Canadian regional climate model: simulation on great slave lake. Environ. Model. Software 22 (6), 787–796. https://doi.org/10.1016/j.envsoft.2006.03.005.
Lesser, G.R., Roelvink, J.A., Keste, T.M.V., Stelling, G.S., 2004. Development and validation of a three-dimensional morphological model. Coast. Eng. 51, 883–915. https://doi.org/10.1016/j.coastaleng.2004.07.014.
Lovato, T., Pecenik, G., 2012. Three-dimensional Modeling of Pollutant Dispersion in Lake Garda (North Italy). Chapter 19 in:Models of the Ecological Hierarchy: from Molecules to the Ecosphere, vol. 25. Elsevier, Amsterdam, The Netherlands, p. 562.
Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. 102 (D14), 16663–16682.
Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50. https://doi.org/10.13031/2013.23153, 05.
Mughal, M.O., Lynch, M., Yu, F., McGann, B., Jeanneret, F., Sutton, J., 2017. Wind modelling, validation and sensitivity study using weather research and forecasting model in complex terrain. Environ. Model. Software 90, 107–125. https://doi.org/10.1016/j.envsoft.2017.01.009, 1364–8152.
Nicolantonio, W.D., Cazzaniga, I., Cacciari, A., Bresciani, M., Giardino, C., 2015. Synergy of multispectral and multisensors satellite observations to evaluate desert aerosol transport and impact of dust deposition on inland waters: study case of Lake Garda. J. Appl. Remote Sens. 9 (1), 1–19. https://doi.org/10.1117/1.JRS.9.095980.
Odermatt, D., Giardino, C., Heege, T., 03 2010. Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. Rem. Sens. Environ. 607–617. https://doi.org/10.1016/j.rse.2009.10.016.
Oke, T.R., 1987. Boundary Layer Climates. Routledge London.
Okubo, A., 1971. Oceanic diffusion diagrams. Deep Sea Res. Oceanogr. Abstr. 18 (8), 789–802.
Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the Earth Sciences. Science 263 (5147), 641–646. https://doi.org/10.1126/science.263.5147.641.
Pan, H., Avissar, R., Haidvoge, D.B., 2002. Summer circulation and temperature structure of Lake Kinneret. J. Phys. Oceanogr. 32 (1), 295–313, 1520-0485(2002)032<0295:SCATSO>2.0.CO;2.
Pareeth, S., Salmaso, N., Adrian, R., Neteler, M., 2016. Homogenised daily lake surface water temperature data generated from multiple satellite sensors: a long-term case study of a large sub-Alpine lake. Sci. Rep. 6, 31251. https://doi.org/10.1038/srep31251.
Perroud, M., Goyette, S., 2010. Impact of warmer climate on Lake Geneva water-temperature profiles. Boreal Environ. Res. 15, 04.
Piccolroaz, S., Toffolon, M., 2013. Deep water renewal in Lake Baikal: a model for long term analyses. J. Geophys. Res. 118 (12), 6717–6733. https://doi.org/10.1002/2013JC009029.
Piccolroaz, S., Toffolon, M., 2018. The fate of Lake Baikal: how climate change may alter deep ventilation in the largest lake on Earth. Climatic Change 150, 181–194. https://doi.org/10.1007/s10584-018-2275-2.
Piccolroaz, S., Amadori, M., Toffolon, M., Dijkstra, H.A., 2019. Importance of planetary rotation for ventilation processes in deep elongated lakes: evidence from Lake Garda (Italy). Sci. Rep. 9, 2045–2322. https://doi.org/10.1038/s41598-019-44730-1.
Pilotti, M., Valerio, G., Leoni, B., 2013. Data set for hydrodynamic lake model calibration: a deep prealpine case. Water Resour. Res. 49 (10), 7159–7163. https://doi.org/10.1002/wrcr.20506.
Prats, J., Reynaud, N., Rebière, D., Peroux, T., Tormos, T., Danis, P.-A., 2018. LakeSST: lake Skin Surface Temperature in French inland water bodies for 1999–2016 from landsat archives. Earth Syst. Sci. Data 10 (2), 727–743. https://doi.org/10.5194/essd-10-727-2018.
Preusse, M., Peeters, F., Lorke, A., 11 2010. Internal waves and the generation of turbulence in the thermocline of a large lake. Limnol. Oceanogr. 55, 2353–2365. https://doi.org/10.4319/lo.2010.55.6.2353.
Råman Vinnå, L., Wüest, A., Bouffard, D., 2017. Physical effects of thermal pollution in lakes. Water Resour. Res. 53 (5), 3968–3987. https://doi.org/10.1002/2016WR019686.
Razmi, A.M., Barry, D.A., Bakhtyar, R., Wüest, A., 2013. Current variability in a wide and open lacustrine embayment in Lake Geneva (Switzerland). J. Great Lake. Res. 39 (3), 455–465. https://doi.org/10.1016/j.jglr.2013.06.011.
Riffler, M., Lieberherr, G., Wunderle, S., 2015. Lake surface water temperatures of European Alpine lakes (1989-2013) based on the Advanced Very High Resolution Radiometer (AVHRR) 1 km data set. Earth Syst. Sci. Data 7 (1), 1–17. https://doi.org/10.5194/essd-7-1-2015.
Rueda, F.J., Schladow, S.G., 2003. Dynamics of large polymictic lake. II: numerical simulations. J. Hydraul. Eng. 129 (2), 92–101. https://doi.org/10.1061/(ASCE)0733-9429(2003)129:2(92).
Salmaso, N., Boscaini, A., Capelli, C., Cerasino, L., 2017. Ongoing ecological shifts in a large lake are driven by climate change and eutrophication: evidences from a three-decade study in Lake Garda. Hydrobiologia 11. https://doi.org/10.1007/s10750-017-3402-1.
Santo, M., Toffolon, M., Zanier, G., Giovannini, L., Armenio, V., 2017. Large Eddy Simulation (LES) of wind-driven circulation in a peri-alpine lake: detection of turbulent structures and implications of a complex surrounding orography. J. Geophys. Res. https://doi.org/10.1002/2016JC012284.
Schlabing, D., Frassl, M.A., Eder, M.M., Rinke, K., Brdossy, A., 2014. Use of a weather generator for simulating climate change effects on ecosystems: a case study on lake constance. Environ. Model. Software 61, 326–338. https://doi.org/10.1016/j.envsoft.2014.06.028, 1364–8152.
Schmidli, J., Böoing, S., Fuhrer, O., 2018. Accuracy of simulated diurnal valley winds in the Swiss Alps: influence of grid resolution, topography filtering, and land surface datasets. Atmosphere 196 (9).
Schwefel, R., Gaudard, A., Wüest, A., Bouffard, D., 2016. Effects of climate change on deepwater oxygen and winter mixing in a deep lake (Lake Geneva): comparing observational findings and modeling. Water Resour. Res. 52 (11), 8811–8826. https://doi.org/10.1002/2016WR019194.
Serafin, S., Adler, B., Cuxart, J., De Wekker, S.F.J., Gohm, A., Grisogono, B., Kalthoff, N., Kirshbaum, D.J., Rotach, M.W., Schmidli, J., Stiperski, I., Vecenaj, v., Zardi, D., 2018. Exchange processes in the atmospheric boundary layer over mountainous terrain. Atmosphere 9 (3). https://doi.org/10.3390/atmos9030102.
Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X.Y., Wang, W., Powers, J.G., 2008. A description of the advanced research WRF version 3. NCAR Technical Note TN-475+STR 125.
Song, Y., Semazzi, F.H.M., Xie, L., Ogallo, L.J., 2004. A coupled regional climate model for the Lake Victoria basin of East Africa. Int. J. Climatol. 24 (1), 57–75. https://doi.org/10.1002/joc.983.
Soulignac, F., Vinçon-Leite, B., Lemaire, B.J., Scarati Martins, J.R., Bonhomme, C., Dubois, P., Mezemate, Y., Tchiguirinskaia, I., Schertzer, D., Tassin, B., Aug 2017. Performance assessment of a 3D hydrodynamic model using high temporal resolution measurements in a shallow urban lake. Environ. Model. Assess. 22 (4), 309–322. https://doi.org/10.1007/s10666-017-9548-4.
Steissberg, T.E., Hook, S.J., Schladow, S.G., 2005. Measuring surface currents in lakes with high spatial resolution thermal infrared imagery. Geophys. Res. Lett. 32 (11), 1–4. Cited By :13.
Strub, P.T., Powell, T.M., 1986. Wind-driven surface transport in stratified closed basins: direct versus residual circulations. J. Geophys. Res. 91 (C7), 8497–8508. https://doi.org/10.1029/JC091iC07p08497.
Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res.: Atmosphere 106 (D7), 7183–7192. https://doi.org/10.1029/2000JD900719.
Toffolon, M., Piccolroaz, S., Dijkstra, H.A., 2017. A plunge into the depths of Italy's Lake Garda. Eos 98. https://doi.org/10.1029/2017EO074499.
Valerio, G., Cantelli, A., Monti, P., Leuzzi, G., 2017. A modeling approach to identify the effective forcing exerted by wind on a prealpine lake surrounded by a complex topography. Water Resour. Res. https://doi.org/10.1002/2016WR020335.
Valerio, G., Pilotti, M., Barontini, S., Leoni, B., 2015. Sensitivity of the multiannual thermal dynamics of a deep pre-alpine lake to climatic change. Hydrol. Process. 29 (5), 767–779. https://doi.org/10.1002/hyp.10183.
van Haren, H., 2018. Philosophy and application of high-resolution temperature sensors for stratified waters. Sensors 18 (10). https://doi.org/10.3390/s18103184.
van Haren, H., Piccolroaz, S., Amadori, M., Toffolon, M., Dijkstra, H.A., 2020. Moored observations of turbulent mixing events in deep Lake Garda, Italy. *Journal of Limnology*, (AOP), Nov. https://doi.org/10.4081/jlimnol.2020.1983.
Wahl, B., Peeters, F., 2014. Effect of climatic changes on stratification and deep-water renewal in Lake Constance assessed by sensitivity studies with a 3D hydrodynamic model. Limnol. Oceanogr. 59 (3), 1035–1052. https://doi.org/10.4319/lo.2014.59.3.1035.
Wang, X., Zhang, J., Babovic, V., Gin, K., 2019. A comprehensive integrated catchment-scale monitoring and modelling approach for facilitating management of water quality. Environ. Model. Software 120, 104489. https://doi.org/10.1016/j.envsoft.2019.07.014.
Wilson, R.C., Hook, S.J., Schneider, P., Schladow, S.G., 2013. Skin and bulk temperature difference at Lake Tahoe: a case study on lake skin effect. J. Geophys. Res.: Atmosphere 118 (18), 10,332–10,346. https://doi.org/10.1002/jgrd.50786.
Woolway, R.I., Maberly, S.C., Jones, I.D., Feuchtmayr, H., 2014. A novel method for estimating the onset of thermal stratification in lakes from surface water measurements. Water Resour. Res. 50 (6), 5131–5140. https://doi.org/10.1002/2013WR014975.
Wüest, A., Lorke, A., 2003. Small scale hydrodynamics in lakes. Annu. Rev. Fluid Mech. 35, 373–412. https://doi.org/10.1146/annurev.fluid.35.101101.161220.
Xue, P., Pal, J.S., Ye, X., Lenters, J.D., Huang, C., Chu, P.Y., 2017. Improving the simulation of large lakes in regional climate modeling: two-way lake-atmosphere

coupling with a 3-D hydrodynamic model of the Great Lakes. J. Clim. https://doi.org/10.1175/JCLI-D-16-0225.1.

Yamashiki, Y., Kato, M., Takara, K., Nakakita, E., Kumagai, M., Jiao, C., 2010. Sensitivity analysis on Lake biwa under the A1B SRES climate change scenario using biwa-3D integrated assessment model. Hydrological Research Letters 4, 45–49. https://doi.org/10.3178/hrl.4.45.

Zilioli, E., Brivio, P., Gomarasca, M., 1994. A correlation between optical properties from satellite data and some indicators of eutrophication in Lake Garda (Italy). Sci. Total Environ. 158, 127–133. https://doi.org/10.1016/0048-9697(94)90052-3.