Novel Air2water Model Variant for Lake Surface Temperature Modeling With Detailed Analysis of Calibration Methods

Adam P. Piotrowski, Jaroslaw J. Napiorkowski [ID], and Senlin Zhu [ID]

Abstract—The air2water model is a simple and efficient tool for modeling surface water temperature in lakes based solely on the air temperature. In this article, we propose to modify the air2water model in such a way that different parameters would be associated with lake stratification of cold waters than with lake stratification of warm waters. The situation of a mix of both cold water and warm water is also considered. The model is tested on 22 lowland Polish lakes against two classical air2water variants. As the new air2water model variant is slightly more complicated than the basic versions, we focus on the importance of the choice of the calibration method. Each variant of the air2water model is calibrated with eight different optimization methods, which are also compared on various benchmark problems. We show that the proposed variant is superior to the classical air2water models on about 90% of tested lakes, but only if the calibration approach is properly selected, which confirms the importance of the links between the model and appropriate optimization procedures. The proposed air2water variant performs well on various lowland lakes, with exception of large but shallow ones, probably due to the weak stratification of the shallow lakes.

Index Terms—Air2water, evolutionary computation, hydrological modeling, lake surface temperatures, limnology, optimization.

I. INTRODUCTION

THE problem of modeling the lake surface water temperature based on the air temperature has been widely discussed in the literature [1], [2], [3]. As water temperature in lakes affects the presence of important gases such as oxygen [4], [5] and impacts the fate of aquatic biota [6], [7], the problem of water temperature modeling is important for both biodiversity [8], [9] and lake water management [10], [11]. Thermal modeling is becoming even more critical during climatic change [12], [13].

The majority of lake water temperature models are either physical-based [14], [15] or empirical [16], [17], [18], [19]. However, recently, some kinds of physically based empirical models have been proposed. One group of such approaches are physics-guided neural networks [6], which have a large number of parameters, another is the air2water model that in its basic version has only eight parameters that represent physical and thermal processes that occur between air, land, and the lake surface [3], [20]. Air2water model has found numerous applications in hydrology and limnology [18], [21], [22], [23], [24]. In general, the air2water model aims at modeling water temperature within some depth of a lake's surface layer. However, it is very difficult to define the bounds of this surface layer [24]. Recently Guo et al. [25] introduced a variant of the air2water model with 15 parameters developed for the thermal surface conditions of frozen lakes. Based on the idea introduced in [25], in this article, we propose a new version of the air2water model that has 12 parameters but is much more flexible in capturing interrelationships between a cold or frozen lake surface and the water body. The variant uses different parameterizations for cold and warm water periods (basically, those with classical thermal stratification and inversion of water temperatures), and considers a situation of a mix of both of them.

In this article, we present a modified version of the air2water model and analyze how the modeling quality will be affected by the choice of the optimizer. Both issues are interrelated, as although the air2water model is physically based, the physical processes in air2water are parameterized, and such parameters require calibration. Simpler optimizers may be unable to show the advantage of the new version of the air2water model over its older counterparts, but more advanced optimization procedures may be able to do so. Hence, proposing a new variant of a model should be associated with a discussion on the appropriate calibration method. The question regarding the relationship between the model and its optimizer may be addressed from a philosophical [26] or empirical point of view (e.g., [27]), but is asked again and again in various fields of science. It is to some extent similar to the problem of the choice of the appropriate machine learning approach for the particular data [19].

Although the literature is full of metaheuristics [28], [29], [30], [31], [32], according to no free lunch theorems, the averaged performance of all possible problems would be equal [33]. Nonetheless, the performances of metaheuristics do differ for the specific problems, and may highly affect applications to such

practical problems like the Internet of Things [34], COVID-19 pandemic [35], robotics [36], photovoltaic systems [37], and various hydrological [38], planetary [39], or remote sensing [40] problems. The question of how to choose an appropriate metaheuristic for a particular application is old and ubiquitous [41], [42], [43], [44], [45]. Various issues may affect this choice. Some of them are related to the flexibility of the specific algorithm, like balancing exploration and exploitation during search [46], [47], [48], [49], fitness-landscape analysis [45], [50], [51], the approach to control parameter adaptation [52], [53], [54], [55], [56], [57], or specification of the population size [58], [59], [60], [61]. The other factors are related to the comparison settings that are set by the user, like the assumed maximum number of allowed function calls [62], [63], the specific benchmarks or criteria based on which the choice is being made [64], [65], [66], the statistical tests used in the evaluation [67], [68], or the nature of the decision space [69], [70]. In choosing the algorithm for the specific application one may even consider how well the particular optimizer matches the source of inspiration (e.g., [71]), or how much inspiration-independent it is (e.g., [72], [73], and [74]). However, there is no general guide for "practitioners," how to choose the "right" metaheuristic (e.g., [65] and [75]).

In the present article, we compare the performance of the new air2water model variant against its two older counterparts [3], [20] on 30-year-long daily air temperatures and water temperatures of surface layers in 22 lakes that are located in the lowland part of Poland. Each air2water variant is trained with eight selected optimization algorithms that are discussed in Section III-A. Before the algorithms are applied to our problem of interest, we test them on benchmarks and real-world problems from the literature to determine which among eight metaheuristics seems to be the "best". To do so, we define a few criteria for comparing optimization algorithms and test the eight metaheuristics on 30 IEEE CEC2017 benchmarks [76] and 22 real-world problems from different fields of sciences introduced in IEEE CEC2011 [77]. We discuss the performance of each optimizer and analyze the ranking of metaheuristics based on particular criteria and benchmarks. We point at algorithms that seem most appropriate based on artificial benchmarks or various real-world problems—following a roughly similar way as in [78]. After that, we get back to the problem of modeling water temperatures in lakes and use each among eight metaheuristics for calibration of the three air2water model variants. We fit parameters of the air2water model variants to observational data from the calibration set that is composed of roughly 20 years of daily measurements; the remaining 10 years of daily data is kept aside as the validation set. Based on the results obtained for calibration and validation sets, for each metaheuristic separately we choose the best variant of the air2water model and discuss the impact of the optimizer on that choice. Finally, we compare the performances obtained by different air2water variants trained by different metaheuristics and point out to what degree the choice of the best metaheuristics based on artificial benchmarks or real-world problems not related to our problem of interest were useful in determining the most appropriate method for calibration of the air2water model.

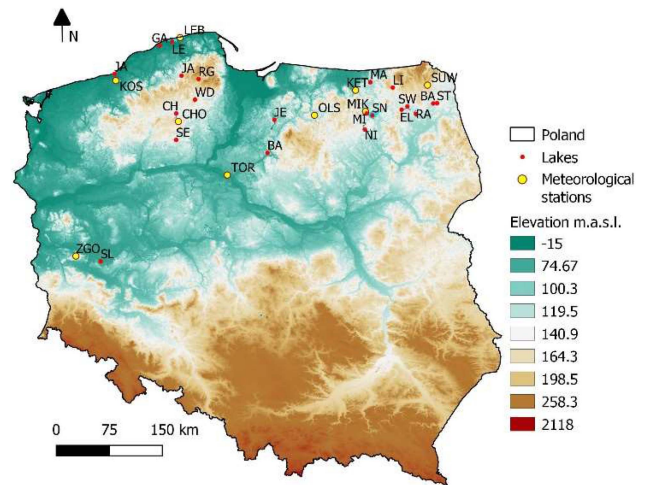To summarize, there are three main goals of the present study:

Fig. 1. Locations of the studied lakes and the corresponding meteorological stations. The numbers and characters are corresponding to that in Table I.

1) to improve the air2water model for lake water temperature modeling, by proposing a new variant;
2) to analyze the impact of existing optimization algorithms on the performance of air2water variants – both the new one and two classical versions from [20];
3) to find out if one could pick up an appropriate optimization algorithm for the specific application—air2water model calibration in our case—based on the results obtained by optimizers on classical benchmark sets: either numerical benchmarks [76] or collections of real-world problems [77].

II. AIR2WATER MODEL VARIANTS AND DATA

A. Data

In this article, 30 years long daily data (1987–2016) from 22 lowland lakes located in the northern and western part of Poland and from nine meteorological stations located in the nearby areas are used (see Table I and Fig. 1). Data from the years 1987–2006 are used for optimization, the remaining 2007–2016 data are left for the model validation. In the case of Lake Bachotek, the data set ends on November 30, 2016, and in the case of Lake Slawskie, the data ends on December 31, 2014; in both cases, the validation periods are shorter, but calibration periods are the same as for the other lakes.

In each lake, the water temperature was measured once a day at 6:00 GMT at a water depth of 0.4 m below the water surface, or below the ice cover during winter months. All lakes are dimictic, which means that the difference in temperature between the surface and bottom layers becomes negligible twice per year (about 4 °C), allowing all strata of the lake's water to circulate vertically. During the year the surface may have water temperatures both lower and higher than 4 °C—this is important for the mixing among lake layers, as 4 °C is the temperature in which the water is the densest.

TABLE I
INFORMATION ABOUT THE 22 LAKES FROM LOWLAND POLAND

No.	Lake name	Depth (m)		Area (km²)	Altitude (masl)	Volume (10^6 m³)	Meteorological station
		mean	max				
1	Sławskie	5.2	12.3	8.23	56.9	42.66	Zielona Góra (A)
2	Jamno	1.4	3.9	22.32	0.1	31.53	Koszalin (B)
3	Gardno	1.3	2.6	23.38	0.3	30.95	Łeba (C)
4	Łebsko	1.6	6.3	70.20	0.3	117.52	Łeba
5	Jasień	8.3	32.2	5.75	112.7	26.10	Chojnice (D)
6	Raduńskie Górne	15.5	43	3.63	161.6	60.16	Chojnice
7	Wdzydze	15.5	69.5	14.17	134	220.80	Chojnice
8	Charzykowskie	9.8	30.5	13.36	120	134.53	Chojnice
9	Sępoleńskie	4.8	10.9	1.58	112.8	7.50	Chojnice
10	Bachotek	7.2	24.3	2.15	70.8	15.39	Toruń (E)
11	Jeziorak	4.1	12.9	31.53	99.2	141.59	Olsztyn (F)
12	Dadaj	12	39.8	9.75	122.5	120.78	Olsztyn
13	Mikołajskie	11.2	25.9	4.24	115.7	55.74	Mikołajki (G)
14	Śniardwy	5.8	23.4	114.88	115.7	660.21	Mikołajki
15	Nidzkie	6.2	23.7	17.5	117.9	113.87	Mikołajki
16	Mamry	9.8	43.8	98.51	115.8	1003.37	Kętrzyn (H)
17	Ełckie	15	55.8	3.85	119.9	57.42	Mikołajki
18	Selmęt Wielki	7.8	21.9	12.08	120.7	9.96	Mikołajki
19	Rajgrodzkie	9.4	52	14.99	118.5	142.62	Suwałki (I)
20	Litygajno	6	16.4	1.55	132.8	9.76	Suwałki
21	Białe Augustowskie	8.7	30.0	4.53	122.2	41.72	Suwałki
22	Studzienicze	8.7	30.5	2.44	123.4	22.07	Suwałki

B. Air2water Model With Eight Parameters

The air2water model, which has been proposed by Piccolroaz et al. [20] and Toffolon et al. [3], is a simple physically inspired model that relates the surface water temperature in lakes only to the air temperature. The model is based on a volume-integrated heat balance equation with linearized heat flux terms. The relation between air and lake water temperatures is expressed as follows [3], [20]:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left(a_1 + a_2 T_a - a_3 T_w + a_5 \cos\left(2\pi\left(\frac{t}{t_y} - a_6\right)\right)\right) \tag{1}$$

where δ is defined as

$$\delta = \exp\left(-\frac{T_w - T_h}{a_4}\right) \text{ if } T_w \geq T_h$$

$$\delta = \exp\left(-\frac{T_h - T_w}{a_7}\right) + \exp\left(-\frac{T_w}{a_8}\right) \text{ if } T_w < T_h. \tag{2}$$

In (1), t is the day of the year and t_y is the number of days in a year. In (1) and (2), T_w is the lake water temperature in the surface layer, T_a is the air temperature, δ is a dimensionless parameter representing the ratio between the volume of the surface lake layer and a reference volume, T_h is a reference value for the deep lake water temperature, and a_1–a_8 are eight model parameters that are to be calibrated for the particular lake. In the

original air2water model with eight parameters, the reference value T_h is fixed depending on the lake type. In the case of Polish lakes that are dimictic, it is set to 4 °C. For warm or cold monomictic lakes, it is set to the minimum or the maximum water temperature, respectively. In the version of the air2water model proposed in [79], T_h is calibrated together with the remaining eight parameters, instead of being fixed. In the present study, we keep the value fixed to 4 °C, as proposed for dimictic lakes in the original air2water model [20].

The air2water model is solved using the 4th-order Runge–Kutta method, and water temperatures below 0 °C are set as 0 °C during computations. To limit the impact of the initial conditions on the results, the first 30 days of each data set are considered a warm-up period and are not used to evaluate the modeling performance. As the basic model has eight parameters, we call it a2w8 in the future part of the text.

C. Air2water Model With Six Parameters

As may be seen from (2), the computation depends on whether the reference value for the deep lake water temperature is higher or lower than the water temperature in the surface lake layer. As discussed in [3] and [20], for some lakes air2water model may be insensitive to the parameters a_7 and a_8, which are used solely for the calculations of δ, and only in a case when the water temperature is between 0 °C and 4 °C. Hence, a simplified

air2water model with six parameters was proposed [3] in the following form:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left(a_1 + a_2 T_a - a_3 T_w + a_5 \cos\left(2\pi\left(\frac{t}{t_y} - a_6\right)\right)\right) \tag{3}$$

$$\delta = \exp\left(-\frac{T_w - T_h}{a_4}\right) \text{ if } T_w \geq T_h$$

$$\delta = 1 \text{ if } T_w < T_h. \tag{4}$$

This way the parameters a_7 and a_8 are ignored and the model requires calibration of just six parameters (hence we call it a2w6). However, the relation between the air and the lake surface temperatures in colder periods of the year may be oversimplified and less adequate.

D. Proposed Air2water Model With 12 Parameters

The classical air2water model relates the air temperature with the temperature of water within the upper layer of the lake, hence the temperature that is observed within a body of liquid water at some depth under the surface. However, a conceptually similar approach could be also used to link air temperature with the temperature of the sheer surface of the lake that may be measured from remote sensing (e.g., [80]). In such a case the surface temperature may be measured at the top of either the liquid water or the ice that may cover the lake. Guo et al. [25] introduced the air2water variant in which heat exchange between the air and the water may in some periods be open, but in others become blocked by the ice. In [25], it was proposed that the temperature at the surface T_w may be computed either as the temperature of the liquid water

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left(a_1 + a_2 T_a - a_3 T_w + a_5 \cos\left(2\pi\left(\frac{t}{t_y} - a_6\right)\right)\right) \tag{5}$$

with

$$\delta = \exp\left(-\frac{T_w - T_h}{a_4}\right) \quad \text{if } T_w \geq T_h$$

$$\delta = \exp\left(-\frac{T_h - T_w}{a_7}\right) + \exp\left(-\frac{T_w}{a_8}\right) \quad \text{if } T_w < T_h \tag{6}$$

or as the temperature of the ice on the lake's surface

$$\frac{dT_i}{dt} = \frac{1}{\delta}\left(a_9 + a_{10} T_a - a_{11} T_w + a_{12} \cos\left(2\pi\left(\frac{t}{t_y} - a_{13}\right)\right)\right) \tag{7}$$

or as a mix of both temperatures. Which equation will be used is governed by water temperature (T_w) and two additional model parameters (a_{14} and a_{15}). If $T_w \geq a_{15}$, (5) is used; if $T_w \leq a_{14}$, (7) is used, and when $a_{15} > T_w > a_{14}$, the mix of both (5) and (7) is applied such that the result from (5) is multiplied by $(1 - K)$ and the result from (7) is multiplied by K, where $K = (a_{15} - T_w)/(a_{15} - a_{14})$ is a ratio between frozen and open water part of the lake surface.

Note that (7) is similar to (5) or (1), but with different parameter names (a_9 instead of a_1, a_{10} instead of a_2, and so on). This allows the model to perform differently in icy and liquid surface

conditions. The approach proposed in [25] was defined for cases when one aims to determine the temperature of the lake surface with satellite data, but it does not skip the problem with the lack of model sensitivity to a_7 and a_8 parameters. However, if we are interested in the temperature of liquid water in a surface layer of the lake (in the case of Polish lakes, either 0.4 m under the open surface or under the ice cover), one may eliminate these two parameters by using two similar equations with a separate parameterization for different lake stratifications in a similar way as proposed in [25]. This new variant developed for in situ water temperature measurements may be proposed in the following form:

$$\frac{dT_w}{dt} = a_1 + a_2 T_a + a_3 T_w + a_4 \cos\left(2\pi\left(\frac{t}{t_y} - a_5\right)\right)$$

if $T_w \geq a_{11} + a_{12}$

$$\frac{dT_w}{dt} = a_6 + a_7 T_a + a_8 T_w + a_9 \cos\left(2\pi\left(\frac{t}{t_y} - a_{10}\right)\right)$$

if $T_w \leq a_{11}$

$$\frac{dT_w}{dt} = (1 - K)$$

$$\times \left(a_1 + a_2 T_a + a_3 T_w + a_4 \cos\left(2\pi\left(\frac{t}{t_y} - a_5\right)\right)\right)$$

$$+ K\left(a_6 + a_7 T_a + a_8 T_w + a_9 \cos\left(2\pi\left(\frac{t}{t_y} - a_{10}\right)\right)\right)$$

if $a_{11} < T_w < a_{11} + a_{12}$ \tag{8}

where $a_{11}, a_{12} \in [0\,°C, 25\,°C]$, and K that appears in (8) represents the ratio between the occurrence of two different stratifications (attributed to cold or warm waters), defined as

$$K = \frac{a_{11} + a_{12} - T_w}{a_{12}}. \tag{9}$$

In the proposed variant of the air2water model [(8) and (9)], there are 12 parameters: a_1–a_5 and a_6–a_{10} represent similar processes, but for warmer (a_1–a_5) or colder (a_6-a_{10}) waters. Similar to the idea presented in [25], the role of K, and hence the parameters a_{11} and a_{12}, is to choose whether the water temperature should be computed according to the version representative for warm waters, cold waters, or a mix of both. This division is important due to different stratification dynamics in colder and warmer waters [81], [82]. Note that δ does not appear in this variant of the air2water model. Because the new variant uses different approximations for cold and warm waters, it may better fit the conditions in the particular lake. The number of parameters (12, hence we call this variant a2w12) is larger than in the case of the classical (8 parameters) or simplified (6 parameters) air2water variants, but lower than in the version designed for satellite measurements [25]. In Supplementary Table I, the bounds for each parameter used in a2w6, a2w8, and a2w12 variants are given.

E. Optimization Criteria

In the present study, we test the performance of three different air2water model variants when each of them is calibrated with eight different optimization algorithms. As a result, we compare both the models and the optimizers. The objective function used during optimization is the mean square error (MSE)

$$\text{MSE} = \frac{\sum_{i=1}^{N} \left(T_w^P (i) - T_w^M (i) \right)^2}{N} \tag{10}$$

where N is the number of observations in the calibration data set, and T^P and T^M are predicted and measured lake temperatures at the ith day, respectively. Obviously, the data from the validation set are used to compute MSE for the validation period (with different days i and a different number of days N).

The maximum number of function calls is set to 100 000 for all considered metaheuristics, irrespective of the variant of the air2water model, and hence of the number of parameters. This approach is similar to the one applied in the CEC2011 competition on real-world problems (where 150 000 function calls were allowed) but different from the majority of CEC competitions held on artificial benchmark problems, in which the number of function calls depends on the problem dimensionality (e.g., [76], [83], and [84]). The fixed number of function calls benefits rather the classical than the new air2water variant. During empirical tests, each optimization algorithm was run 30 times on every air2water variant and lake.

III. OPTIMIZATION METAHEURISTICS

A. Chosen Optimization Algorithms

In [85], 12 various metaheuristics were tested for the calibration of the basic air2water model with 8 parameters. In the present study, we introduce a new air2water model variant and analyze how to choose the right algorithm to train it. As a result, we need to apply recent and competing algorithms that belong to various families of methods. For historical reasons, the basic PSO is also tested, as it is the first method used for air2water optimization in the past. The eight metaheuristics tested, include the following:

1) The basic PSO algorithm [86] that has been used for the calibration of the basic air2water model in the past papers [87]. The PSO version with global topology [86], $c_1 = c_2 = 1.49445$ [88], inertia weight decreasing linearly with time from 0.9 at the start of the search to 0.4 at the end [89] and swarm size set to 50 [59] has been applied.
2) Triple archives PSO (TAPSO) [90] is a new version of PSO that divides the particles into different subsets depending on their fitness and the speed of progress and proceeds with both swarm intelligence and genetic operators. We used 60 particles in the experiment, as there was no guide provided in [90] on how to set the population size for a particular type of problems.
3) HARD-DE [91] that turned out the best method for the classical eight-parameter version of the air2water model in [85]. The good performance of HARD-DE was also shown in [91] against a number of other DE variants on various

benchmarks. HARD-DE uses nonlinear, parabolic population size reduction, from initial $25 \times lnD \times sqrt(D)$ (but, in the present study, not fewer than 25) to 4 at the end of the search.

4) EL-SHADE-SPACMA [92], which was ranked the third best method in the CEC2018 competition. This algorithm is an updated version of L-SHADE-SPACMA [93], and is a member of the L-SHADE family of methods [61] that was developed by step-by-step improvements of older algorithms [94] and performs exceptionally well in many competitions between metaheuristics [95]. This variant uses linear population size reduction during the run; the algorithm starts from the population size set to 18D and ends with four individuals.
5) L-SHADE-50-PWI [96] that, unusually, is a simplified version of some L-SHADE variants, developed by reducing some of their unneeded elements, but adding an inertia weight that is so widely used in PSO algorithms [89]. In the source paper, it outperformed a number of other L-SHADE variants on CEC2017, CEC2014, and various real-world problems. L-SHADE-50-PWI found a successful application in the soil erosion analysis [97] and achieved reasonable results for brain magnetic resonance image segmentation [98]. The use of inertia weight within L-SHADE was also considered a promising solution in space flight trajectory design [99].
6) djDE with diversity-based adaptive population size [60]. This approach combines one of the best DE variants from early 2000's, jDE algorithm [100] with a new population size adaptivity method that allows to balance exploration and exploitation capabilities by controlling the population diversity [101].
7) Stochastic fractional search algorithm with fitness-distance balance (FDB) [102], which is a fresh approach based on the fitness-distance balance algorithm proposed in 2020 [103]. The population size is set to 50.
8) Self-organizing migrating algorithm-T3A – (SOMA, [104]), an adaptive extension of an algorithm proposed in [105], that has been ranked 4th out of 18 algorithms for the 2019 100-Digit Challenge on real-parameter single objective optimization [63], a competition with (almost) unlimited time to solve 10 hard problems with precision up to 10 digits of accuracy. SOMA solved 9 problems out of 10 with the required accuracy, the only algorithm that solved all problems with assumed precision was a variant of jDE called jDE100 [106]—a modification of which, called djDE [60], has already been included in our comparison.

Three among eight tested algorithms are based on the famous L-SHADE [61] optimizer that won IEEE Competition in Evolutionary Computation in 2014. We have used three algorithms from this family of methods, as several L-SHADE-based variants won various IEEE Competitions or achieved very good performances in comparison studies [94]. The efficient L-SHADE algorithms include the three variants applied in this study, as well as SPS-L-SHADE-EIG that uses rotation invariant mechanism [107], L-SHADE-cnEpSin [108] that uses an ensemble

of sinusoidal approaches in parameter adaptation, L-SHADE-X [109] that introduces dimensionality-dependent spread parameter in distribution from which scaling factor is generated and an archive with offsprings that did not reach the main population despite having good performance, or OLSHADE-CS [110] that uses novel initialization and conservative selection mechanisms. In our opinion, large successes of L-SHADE-based algorithms justify testing various variants on different practical applications, as in [57], [111], and [112].

All the metaheuristics were run with default control parameters as proposed by their inventors, with a few exceptions noted in the above description. In our tests, first, each method is tested on 22 real-world problems from CEC2011, and, due to the low-dimensionality of the air2water model–on 10-dimensional CEC2017 artificial benchmarks. After that, each method is applied to every air2water variant for each lake. A comparison of the results would show us to what extent the choice of metaheuristics based on either various real-world problems or CEC2017 artificial benchmarks will be useful for choosing the appropriate approach for a particular air2water variant.

On CEC2011 and CEC2017 problems, each algorithm is run 51 times as suggested in [76]. We compare the optimization algorithms in three ways.

First, the averaged performance of each algorithm on every problem is computed, and algorithms are ranked from the best (rank 1) to the worst (rank 8) on each problem. The rankings are then averaged over all problems in the set, and Friedman's test is used with the *post hoc* Shaffer's static procedure at $\alpha = 0.05$ to choose if there are statistically significant differences between algorithms [113], [114]. This is the classical way of comparing metaheuristics, used in plenty of papers. However, it compares just average performance, and the specialization of a particular metaheuristic in a particular type of problem is often lost; as a result, the algorithm that performs above the average for all problems may be ranked best, even if it is not the best approach for any specific problem.

Second, we count the number of problems for which the particular algorithm has won. The algorithm that achieved the best-averaged performance on the particular problem is termed a winner. We set a range of precision to 10^{-19}. If the difference between performances achieved by some algorithms is lower than 10^{-19}, and no other algorithm performed better, we consider all such equally performing algorithms as winners. As a result, there may be more than 1 winner for the particular problem. By counting the number of problems for which the particular algorithm turns out a winner we find whether the algorithm may be especially useful for some problems. However, the possible failures of the algorithm in finding any reasonable solutions for some problems are neglected by this measure.

Thirdly, instead of averaged ranking, we count the number of wins differently. For each problem, we find the best run achieved by the particular algorithm. Then, we compare the best runs achieved by different algorithms and find a winner for a particular problem. After that, we count the number of problems for which the particular algorithm turned out a winner using the best run performance. This way we may look at algorithms that achieve diversified performance in each run,

and may occasionally lead to very good results, even though some poor runs may highly affect their average performance. The algorithm chosen in this way would not necessarily be reliable for a single application and may perform poorly for some problems. However, for other problems, if run many times, it may find a solution better than algorithms that were chosen according to the average performance. Here we do not claim which way of comparison is a better option for choosing the right approach for the specific problem—in our case the calibration of the air2water variants. Our only intent is to verify how consistent (or not) the results will be if we use different ways to compare metaheuristics.

B. Comparing Metaheuristics on CEC2017 Benchmarks

The CEC2017 set contains 30 problems with various difficulties [76]. Because tested air2water model variants have 6 to 12 parameters, we have selected 10-dimensional versions of CEC2017 problems.

The average results obtained by eight optimization algorithms on each among thirty 10-dimensional CEC2017 problems are given in Supplementary Table II. The averaged ranking of algorithms, and the number of wins based on the averaged performance is given in Table II. As may be seen, based on the ranking averaged over all problems HARD-DE turned out to be the best method (with mean ranking = 2.4), followed by EL-SHADE-SPACMA (mean ranking = 2.72), djDE (3.0), and L-SHADE-50-PWI (3.5). Clearly, the worst method was SOMA (mean ranking = 7.44). In Supplementary Table III, the statistical significance of the differences achieved by all algorithms, computed according to Friedman's test with Shaffer's *post hoc* method with $\alpha = 0.05$, is presented. As may be seen, the differences between the best 4 algorithms are not statistically significant. FDB, ranked fifth, is significantly poorer than HARD-DE, but the difference between FDB and the other three well-performing methods is statistically insignificant.

Marginally different results are obtained if we count the number of wins for each algorithm when the win is based on the averaged performance for the particular problem. In such a case HARD-DE and EL-SHADE-SPACMA perform equally well with 12 wins, and djDE and L-SHADE-50-PWI share 10 wins. Among the other four methods, FDB achieves 7 wins, whereas other methods achieve at best a single win among 30 problems.

Based on these two ways of comparison, HARD-DE may be chosen as the most appropriate method for 10-dimensional benchmark functions, closely followed by EL-SHADE-SPACMA. Two other algorithms, i.e., L-SHADE-50-PWI and djDE, perform only marginally poorer. Other methods are clearly inferior, and only FDB may be considered somehow competitive.

Supplementary Table IV contains the best results obtained by particular algorithms on the specific problems in 51 runs. The number of wins counted with respect to this best performance is also given in Table II. If we count the number of problems for which the particular algorithm can find a better solution than the

TABLE II
SUMMARY OF THE RESULTS OBTAINED FOR 10 DIMENSIONAL CEC 2017 BENCHMARKS AND 22 CEC2011 REAL-WORLD PROBLEMS (DETAILS ARE GIVEN IN SUPPLEMENTARY TABLES)

	PSO	SOMA	TAPSO	FDB	L-SHADE-50-PWI	djDE	EL-SHADE-SPACMA	HARD-DE
10 dimensional CEC 2017 problems								
all-problems averaged rank	6.37	7.43	6.30	4.29	3.50	3.00	2.72	2.40
nr of wins based on average ranking	1	0	1	7	10	10	12	12
nr of wins based on best performance	8	0	8	12	17	15	17	15
CEC 2011 real-world problems								
all-problems averaged rank	3.64	7.48	5.05	5.43	2.57	5.16	3.64	3.05
nr of wins based on average ranking	2	1	4	3	10	2	2	6
nr of wins based on best performance	2	1	8	5	9	2	7	6

TABLE III
NUMBER OF TIMES PARTICULAR Air2water VARIANT TURNS OUT THE BEST DEPENDING ON THE METAHEURISTIC USED FOR OPTIMIZATION WHEN MEAN PERFORMANCE (MSE) FROM ALL RUNS IS COMPARED

	CALIBRATION DATA			TESTING DATA		
problem	a2w12	a2w8	a2w6	a2w12	a2w8	a2w6
PSO	3	0	19	0	0	22
SOMA	0	1	21	0	3	19
TAPSO	8	0	14	6	0	16
FDB	13	3	6	10	1	11
L-SHADE-50-PWI	13	2	7	11	1	10
djDE	15	1	6	12	0	10
EL-SHADE-SPACMA	16	2	4	15	1	6
HARD-DE	21	0	1	19	0	3

TABLE IV
NUMBER OF TIMES PARTICULAR Air2water VARIANT TURNS OUT THE BEST DEPENDING ON THE METAHEURISTIC USED FOR OPTIMIZATION WHEN THE PERFORMANCE (MSE) FROM THE BEST RUN IS CONSIDERED FOR COMPARISON

	CALIBRATION DATA			TESTING DATA		
problem	a2w12	a2w8	a2w6	a2w12	a2w8	a2w6
PSO	22	0	0	20	0	2
SOMA	2	4	16	0	5	17
TAPSO	22	0	0	20	0	2
FDB	21	0	1	19	0	3
L-SHADE-50-PWI	21	0	1	18	0	4
djDE	21	0	1	20	0	2
EL-SHADE-SPACMA	21	0	1	18	0	4
HARD-DE	21	0	1	19	0	3

other algorithms (including ties) in the best run, both L-SHADE-based algorithms, namely L-SHADE-50-PWI and EL-SHADE-SPACMA, would be ranked the best, with 17 wins (including ties) each. HARD-DE and djDE are slightly less efficient, with 15 wins each. FDB achieves 12 wins, both PSO and TAPSO – 8; only SOMA remains without any win.

Although these findings are not in contradiction to the results based on the average ranking, they nonetheless clearly differ. First of all, L-SHADE-50-PWI and EL-SHADE-SPACMA turn out more efficient in finding the best solutions than HARD-DE. Secondly, the best solutions for specific problems are found by seven various algorithms, including FDB, PSO, and TAPSO.

TABLE V
MEAN PERFORMANCE OF METAHEURISTICS ON CALIBRATION OF a2w12 VARIANT FOR EACH LAKE (VALIDATION SET)

problem	PSO	SOMA	TAPSO	FDB	L-SHADE-50-PWI	djDE	EL-SHADE-SPACMA	HARD-DE
Slawskie	8.41E-01	6.04E-01	8.57E-01	5.43E-01	6.04E-01	5.43E-01	5.43E-01	5.43E-01
Jamno	2.02E+00	1.90E+00	2.27E+00	1.83E+00	1.83E+00	1.83E+00	1.83E+00	1.83E+00
Gardno	1.67E+00	1.59E+00	1.62E+00	1.59E+00	1.59E+00	1.59E+00	1.59E+00	1.59E+00
Lebsko	1.55E+00	1.36E+00	1.36E+00	1.36E+00	1.36E+00	1.36E+00	1.36E+00	1.36E+00
Jasien	3.26E+00	3.12E+00	3.55E+00	2.99E+00	2.99E+00	3.01E+00	2.86E+00	2.83E+00
Radunskie	9.09E-01	1.19E+00	1.29E+00	9.19E-01	9.37E-01	1.21E+00	8.28E-01	7.92E-01
Wdzydze	9.38E-01	1.35E+00	1.36E+00	1.08E+00	1.04E+00	1.29E+00	1.02E+00	1.02E+00
Charzykowskie	1.15E+00	1.31E+00	1.43E+00	9.79E-01	1.09E+00	1.15E+00	9.23E-01	9.23E-01
Sepolonskie	9.42E-01	8.76E-01	1.57E+00	7.33E-01	7.00E-01	7.00E-01	7.00E-01	7.00E-01
Bachotek	8.79E-01	7.24E-01	1.10E+00	6.12E-01	5.85E-01	6.92E-01	5.58E-01	5.58E-01
Jeziorak	1.70E+00	1.48E+00	1.98E+00	1.38E+00	1.38E+00	1.38E+00	1.38E+00	1.38E+00
Dadaj	1.21E+00	1.72E+00	1.72E+00	1.39E+00	1.22E+00	1.60E+00	1.15E+00	1.05E+00
Mikolajskie	1.08E+00	1.03E+00	1.46E+00	8.18E-01	1.08E+00	1.13E+00	7.38E-01	6.86E-01
Sniardwy	9.19E-01	7.94E-01	9.24E-01	7.79E-01	7.95E-01	7.80E-01	7.81E-01	7.79E-01
Nidzkie	9.14E-01	7.77E-01	1.14E+00	6.77E-01	6.77E-01	6.76E-01	6.76E-01	6.76E-01
Mamry	9.87E-01	9.54E-01	1.38E+00	7.93E-01	7.93E-01	8.20E-01	7.93E-01	7.93E-01
Elckie	9.70E-01	1.36E+00	1.30E+00	9.74E-01	9.13E-01	1.12E+00	8.29E-01	7.46E-01
Selmet	8.87E-01	9.68E-01	1.20E+00	8.34E-01	8.57E-01	9.56E-01	6.60E-01	6.60E-01
Rajgrodzkie	1.24E+00	1.50E+00	1.62E+00	1.11E+00	1.34E+00	1.41E+00	9.86E-01	9.86E-01
Litygajno	1.01E+00	9.41E-01	1.36E+00	7.38E-01	7.66E-01	8.20E-01	7.38E-01	7.38E-01
Biale	9.25E-01	1.12E+00	1.31E+00	8.15E-01	9.18E-01	1.04E+00	7.75E-01	7.74E-01
Studzienicze	9.39E-01	1.11E+00	1.29E+00	7.96E-01	8.91E-01	1.02E+00	7.97E-01	7.96E-01
mean rank	5.73	6.27	7.86	3.80	3.93	4.68	2.23	1.50

Only SOMA remains fully unable to locate the best solution for any problems. This suggests that the good performance of HARD-DE observed when the performance is averaged over all runs may be due to the consistent and reliable performance in all runs, rather than the ability to find a solution that cannot be found by other optimizers.

According to the all above discussion based on artificial benchmarks of similar dimensionality to air2water models, one could conclude that one of four algorithms, i.e., HARD-DE, EL-SHADE-SPACMA, L-SHADE-50-PWI, and djDE, should be applied to the calibration of air2water model variants.

C. Comparing Metaheuristics on 22 Real-World Problems From Different Fields of Science

In Supplementary Tables V–VII, the same eight algorithms are compared on 22 real-world problems of different dimensionality that originate from various fields of science or industry. Supplementary Table V contains the averaged performance obtained on a particular problem, the ranking of algorithms averaged over all problems, and the number of wins achieved according to the average performance. Supplementary Table VI contains information on the significance of the differences between performances obtained by various algorithms according to Friedman's test with Shaffer's *post hoc* procedure with $\alpha = 0.05$. Supplementary Table VII shows the information on the

best performance in 51 runs achieved by a particular algorithm on a specific problem, and the number of wins counted with respect to the best performance. The main results are, like in the case of CEC 2017 benchmarks, summarized in Table II.

According to the averaged ranks based on the mean performance, on real-world problems, L-SHADE-50-PWI turns out the best method (mean ranking = 2.57). L-SHADE-50-PWI is followed by HARD-DE (mean ranking = 3.05), PSO (3.64), and EL-SHADE-SPACMA (3.64). The djDE algorithm does not perform well on real-world problems and is only ranked 6th (mean ranking = 5.43). SOMA remains the poorest approach (with a mean ranking of 7.48).

Like in the case of CEC2017 problems, the differences between the four best methods are not statistically significant at $\alpha = 0.05$. However, L-SHADE-50-PWI is significantly better than SOMA, djDE, TAPSO, and FDB algorithms.

If we count the number of problems for which the particular algorithm is the winner according to the mean performance (see Table II and Supplementary Table V), L-SHADE-50-PWI is clearly the best approach (ten wins), followed by HARD-DE (six wins) and, surprisingly, TAPSO (four wins) and FDB (three wins). This shows the superiority of L-SHADE-50-PWI and, to some extent HARD-DE, over the other methods. However, it also shows that some algorithms that do not perform well across all considered real-world problems (i.e., TAPSO and FDB) may be well suited for the specific goals: TAPSO is the best method for

TABLE VI
MEAN PERFORMANCE OF METAHEURISTICS ON CALIBRATION OF a2w8 VARIANT FOR EACH LAKE (VALIDATION SET)

problem	PSO	SOMA	TAPSO	FDB	L-SHADE-50-PWI	djDE	EL-SHADE-SPACMA	HARD-DE
Slawskie	7.89E-01	8.59E-01	7.15E-01	7.11E-01	5.96E-01	6.29E-01	5.96E-01	5.96E-01
Jamno	1.71E+00	1.69E+00	1.77E+00	1.84E+00	1.84E+00	1.83E+00	1.86E+00	1.86E+00
Gardno	1.55E+00	1.59E+00	1.46E+00	1.51E+00	1.50E+00	1.44E+00	1.41E+00	1.41E+00
Lebsko	1.53E+00	1.54E+00	1.54E+00	1.54E+00	1.54E+00	1.53E+00	1.54E+00	1.54E+00
Jasien	1.54E+00	1.76E+00	1.42E+00	1.39E+00	1.37E+00	1.38E+00	1.38E+00	1.37E+00
Radunskie	1.09E+00	1.29E+00	1.02E+00	1.11E+00	1.36E+00	1.26E+00	1.26E+00	8.73E-01
Wdzydze	1.08E+00	1.86E+00	8.39E-01	1.24E+00	7.43E-01	7.46E-01	7.42E-01	7.42E-01
Charzykowskie	1.83E+00	1.82E+00	1.42E+00	1.39E+00	1.71E+00	1.45E+00	1.46E+00	1.34E+00
Sepolonskie	9.08E-01	1.17E+00	8.58E-01	7.79E-01	7.80E-01	7.75E-01	7.74E-01	7.74E-01
Bachotek	1.37E+00	1.62E+00	1.30E+00	1.21E+00	1.19E+00	1.19E+00	1.19E+00	1.19E+00
Jeziorak	2.97E+00	1.93E+00	1.89E+00	1.83E+00	1.83E+00	1.75E+00	1.75E+00	1.81E+00
Dadaj	1.52E+00	1.99E+00	1.45E+00	1.50E+00	1.25E+00	1.49E+00	1.22E+00	1.21E+00
Mikolajskie	8.61E-01	1.19E+00	7.66E-01	7.32E-01	6.77E-01	6.96E-01	6.76E-01	6.40E-01
Sniardwy	9.53E-01	1.05E+00	8.88E-01	8.80E-01	1.51E+01	1.46E+00	1.46E+00	8.86E-01
Nidzkie	9.05E-01	9.91E-01	8.78E-01	7.90E-01	7.90E-01	7.75E-01	7.75E-01	7.75E-01
Mamry	1.01E+00	1.19E+00	1.07E+00	8.93E-01	8.16E-01	8.51E-01	8.38E-01	7.72E-01
Elckie	7.54E-01	1.20E+00	6.86E-01	5.92E-01	5.91E-01	5.92E-01	5.91E-01	5.92E-01
Selmet	8.06E-01	1.19E+00	7.13E-01	7.75E-01	6.44E-01	6.82E-01	6.63E-01	6.05E-01
Rajgrodzkie	2.94E+00	2.95E+00	2.32E+00	2.11E+00	1.00E+06	2.09E+00	2.08E+00	2.08E+00
Litygajno	8.04E-01	1.16E+00	7.24E-01	6.33E-01	6.20E-01	6.22E-01	6.23E-01	6.25E-01
Biale	9.50E-01	1.32E+00	8.25E-01	7.20E-01	8.39E-01	7.17E-01	7.36E-01	6.96E-01
Studzienicze	9.40E-01	1.31E+00	9.10E-01	7.22E-01	7.18E-01	7.16E-01	7.17E-01	7.19E-01
mean rank	6.14	7.32	4.95	4.77	3.91	3.36	3.14	2.41

TABLE VII
MEAN PERFORMANCE OF METAHEURISTICS ON CALIBRATION OF a2w6 VARIANT FOR EACH LAKE (VALIDATION SET)

problem	PSO	SOMA	TAPSO	FDB	L-SHADE-50-PWI	djDE	EL-SHADE-SPACMA	HARD-DE
Slawskie	6.92E-01	1.45E+00	6.92E-01	6.92E-01	6.92E-01	6.92E-01	6.92E-01	6.92E-01
Jamno	1.70E+00	1.61E+00	1.70E+00	1.70E+00	1.70E+00	1.70E+00	1.70E+00	1.70E+00
Gardno	1.47E+00	1.49E+00	1.47E+00	1.47E+00	1.47E+00	1.47E+00	1.47E+00	1.47E+00
Lebsko	1.45E+00	1.45E+00	1.46E+00	1.45E+00	1.45E+00	1.45E+00	1.45E+00	1.45E+00
Jasien	1.40E+00	1.59E+00	1.42E+00	1.40E+00	1.40E+00	1.40E+00	1.40E+00	1.40E+00
Radunskie	8.77E-01	1.08E+00	9.07E-01	8.78E-01	8.78E-01	8.78E-01	8.78E-01	8.78E-01
Wdzydze	8.84E-01	1.28E+00	8.84E-01	8.84E-01	8.84E-01	8.84E-01	8.84E-01	8.84E-01
Charzykowskie	1.42E+00	1.57E+00	1.42E+00	1.42E+00	1.42E+00	1.42E+00	1.42E+00	1.42E+00
Sepolonskie	8.35E-01	9.38E-01	8.43E-01	8.35E-01	8.35E-01	8.35E-01	8.35E-01	8.35E-01
Bachotek	1.33E+00	1.45E+00	1.33E+00	1.33E+00	1.33E+00	1.33E+00	1.33E+00	1.33E+00
Jeziorak	1.48E+00	1.56E+00	1.49E+00	1.48E+00	1.48E+00	1.48E+00	1.48E+00	1.48E+00
Dadaj	1.39E+00	1.55E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00	1.39E+00
Mikolajskie	6.85E-01	7.43E-01	6.85E-01	6.85E-01	6.85E-01	6.85E-01	6.85E-01	6.85E-01
Sniardwy	9.36E-01	9.52E-01	9.36E-01	9.36E-01	9.36E-01	9.36E-01	9.36E-01	9.36E-01
Nidzkie	7.88E-01	8.21E-01	7.88E-01	7.88E-01	7.88E-01	7.88E-01	7.88E-01	7.88E-01
Mamry	8.14E-01	9.34E-01	8.15E-01	8.15E-01	8.15E-01	8.15E-01	8.15E-01	8.15E-01
Elckie	6.38E-01	7.63E-01	6.38E-01	6.38E-01	6.38E-01	6.38E-01	6.38E-01	6.38E-01
Selmet	6.80E-01	8.22E-01	6.80E-01	6.80E-01	6.80E-01	6.80E-01	6.80E-01	6.80E-01
Rajgrodzkie	2.30E+00	2.56E+00	2.30E+00	2.30E+00	2.30E+00	2.30E+00	2.30E+00	2.30E+00
Litygajno	7.09E-01	8.52E-01	7.09E-01	7.09E-01	7.09E-01	7.09E-01	7.09E-01	7.09E-01
Biale	8.37E-01	9.80E-01	8.37E-01	8.37E-01	8.37E-01	8.37E-01	8.37E-01	8.37E-01
Studzienicze	8.35E-01	9.55E-01	8.35E-01	8.35E-01	8.35E-01	8.35E-01	8.35E-01	8.35E-01
mean rank	2.59	7.64	3.95	4.59	4.05	5.00	4.36	3.82

30-dimensional problem 2 and 110- and 216-dimensional problems 11.1 and 11.2; FDB—is the best method for 1-dimensional problem 4 and 96-dimensional problem 11.10. Note also that EL-SHADE-SPACMA turned out inefficient in this competition.

When we compare the number of wins for the best among 51 runs performed on a particular problem, the results are much different. Although L-SHADE-50-PWI remains the best approach with nine wins, it is only marginally better than TAPSO with eight wins. EL-SHADE-SPACMA has seven wins, HARD-DE has six, and FDB has five. Hence, five out of eight algorithms can find solutions for the specific problem that is better (including ties) than all other solutions for this problem found by the other methods. We may again note that the specialization of algorithms for a particular kind of search becomes much more evident when the best runs, instead of the mean performances, are compared.

Tests performed on 22 CEC2011 real-world problems do not confirm the superiority of HARD-DE, and do not point at djDE as a promising method. Instead, overall for real-world problems L-SHADE-50-PWI seems to be the best initial choice. However, for many specific real-world problems HARD-DE, TAPSO, EL-SHADE-SPACMA, or even basic PSO may be considered the second-best approach. It seems that choosing the appropriate metaheuristics for the air2water model calibration based on the results obtained for various-dimensional real-world problems is more difficult than making this choice based on artificial benchmarks of fixed dimensionality.

IV. RESULTS AND DISCUSSION

A. Choosing the Best Air2water Model Variant Using Different Optimizers

In this section we use, one by one, each among eight metaheuristics to calibrate the three air2water variants on every among 22 lakes. We run each metaheuristic 30 times for every variant and lake. We consider two ways of comparison—mean and best performance among 30 runs. MSE (10) is used as an objective function and as a comparison criterion. We separately compare air2water variants for calibration and for testing data sets.

The graphical comparison of averaged MSE obtained for three air2water variants calibrated by each metaheuristic is given in Fig. 2; in Supplementary Fig. 1, a similar comparison is shown solely for the MSE obtained in the best run. The number of times the particular air2water variant turns out the best according to each comparison scenario is given in Tables III and IV—we seek the best air2water variant for each lake, and then sum up the number of lakes for which particular air2water variant is the best choice.

From Fig. 2 and Supplementary Fig. 1 and Tables III and IV, we note that the relative performance of different air2water variants indeed depends on the metaheuristic used for comparison. The results also differ depending on whether we compare the averaged MSE from all runs, or the best result obtained in all runs.

When MSE is averaged over all runs, three algorithms, namely PSO, SOMA, and TAPSO point at the superiority of a2w6, the variant with the lowest number of parameters, over a2w8 and

the newly proposed a2w12 (see Table III). The choice based on the calibration data is, for these three algorithms, confirmed on the testing data set. When the FDB algorithm is used for calibration, a2w12 seems to be the best variant according to the calibration data, but this result is not confirmed on the testing set. On the contrary, when air2water model variants are calibrated with L-SHADE-50-PWI or djDE, the a2w12 variant seems to be the best according to the calibration data, and the validation data set does not change the outcome (although the number of wins achieved by a2w12 and a2w6 is close to each other, a2w12 is marginally better). Finally, HARD-DE and EL-SHADE-SPACMA algorithms show the clear advantage of a2w12 over the two simpler variants. When HARD-DE is used for calibration, a2w12 is better than both a2w6 and a2w8 on 21 out of 22 lakes according to the calibration data, and on 19 lakes according to the validation set.

From such results, we see that the a2w12 variant may outperform the other two versions on almost all lakes providing that the appropriate optimization method is chosen for calibration. In such a case the superiority of a2w12 is confirmed on independent validation data set. However, if for example the classical PSO algorithm is used to calibrate the models—as in various recent studies on air2water—the simplest air2water variant with 6 parameters will be pointed out as the best option. It seems that simpler algorithms are able to find appropriate solutions for simpler air2water variants, but fail on more difficult ones. Only advanced metaheuristics are able to find appropriate solutions also for other air2water variants.

Interestingly, the results would change if one compares only solutions obtained in the best run (out of 30) of the calibration algorithm. In such a case only SOMA would point to a2w6 as the best choice (see Supplementary Fig. 1 and Table IV), all other algorithms clearly show the advantage of a2w12 over simpler variants, and this result is confirmed on both calibration and testing data sets. It means that almost every algorithm was able to converge to the better solution for the a2w12 model than for the a2w6 variant, but finding such good solutions for a2w12 was relatively harder than for a2w6, and was done successfully only in some runs. As a result, in the case of PSO, TAPSO, and FDB and to some extent L-SHADE-50-PWI and djDE, the conclusions based on the performance from the mean and the best run do differ. Only for HARD-DE and EL-SHADE-SPACMA, the final results are basically the same for both cases (compare Tables III and IV).

From Fig. 2 and Supplementary Fig. 1, one may note that when the HARD-DE algorithm is used for calibration, the results obtained by the a2w12 variant are clearly better than the results obtained by a2w8 and a2w6 calibrated by any algorithm. In that case a2w12 variant may not outperform other air2water variants only on three lakes: Jamno (lake no. 2 in Table I and Fig. 2 and Supplementary Fig. 1), Lebsko (lake no. 4), and Jeziorak (lake no. 11). For all other 19 lakes the performance of a2w12 is superior to the performance of a2w8 and a2w6 variants. May we find any common feature of these three lakes on which a2w12 fails? All three lakes are relatively shallow and relatively large compared to the other lakes under study (see Table I). Jamno and Lebsko are two of three most shallow lakes, out of 22, and
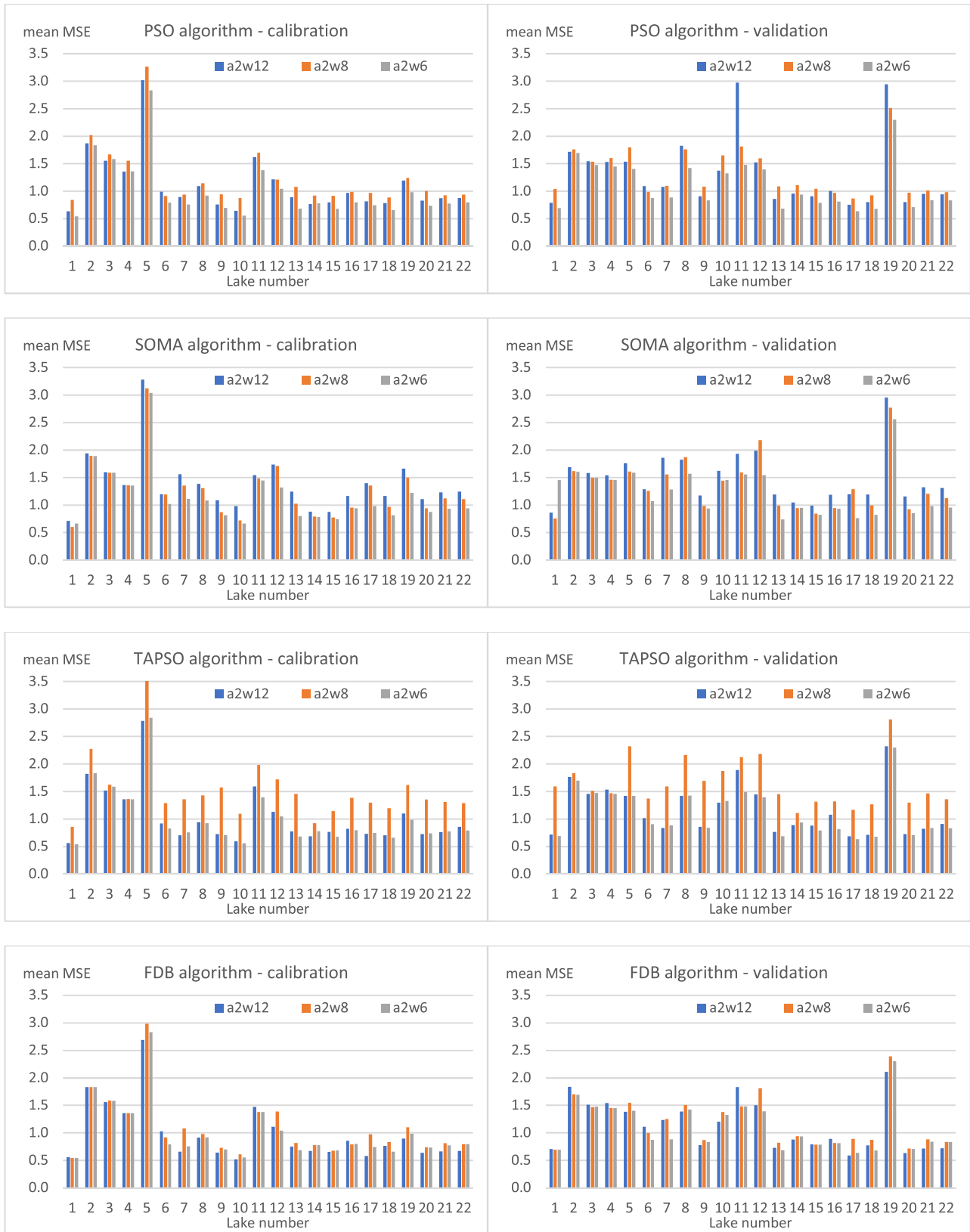
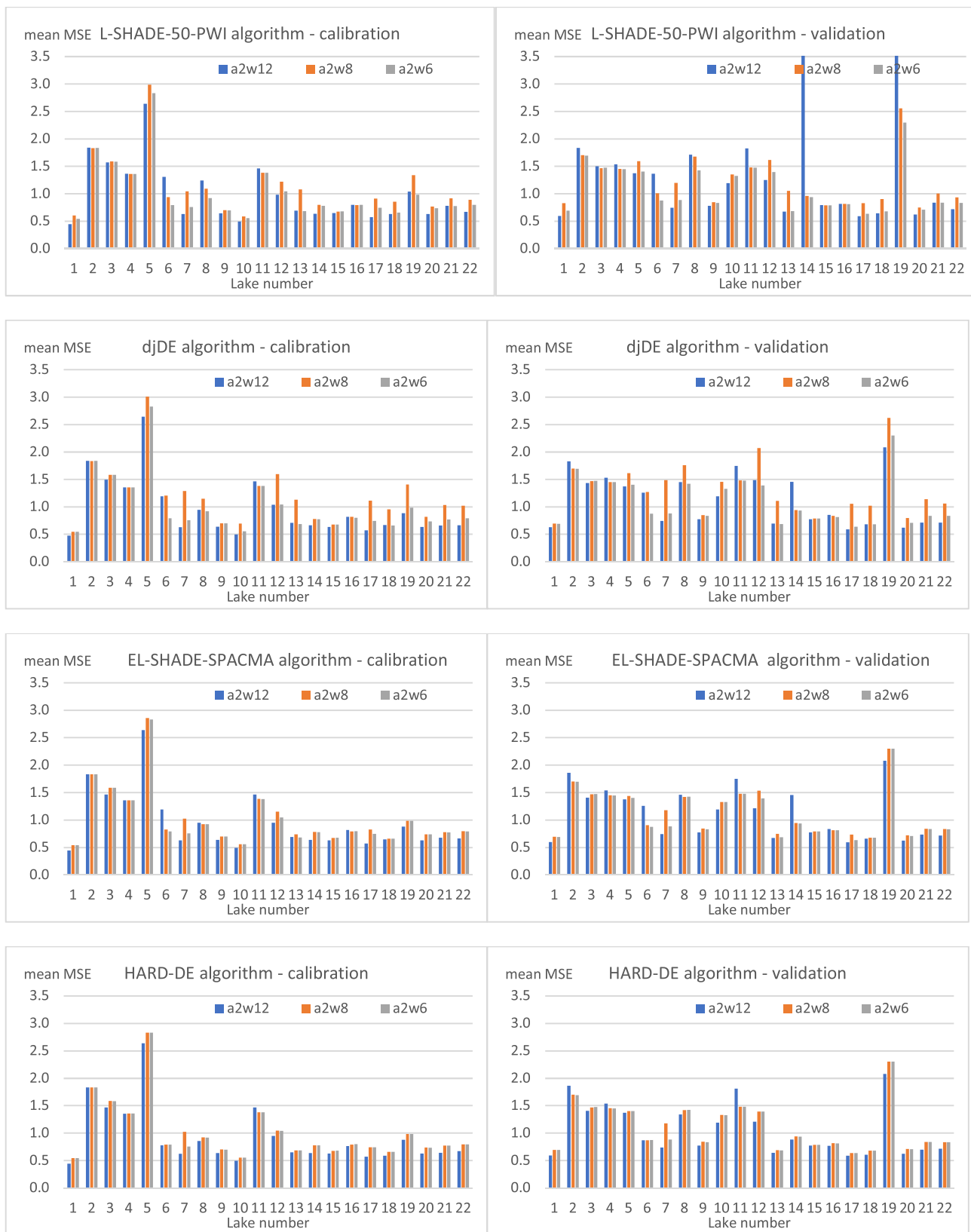Fig. 2.   Averaged MSE obtained by each air2water variant on each lake.

Fig. 2.   *(Continued.)*

Jeziorak is 4th–6th the most shallow lake, depending on whether one compares maximum or average depth. Lebsko lake is the 3rd largest lake, and both Jamno and Jeziorak are among the 6 largest lakes, out of 22 tested. No other common features of the three lakes may be found. This suggests that the a2w12 variant is not appropriate for large but shallow lakes. This observation may be confirmed (see Fig. 2 and Supplementary Fig. 1) by the very good performance of the a2w12 variant on the two deepest lakes—Wdzydze (lake 7) which is of moderate size, and Elckie (lake 17), which is very small. We suppose that the reason is in weak stratification of large shallow lakes, which results in low importance of seasonal mixing between deeper and surface lake layers. As a result, we may recommend an a2w12 variant of the air2water model for various lowland lakes with exception of large shallow ones.

B. Comparison Between Metaheuristics for Each Air2water Variant

In the previous section, we analyzed how the choice of calibration method would affect the choice of appropriate air2water variant. In this section, we compare the performance of metaheuristics for each individual air2water variant.

In Tables V–VII, the averaged over 30 runs performance of each algorithm achieved for every air2water variant is given for the validation set. The respective results for the calibration set are given in Supplementary Tables VIII–X. The results are shown separately for calibration and testing (validation) data sets. From the metaheuristics' point of view, the results for the calibration set are more meaningful, as the aim of the optimizer is to fit the model to these data. However, from a hydrological point of view results obtained for the testing set are much more important—as what is finally needed is the extrapolation of the calibrated model for the future, unknown data. In the bottom of Tables V–VII and Supplementary Tables VIII—X, the mean ranking (averaged over all 22 lakes) achieved by each metaheuristic is given. The lower the ranking, the better metaheuristic. We set the threshold in comparison among metaheuristics to 10^{-19}; if the difference in averaged performance is lower than this threshold for the particular lake, the methods are given an equal rank.

From Tables V–VII and Supplementary Tables VIII—X, we see that HARD-DE is the best choice to calibrate a2w12 and a2w8 variants. This result is consistent for calibration (see Supplementary Tables VIII—X) and validation (see Tables V–VII) data sets. In the case of the a2w8 variant, HARD-DE performs equally well for the calibration set as EL-SHADE-SPACMA and L-SHADE-50-PWI, for the other cases – it is the sole winner. EL-SHADE-SPACMA is constantly the second-best approach, and L-SAHDE-50-PWI and djDE compete for third place. All four metaheuristics perform well, and the differences between their results are relatively minor for the majority of lakes.

For a2w12 and a2w8 variants, the FDB algorithm is to some extent competitive, but the three remaining algorithms (TAPSO, PSO, and SOMA) perform much poorer. SOMA is the weakest approach.

The ranking of algorithms remains roughly similar for the a2w6 variant when calibration data are considered, with HARD-DE and EL-SHADE-SPACMA as two clear winners. Interestingly, for the validation set, PSO seems to be the best approach to calibrate the a2w6 variant. This may be highly surprising, but a quick look at the specific results given in Table VII shows that, from the practical point of view, the performance of all algorithms apart from SOMA is almost equal. Only the marginal differences noted for some lakes resulted in a better ranking of PSO in this specific case. The possible reason for the marginal superiority of PSO for validation data in the case of the air2water variant with just six parameters is that PSO was sufficient to find a good solution but unable to overfit to the specific training data; other algorithms (with exception of SOMA) fitted to both signal and noise present in the training sample.

Supplementary Figs. 2 and 3 allow visual comparison between observed and simulated lake water temperatures for the year 2016, which belongs to the validation period, on two typical lakes: Mikolajskie and Gardno. The simulated values are single realizations, out of 30 runs, from each metaheuristic for every air2water variant. One may note that simulations obtained by different metaheuristics for the a2w6 variant are almost identical (overlap on Supplementary Figs 2 and 3). On the contrary, simulations obtained by different metaheuristics for a2w8 and a2w12 variants are highly diversified for Lake Mikolajskie, and also differ, although on a much lower scale, for Lake Gardno. The differences between simulations obtained with the use of different optimizers may be large, and for some days exceed 3 °C. This confirms that the calibration of the a2w6 variant is rather straightforward, but the a2w8 and a2w12 variants are much harder to calibrate and require much more attention in choosing the appropriate calibration method.

Having all this in mind, one could conclude that, for the problem of air2water model calibration, the choice of the optimizer based on CEC2017 benchmarks seems more appropriate than the choice based on real-world problems of various dimensionalities that come from versatile fields of knowledge.

V. CONCLUSION

In this article, we have introduced a novel air2water model variant for modeling lake surface water temperatures based on air temperatures, and have shown that the choice of the optimization method may be crucial to confirm the good performance of the new model variant. The new variant uses separate parameterization for cold and warm water periods. If inappropriate metaheuristics are used, they point at simpler versions of the model as the better. Only more flexible optimization algorithms may find appropriate solutions for more complicated air2water model variants.

Which model variant performs the best may also depend on the performance measure. If instead of the averaged performance from all runs, one compares only the best performance achieved by a particular algorithm in all runs, seven out of eight considered metaheuristics would show the superiority of the new air2water model variant over its older versions, and only one would suggest the superiority of the simplest variant with six parameters. It

turns out that almost every algorithm was able to converge at least once to the sufficiently good solution of the new air2water variant to show its superiority. However, finding such a good solution was relatively hard, and many metaheuristics were able to achieve it only occasionally. Nonetheless, repeating computation many times and finding the best solution may be a simpler and cheaper option for many practitioners than seeking the method that will find a similarly good solution—but in every run.

We have also found that it is not easy to choose the right metaheuristics based on available benchmark problems from the literature. Depending on whether the tests are to be performed on 10-dimensional IEEE CEC2017 problems (which are similar to the air2water model variants in dimensionality), or on 22 real-world problems from the IEEE CEC2011 set (which, like optimization of the air2water model, represent real-world problems, not artificial functions), different metaheuristics may be considered as the best ones. Depending on the criteria of comparison, on CEC2017 problems HARD-DE or EL-SHADE-SPACMA are pointed out as the best algorithms. On the contrary, on CEC2011 real-world problems L-SHADE-50-PWI performs best and is followed by TAPSO, HARD-DE, or EL-SHADE-SPACMA.

In the practical application of the calibration of air2water model variants, the HARD-DE algorithm turns out clearly the best approach, confirming the choice that would be based on 10-dimensional artificial benchmarks, rather than various-dimensional real-world problems from diversified fields of science. When HARD-DE is used, the air2water model with 12 parameters turns out clearly better than the older air2water versions. However, if one would calibrate air2water models with L-SHADE-50-PWI, the superiority of the new air2water variant is weaker. If one decides on PSO for model calibration, the simplest air2water variant with six parameters would seem to be the best. Hence, the choice of metaheuristics for model calibration may be a crucial factor affecting the modeling outcome.

The general conclusion from our work regarding the search for the appropriate optimizer for the specific application based on the results obtained on benchmark problems is to look for algorithms that perform best on problems of similar dimensionality. Results obtained on versatile real-world problems of various dimensionalities may be misleading. If an analysis of the results for similarly-dimensional problems does not lead to any conclusions—a number of tests with different optimizers for the particular application of interest is needed.

We may finally note that if the right calibration method is chosen, the proposed air2water model variant with 12 parameters outperforms other air2water model variants on 19 out of 22 lakes. The new variant is not recommended for larger but shallow lakes but performs well on other kinds of lowland lakes. For large shallow lakes, we suggest using the simplified air2water model version with 6 parameters.

REFERENCES

[1] T. Jia, K. Yang, Z. Q. Peng, L. Tang, H. Duan, and Y. Luo, "Review on the change trend, attribution analysis, retrieval, simulation, and prediction of lake surface water temperature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6324–6355, 2022.

[2] S. Sharma, S. C. Walker, and D. A. Jackson, "Empirical modelling of lake water-temperature relationships: A comparison of approaches," *Freshwater Biol.*, vol. 53, no. 5, pp. 897–911, 2008.

[3] M. Toffolon et al., "Prediction of surface temperature in lakes with different morphology using air temperature," *Limnol. Oceanogr.*, vol. 59, no. 6, pp. 2185–2202, 2014.

[4] D. M. Farmer and E. Carmack, "Wind mixing and restratification in a lake near the temperature of maximum density," *J. Phys. Oceanogr.*, vol. 11, no. 11, pp. 1516–1533, 1981.

[5] A. Liikanen, T. Murtoniemi, H. Tanskanen, T. Vaisanen, and P. J. Martikainen, "Effects of temperature and oxygen availability on greenhouse gas and nutrient dynamics in sediment of a eutrophic mid-boreal lake," *Biogeochemistry*, vol. 59, pp. 269–286, 2002.

[6] X. Jia et al., "Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles," *ACM Trans. Data Sci.*, vol. 2, no. 3, 2021, Art. no. 20.

[7] A. Mazumder, W. D. Taylor, D. J. McQueen, and D. R. S. Lean, "Effects of fish and plankton on lake temperature and mixing depth," *Science*, vol. 247, pp. 312–315, 1990.

[8] M. Stomp, J. Huisman, G. G. Mittelbach, E. Litchman, and C. A. Klausmeier, "Large-scale biodiversity patterns in freshwater phytoplankton," *Ecology*, vol. 92, no. 11, pp. 2096–2107, 2011.

[9] J. Wang, F. Pan, J. Soininen, J. Heino, and J. Shen, "Nutrient enrichment modifies temperature-biodiversity relationships in large-scale field experiments," *Nature Commun.*, vol. 7, 2016, Art. no. 13960.

[10] O. Bozorg-Haddad, P. Dehghan, B. Zolghadir-Asli, V. P. Singh, X. Chu, and H. A. Loaiciga, "System dynamics modeling of lake water management under climate change," *Sci. Rep.*, vol. 12, 2022, Art. no. 5828.

[11] D. Trolle, D. P. Hamilton, C. A. Pilditch, I. C. Duggan, and E. Jeppesen, "Predicting the effects of climate change on trophic status of three morphologically varying lakes: Implications for lake restoration and management," *Environ. Model. Softw.*, vol. 26, no. 4, pp. 354–370, 2011.

[12] B. M. Kraemer et al., "Climate change drives widespread shift in lake thermal habitat," *Nature Climate Change*, vol. 11, pp. 521–529, 2021.

[13] S. Piccolroaz, S. Zhu, M. Ptak, M. Sojka, and X. Du, "Warming of lowland Polish lakes under future climate change scenarios and consequences for ice cover and mixing dynamics," *J. Hydrol., Regional Stud.*, vol. 34, 2021, Art. no. 100780.

[14] M. R. Hipsey, L. C. Bruce, and D. P. Hamilton, "GLM- General lake model: Model overview and user information," Univ. Western Aust., Perth, WA, Australia, AED Rep. 26, 2014.

[15] M. Hondzo and H. G. Stefan, "Lake water temperature simulation model," *J. Hydraulic Eng.*, vol. 119, no. 11, pp. 1251–1273, 1993.

[16] W. C. Liu and W. B. Chen, "Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models," *Comput. Geosci.*, vol. 45, pp. 13–25, 2012.

[17] J. S. Read et al., "Process-guided deep learning predictions of lake water temperature," *Water Resour. Res.*, vol. 55, no. 11, pp. 9173–9190, 2019.

[18] S. Zhu, M. Ptak, Z. M. Yaseen, J. Dai, and B. Sivakumar, "Forecasting surface water temperature in lakes: A comparison of approaches," *J. Hydrol.*, vol. 585, 2020, Art. no. 124809.

[19] A. Yousefi and M. Toffolon, "Critical factors for the use of machine learning to predict lake surface water temperature," *J. Hydrol.*, vol. 606, 2022, Art. no. 127418.

[20] S. Piccolroaz, M. Toffolon, and B. Majone, "A simple lumped model to convers air temperature into surface water temperature in lakes," *Hydrol. Earth Syst. Sci.*, vol. 17, pp. 3323–3338, 2013.

[21] E. Calamita, S. Piccolroaz, B. Majone, and M. Toffolon, "On the role of local depth and latitude on surface warming heterogeneity in the Laurentian Great Lakes," *Inland Waters*, vol. 11, no. 2, pp. 208–222, 2021.

[22] G. Flaim, D. Andreis, S. Piccolroaz, and U. Obertegger, "Ice cover and extreme events determine dissolved oxygen in a placid mountain lake," *Water Resour. Res.*, vol. 56, 2020, Art. no. e2020WR027321.

[23] M. Schmid and O. Koster, "Excess warming of a central European lake driven by solar brightening," *Water Resour. Res.*, vol. 52, no. 10, pp. 8103–8116, 2016.

[24] M. Toffolon, A. Yousefi, and S. Piccolroaz, "Estimation of the thermally reactive layer in lakes based on surface water temperature," *Water Resour. Res.*, vol. 58, no. 6, 2022, Art. no. e2021WR031755.

[25] L. Guo et al., "Responses of lake ice phenology to climate change at Tibetan Plateau," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3856–3861, 2020.

[26] M. Aizeman and A. Malishevski, "General theory of best variant choice: Some aspects," *IEEE Trans. Autom. Control*, vol. 26, no. 5, pp. 1030–1040, Oct. 1981.

[27] N. Sherborne, N. Galic, and R. Ashauer, "Sublethal effect modelling for environmental risk assessment of chemicals: Problem definitions, model variants, applications and challenges," *Sci. Total Environ.*, vol. 745, 2020, Art. no. 141027.

[28] I. Boussaid, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," *Inf. Sci.*, vol. 237, pp. 82–117, 2013.

[29] J. Del Sar et al., "Bio-inspired computation: Where we stand and what's next," *Swarm Evol. Comput.*, vol. 48, pp. 220–250, 2019.

[30] O. C. M. Ross, "A review of quantum-inspired metaheuristics: Going from classical computers to real quantum computers," *IEEE Access*, vol. 8, pp. 814–838, 2019.

[31] F. Fausto, A. Reyna-Orta, E. Cuevas, A. G. Andrade, and M. Perez-Cisneros, "From ants to whales: Metaheuristics for all tastes," *Artif. Intell. Rev.*, vol. 53, pp. 753–810, 2020.

[32] T. Wei, S. Wang, J. Zhong, D. Liu, and J. Zhang, "A review on evolutionary multi-task optimization: Trends and challenges," *IEEE Trans. Evol. Comput.*, vol. 26, no. 5, pp. 941–960, Oct. 2022, doi: 10.1109/TEVC.2021.3139437.

[33] D. H. Wolpert and W. G. MacReady, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.

[34] B. Cao, Y. Zhang, J. Zhao, X. Liu, L. Skonieczny, and Z. Lv, "Recommendation based on large-scale many-objective optimization for the intelligent Internet of Things system," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 15030–15038, Aug. 2022.

[35] A. P. Piotrowski and A. E. Piotrowska, "Differential evolution and particle swarm optimization against COVID-19," *Artif. Intell. Rev.*, vol. 55, pp. 2149–2219, 2022.

[36] F. Neri and E. Mininno, "Memetic compact differential evolution for cartesian robot control," *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 54–65, May 2010.

[37] R. K. Pachauri et al., "Impact of partial shading on various PV array configurations and different modeling approaches: A comprehensive review," *IEEE Access*, vol. 8, pp. 181375–181403, 2020.

[38] H. R. Maier et al., "Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions," *Environ. Model. Softw.*, vol. 62, pp. 271–299, 2014.

[39] S. Paterna, M. Santoni, and L. Bruzzone, "An approach based on multi-objective genetic algorithms to schedule observations in planetary remote sensing missions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4714–4727, 2020.

[40] Y. Zhong, S. Zhang, and L. Zhang, "Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2290–2301, Oct. 2013.

[41] E. K. Burke et al., "Hyper-heuristics: A survey of the state of the art," *J. Oper. Res. Soc.*, vol. 64, pp. 1695–1724, 2013.

[42] P. Kerschke, H. H. Hoos, F. Neumann, and H. Trautmann, "Automated algorithm selection: Survey and perspectives," *Evol. Comput.*, vol. 27, no. 1, pp. 3–45, 2019.

[43] J. R. Rice, "The algorithm selection problem," *Adv. Comput.*, vol. 15, pp. 65–118, 1976.

[44] Y. Tian, S. Peng, X. Zhang, T. Rodemann, K. C. Tan, and Y. Jin, "A recommender system for metaheuristic algorithms for continuous optimization based on deep recurrent neural networks," *IEEE Trans. Artif. Intell.*, vol. 1, no. 1, pp. 5–18, Aug. 2020.

[45] S. Zhu, H. R. Maier, and A. C. Zecchin, "Identification of metrics suitable for determining the features of real-world optimisation problems," *Environ. Model. Softw.*, vol. 148, 2022, Art. no. 105281.

[46] J. Chen, B. Xin, Z. Peng, L. Dou, and J. Zhang, "Optimal contraction theorem for exploration–exploitation tradeoff in search and optimization," *IEEE Trans. Syst., Man, Cybern. A, Syst. Hum.*, vol. 39, no. 3, pp. 680–691, May 2009.

[47] M. Crepinsek, S. H. Liu, and M. Mernik, "Exploration and exploitation in evolutionary algorithms: A survey," *ACM Comput. Surv.*, vol. 45, no. 3, 2013, Art. no. 35.

[48] A. E. Eiben and C. A. Schippers, "On evolutionary exploration and exploitation," *Fundamenta Informaticae*, vol. 35, no. 1–4, pp. 35–50, 1998.

[49] J. Jerebic et al., "A novel direct measure of exploration and exploitation based on attraction basins," *Expert Syst. Appl.*, vol. 167, 2021, Art. no. 114353.

[50] R. Cheng, M. Li, K. Li, and X. Yao, "Evolutionary multiobjective optimization-based multimodal optimization: Fitness landscape approximation and peak detection," *IEEE Trans. Evol. Comput.*, vol. 22, no. 5, pp. 692–706, Oct. 2018.

[51] H. Richter and A. P. Engelbrecht, *Recent Advances in the Theory and Application of Fitness Landscapes*. Berlin, Germany: Springer-Verlag, 2014.

[52] J. Brest, M. S. Maučec, and B. Bošković, "Single objective real-parameter optimization: Algorithm jSO," in *Proc. IEEE Congr. Evol. Comput.*, 2017, pp. 1311–1318.

[53] M. G. P. de Lacerda, L. F. de Araujo Pessoa, F. B. de Lima Neto, T. B. Ludermir, and H. Kuchen, "A systematic literature review on general parameter control for evolutionary and swarm-based algorithms," *Swarm Evol. Comput.*, vol. 60, 2021, Art. no. 100777.

[54] R. S. Parpinelli, G. F. Plichoski, R. S. da Silva, and P. H. Narloch, "A review of techniques for on-line control of parameters in swarm intelligence and evolutionary computation algorithms," *Int. J. Bio-Inspired Comput.*, vol. 13, no. 1, pp. 1–17, 2019.

[55] R. Tanabe and A. Fukunaga, "Reviewing and benchmarking parameter control methods in differential evolution," *IEEE Trans. Cybern.*, vol. 50, no. 3, pp. 1170–1184, Mar. 2020.

[56] A. Zamuda and J. Brest, "Self-adaptive control parameters' randomization frequency and propagations in differential evolution," *Swarm Evol. Comput.*, vol. 25, pp. 72–99, 2015.

[57] M. Zuo, G. Dai, L. Peng, Z. Tang, D. Gong, and Q. Wang, "A differential evolution algorithm with the guided movement for population and its application to interplanetary transfer trajectory design," *Eng. Appl. Artif. Intell.*, vol. 110, 2022, Art. no. 104727.

[58] T. Jansen, K. A. De Jong, and I. Wegener, "On the choice of the offspring population size in evolutionary algorithms," *Evol. Comput.*, vol. 13, no. 4, pp. 413–440, 2005.

[59] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, "Population size in particle swarm optimization," *Swarm Evol. Comput.*, vol. 58, 2020, Art. no. 100718.

[60] R. Polakova, J. Tvrdik, and P. Bujok, "Differential evolution with adaptive mechanism of population size according to current population diversity," *Swarm Evol. Comput.*, vol. 50, 2019, Art. no. 100519.

[61] R. Tanabe and A. Fukunaga, "Improving the search performance of SHADE using linear population size reduction," in *Proc. IEEE Congr. Evol. Comput.*, 2014, pp. 1658–1665.

[62] A. P. Piotrowski, M. J. Napiorkowski, J. J. Napiorkowski, and P. M. Rowinski, "Swarm intelligence and evolutionary algorithms: Performance versus speed," *Inf. Sci.*, vol. 384, pp. 34–85, 2017.

[63] K. V. Price, N. H. Awad, M. Z. Ali, and P. N. Suganthan, "The 2019 100-Digit Challenge on real-parameter, single-objective optimization: Analysis of results," Nanyang Technol. Univ., Singapore, 2019, [Online]. Available: http://www.ntu.edu.sg/home/epsugan

[64] I. Fister, J. Brest, A. Iglesias, A. Galvez, S. Deb, and I. Fister, Jr., "On selection of a benchmark by determining the algorithms' qualities," *IEEE Access*, vol. 9, pp. 51166–51178, 2021.

[65] A. La Torre, D. Molina, E. Osaba, J. Poyatos, J. del Ser, and F. Herrera, "A prescription of methodological guidelines for comparing bio-inspired optimization algorithms," *Swarm Evol. Comput.*, vol. 67, 2021, Art. no. 100973.

[66] O. Mersmann, M. Preuss, H. Trautmann, B. Bischl, and C. Weihs, "Analyzing the BBOB results by means of benchmarking concepts," *Evol. Comput.*, vol. 23, no. 1, pp. 161–185, 2015.

[67] J. Carrasco, S. Garcia, M. M. Rueda, S. Das, and F. Herrera, "Recent trends in the use of statistical tests for comparing swarm and evolutionary computing algorithms: Practical guidelines and a critical review," *Swarm Evol. Comput.*, vol. 54, 2020, Art. no. 100665.

[68] N. Vecek, M. Mernik, and M. Crepinsek, "A chess rating system for evolutionary algorithms: A new method for the comparison and ranking of evolutionary algorithms," *Inf. Sci.*, vol. 277, pp. 656–679, 2014.

[69] K. B. Cantún-Avila, D. González-Sánchez, S. Díaz-Infante, and F. Peñuñuri, "Optimizing functionals using differential evolution," *Eng. Appl. Artif. Intell.*, vol. 97, 2021, Art. no. 104086.

[70] Y. Liu, H. Ishibuchi, G. G. Yen, Y. Nojima, and N. Masuyama, "Handling imbalance between convergence and diversity in the decision space in evolutionary multimodal multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 24, no. 3, pp. 551–565, Jun. 2020.

[71] R. Miikkulainen and S. Forrest, "A biological perspective on evolutionary computation," *Nature Mach. Intell.*, vol. 3, pp. 9–15, 2021.

[72] J. de Armas, E. Lalla-Ruiz, S. L. Tilahun, and S. Voß, "Similarity in metaheuristics: A gentle step towards a comparison methodology," *Neural Comput.*, vol. 21, pp. 265–287, 2022.

[73] D. Molina, J. Poyatos, J. del Seer, S. Garcia, A. Hussain, and F. Herrera, "Comprehensive taxonomies of nature- and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations," *Cogn. Comput.*, vol. 12, pp. 897–939, 2020.

[74] K. Sorensen, "Metaheuristics – The metaphor exposed," *Int. Trans. Oper. Res.*, vol. 22, pp. 3–18, 2015.

[75] E. Osaba et al., "A tutorial on the design, experimentation and application of metaheuristic algorithms to real-world optimization problems," *Swarm Evol. Comput.*, vol. 64, 2021, Art. no. 100888.

[76] N. H. Awad, M. Z. Ali, J. J. Liang, B. Y. Qu, and P. N. Suganthan, "Problem defnitions and evaluation criteria for the CEC 2017 special session and competition on single objective real-parameter numerical optimization," Nanyang Technol. Univ., Singapore, 2016.

[77] S. Das and P. N. Suganthan, "Problem definitions and evaluation criteria for CEC 2011 competition on testing evolutionary algorithms on real world optimization problems," Jadavpur Univ., Nanyang Technol. Univ., Kolkata, India, 2010.

[78] P. Bujok, J. Tvrdik, and R. Polakova, "Comparison of nature-inspired population-based algorithms on continuous optimisation problems," *Swarm Evol. Comput.*, vol. 50, 2019, Art. no. 100490.

[79] A. P. Piotrowski, S. Zhu, and J. J. Napiorkowski, "Air2water model with nine parameters for lake surface temperature assessment," *Limnologica*, vol. 94, 2022, Art. no. 125967.

[80] A. Javaheri, M. Babbar-Sebens, and R. N. Miller, "From skin to bulk: An adjustment technique for assimilation of satellite-derived temperature observations in numerical models of small inland water bodies," *Adv. Water Resour.*, vol. 92, pp. 284–298, 2016.

[81] M. Toffolon, L. Cortese, and D. Bouffard, "SELF v1.0: A minimal physical model for predicting time of freeze-up in lakes," *Geosci. Model Develop.*, vol. 14, pp. 7527–7543, 2021.

[82] R. I. Woolway et al., "Phenological shifts in lake stratification under climate change," *Nature Commun.*, vol. 12, 2021, Art. no. 2318.

[83] P. N. Suganthan et al., "Problem definitions and evaluation criteria for the CEC 2005 special session on real-parameter optimization," IIT Kanpur, Kanpur, India, Tech. Rep. 2005005, 2005.

[84] C. Yue, B. Qu, K. Yu, J. Liang, and X. Li, "A novel scalable test problem suite for multimodal multiobjective optimization," *Swarm Evol. Comput.*, vol. 48, pp. 62–71, 2019.

[85] S. Zhu, A. P. Piotrowski, M. Ptak, J. J. Napiorkowski, J. Dai, and Q. Ji, "How does the calibration method impact the performance of the air2water model for the forecasting of lake surface water temperatures?," *J. Hydrol.*, vol. 597, 2021, Art. no. 126219.

[86] J. Kennedy and R. C. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4, 1995, pp. 1942–1948.

[87] S. Piccolroaz et al., "On the predictability of lake surface temperature using air temperature in a changing climate: A case study for Lake Tahoe (USA)," *Limnol. Oceanogr.*, vol. 63, pp. 243–261, 2018.

[88] M. Clerc and J. Kennedy, "The particle swarm – explosion, stability, and convergence in multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[89] Y. Shi and R. C. Eberhart, "A modified particle swarm optimizer," in *Proc. IEEE Congr. Evol. Comput.*, 1998, pp. 69–73.

[90] X. Xia et al., "Triple archives particle swarm optimization," *IEEE Trans. Cybern.*, vol. 50, no. 12, pp. 4862–4875, Dec. 2020.

[91] Z. Meng and J.-S. Pan, "HARD-DE: Hierarchical archive based mutation strategy with depth information of evolution for the enhancement of differential evolution on numerical optimization," *IEEE Access*, vol. 7, pp. 12832–12854, 2019.

[92] A. A. Hadi, A. W. Mohamed, and K. M. Jambi, "Single-objective real-parameter optimization: Enhanced LSHADE-SPACMA algorithm," *Heuristics Optim. Learn. Stud. Comput. Intell.*, vol. 906, pp. 103–121, 2021.

[93] A. W. Mohamed, A. A. Hadi, A. M. Fattouh, and K. M. Jambi, "LSHADE with semi-parameter adaptation hybrid with CMA-ES for solving CEC 2017 benchmark problems," in *Proc. IEEE Congr. Evol. Comput.*, 2017, pp. 145–152, doi: 10.1109/CEC.2017.7969307.

[94] A. P. Piotrowski and J. J. Napiorkowski, "Step-by-step improvement of JADE and SHADE-based algorithms: Success or failure?," *Swarm Evol. Comput.*, vol. 43, pp. 88–108, 2018.

[95] M. Pant, H. Zaheer, L. Garcia-Hernandez, and A. Abraham, "Differential evolution: A review of more than two decades of research," *Eng. Appl. Artif. Intell.*, vol. 90, 2020, Art. no. 103479.

[96] A. P. Piotrowski, "L-SHADE optimization algorithms with population-wide inertia," *Inf. Sci.*, vol. 468, pp. 117–141, 2018.

[97] T. Vu Dinh, H. Nguyen, X. L. Tran, and N. D. Hoang, "Predicting rainfall-induced soil erosion based on a hybridization of adaptive differential evolution and support vector machine classification," *Math. Problems Eng.*, vol. 2021, 2021, Art. no. 6647829.

[98] R. Bandyopadhyay, R. Kundu, D. Oliva, and R. Sarkar, "Segmentation of brain MRI using an altruistic Harris Hawks' optimization algorithm," *Knowl.-Based Syst.*, vol. 232, 2021, Art. no. 107468.

[99] M. Zuo, G. Dai, L. Peng, M. Wang, Z. Liu, and C. Chen, "A case learning-based differential evolution algorithm for global optimization of interplanetary trajectory design," *Appl. Soft Comput.*, vol. 94, 2020, Art. no. 106451.

[100] J. Brest, S. Greiner, B. Boskovic, M. Mernik, and V. Zumer, "Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems," *IEEE Trans. Evol. Comput.*, vol. 10, no. 6, pp. 646–657, Dec. 2006.

[101] R. Polakova and P. Bujok, "Popular optimisation algorithms with diversity-based adaptive mechanism for population size," *Recent Adv. Soft Comput. Cybern., Stud. Fuzziness Soft Comput.*, vol. 403, pp. 171–182, 2021.

[102] S. Aras, E. Gedikli, and T. H. Kahraman, "A novel stochastic fractal search algorithm with fitness-distance balance for global numerical optimization," *Swarm Evol. Comput.*, vol. 61, 2021, Art. no. 100821.

[103] H. T. Kahraman, S. Aras, and E. Gedikli, "Fitness-distance balance (FDB): A new selection method for meta-heuristic search algorithms," *Knowl.-Based Syst.*, vol. 190, 2020, Art. no. 105169.

[104] Q. B. Diep, I. Zelinka, S. Das, and R. Senkerik, "SOMA T3A for solving the 100-digit challenge," in *Proc. Int. Conf. Swarm, Evol. Memetic Comput. Conf.*, 2019, pp. 155–165.

[105] I. Zelinka, "SOMA – Self-organizing migrating algorithm," in *New Optimization Techniques in Engineering. Studies in Fuzziness and Soft Computing*, vol. 141, Berlin, Heidelberg: Springer, 2004. [Online]. Available: https://doi.org/10.1007/978-3-540-39930-8_7

[106] J. Brest, M. S. Maucec, and B. Boskovic, "The 100-Digit Challenge: Algorithm jDE100," in *Proc. IEEE Congr. Evol. Comput.*, 2019, pp. 19–26, doi: 10.1109/CEC.2019.8789904.

[107] S.-M. Guo, J. S.-H. Tsai, C.-C. Yang, and P.-H. Hsu, "A self-optimization approach for L-SHADE incorporated with eigenvector-based crossover and successful-parent-selecting framework on CEC 2015 benchmark set," in *Proc. IEEE Congr. Evol. Comput.*, 2015, pp. 1003–1010.

[108] N. H. Awad, M. Z. Ali, and P. N. Suganthan, "Ensemble sinusoidal differential covariance matrix adaptation with Euclidean neighborhood for solving CEC2017 benchmark problems," in *Proc. IEEE Congr. Evol. Comput.*, 2017, pp. 372–379, doi: 10.1109/CEC.2017.7969336.

[109] M. Zuo, G. Dai, and L. Peng, "A new mutation operator for differential evolution algorithm," *Soft Comput.*, vol. 25, pp. 13595–13615, 2021.

[110] A. Kumar, P. P. Biswas, and P. N. Suganthan, "Differential evolution with orthogonal array-based initialization and a novel selection strategy," *Swarm Evol. Comput.*, vol. 68, 2022, Art. no. 101010.

[111] P. P. Biswas, P. N. Suganthan, and G. A. J. Amaratunga, "Minimizing harmonic distortion in power system with optimal design of hybrid active power filter using differential evolution," *Appl. Soft Comput.*, vol. 61, pp. 486–496, 2017.

[112] M. Zuo, G. Dai, L. Peng, M. Wang, Z. Liu, and C. Chen, "A case learning-based differential evolution algorithm for global optimization of interplanetary trajectory design," *Appl. Soft Comput.*, vol. 94, 2020, Art. no. 106451.

[113] S. Garcia and F. Herrera, "An extension on "Statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons," *J. Mach. Learn. Res.*, vol. 9, pp. 2677–2694, 2019.

[114] A. Ulas, O. T. Yildiz, and E. Alpaydin, "Cost-conscious comparison of supervised learning algorithms over multiple data sets," *Pattern Recognit.*, vol. 45, pp. 1772–1781, 2012.

Adam P. Piotrowski received the Ph.D. and Habilitation degrees in earth sciences from the Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland, in 2005 and 2014, respectively.

He is currently working on evolutionary computation, neural networks, thermal aspects of natural water bodies, and rainfall-runoff modeling. He is the first author of more than 40 journal papers, and is currently an Associate Editor of *Swarm and Evolutionary Computation* journal.

Jaroslaw J. Napiorkowski received the M.Sc. degree in electronics from the Department of Electronics, Warsaw University of Technology, Warsaw, Poland, and the Ph.D. degree in earth sciences from the Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland.

He was a postdoctoral scholar at the University College Dublin and as Humboldt Fellow at the Ruhr University Bochum. Since 2004, he has been the Head of Department of Hydrology and Hydraulics at the Institute of Geophysics Polish Academy of Sciences. Dr. Napiorkowski was the winner of the Tison Award.

Senlin Zhu received the Ph.D. degree in hydraulics and river dynamics from Wuhan University, Wuhan, China, in 2017.

He is currently the Assistant Professor with the College of Hydraulic Science and Engineering, Yangzhou University, Yangzhou, China. His current projects are river/lake environment monitoring and modeling and climate change on aquatic system. His research interests include climate change, water quality modeling, and hydrological modeling.