COPYRIGHT NOTICE

FedUni ResearchOnline
http://researchonline.federation.edu.au

This is the published version of:

Moore, W., et al. (2016) Comparison of multiple surrogates for 3D CFD model in tidal farm optimisation. *Procedia Engineering*, 154(2016), 1132-1139.

Available online at http://doi.org/10.1016/j.proeng.2016.07.523

Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 154 (2016) 1132 – 1139

Procedia Engineering

www.elsevier.com/locate/procedia

12th International Conference on Hydroinformatics, HIC 2016

Comparison of multiple surrogates for 3D CFD model in tidal farm optimisation

William Ogaday Willers Moore[a,*], Helena Mala-Jetmarova[a,b], Mulualem Gebreslassie[c], Gavin R. Tabor[a], Michael R. Belmont[a], Dragan A. Savic[a]

[a]*College of Engineering, Mathematics and Physical Sciences, University of Exeter, Streatham Campus, North Park Road, Exeter, Devon EX4 4QF, United Kingdom*

[b]*Faculty of Science and Technology, Federation University Australia, Mt Helen Campus, University Drive, Ballarat, Victoria 3350, Australia*

[c]*School of Mechanical and Industrial Engineering, Mekelle University, Endayesus Campus, PO Box 3086, Mekelle, Tigray, Ethiopia*

Abstract

Marine currents have been identified as a considerable renewable energy source. Therefore, in recent years, research on optimising tidal stream farm layouts in order to maximise power output has emerged. Traditionally, computational fluid dynamics (CFD) models are used to model power output, but their computational cost is prohibitive within an optimisation algorithm. This paper uses surrogate models in place of CFD simulations to optimise the layout of tidal stream farm layouts. Surrogates are functions which are designed to emulate the behaviour of other models with radically reduced computational expense. Two surrogate models are applied and compared: artificial neural network (ANN) and k-nearest neighbours regression (k-NN). We measure their suitability by four criteria: accuracy, efficiency, robustness and performance within an optimisation algorithm. The results reveal that the ANN surrogate is superior in every criteria to the k-NN surrogate. However, the k-NN surrogate is also able to perform adequate optimisation. Finally, we demonstrate that optimisation relying solely on surrogate models is a viable approach, with dramatically reduced computational expense of optimisation.

Peer-review under responsibility of the organizing committee of HIC 2016

Keywords: Tidal stream farm layout; optimisation; surrogate model; artificial neural network; k-nearest neighbours regression

* Corresponding author. Tel.: +44-7724-739-513.
E-mail address: O.WillersMoore@exeter.ac.uk

Peer-review under responsibility of the organizing committee of HIC 2016
doi:10.1016/j.proeng.2016.07.523

1. Introduction

Energy extraction from tidal streams has been the subject of increased research efforts over the past few years. To extract an economically viable amount of energy, hundreds of tidal turbines are required in arrays within a tidal farm [1]. In prior research (for example, [2, 3]), when optimising the layouts of such tidal farms in order to maximise power output, computational fluid dynamics (CFD) models have been used to calculate the energy yield and to understand flow interactions between individual turbines.

However, these CFD simulations may become computationally prohibitive especially as the fidelity of the model and the number of devices increase. Two broad approaches are currently applied to reduce computational cost. Firstly, pseudo elements are used to represent turbines in the model, which reduce computational burden yet are still able to satisfactory reproduce flow structures. They include, for example, actuator discs [4, 5], friction elements [1, 2] and frozen rotors [3]. Secondly, simplified CFD models are applied, which reduce dimensionality of complex three-dimensional (3D) flows [1, 6] and/or linearize governing flow equations [2]. In spite of the above measures, CFD simulations still represent a significant computational expense, even for small arrays of turbines.

An alternative approach is proposed in this paper, which replaces 3D CFD simulations in an optimisation algorithm with a surrogate model. Surrogate models (metamodels) were introduced by Blanning in 1975 [7] and their foremost advantage is a substantially reduced computational cost in comparison with simulation models such as CFD models. Surrogate models do not approximate the system's mechanisms. Instead, they model the relationship between the inputs (i.e. decision/control variables) and the outputs (i.e. state variables) of the original model [8]. Surrogate models can subsequently be used as objective functions in order to calculate a solution's “fitness” in place of the original model in an optimisation algorithm.

This paper is believed to be one of the first applications of the surrogate model to replace the 3D CFD simulation in tidal farm optimisation. Two different types of surrogate model are compared: an artificial neural network (ANN) and k-nearest neighbours regression (k-NN). These models are used because they have been shown to provide good performance in other applications with computationally intensive objective function evaluation [8, 9]. These surrogate models are trained on limited data produced by a 3D CFD model and compared according to four criteria: their (i) accuracy, (ii) efficiency, (iii) robustness and (iv) performance within an optimisation algorithm.

2. Turbine and tidal stream farm layout

The Momentum-Reversal-Lift (MRL) turbine [10], which is in the prototype stage, is used (Fig. 1). The MRL turbine has been designed for tidally reverse current conditions (i.e. estuaries). The turbine rotates around its horizontal axis and has three horizontally oriented blades which rotate around both turbine's primary axis and their own blade axes. The diameter (D) of the prototype turbine is 0.2 m and length (L) is 0.3 m. In relation to depth of stream flow, the turbine is designed to operate while floating below water level and anchored to the estuary bed.

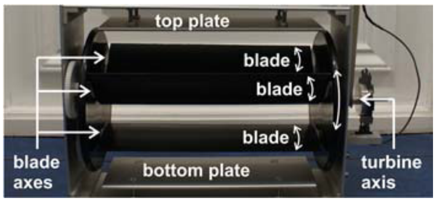

Fig. 1. Momentum-Reversal-Lift turbine.

A staggered tidal farm layout (Fig. 2) is used as it is hypothesised to produce higher power output [11]. A cluster of only four turbines arranged in three rows is represented in a 3D CFD model. This limits the simulation time so it is possible to generate enough data in a reasonable time frame. The model represents turbine T_1 in the first row, half turbines T_2 and T_3 in the second row and turbine T_4 in the third row. The model was developed as part of previous work [11, 12]. It implements a symmetry boundary, which intersects turbines T_2 and T_3, to potentially expand the number of devices in the lateral direction beyond the current CFD model domain.

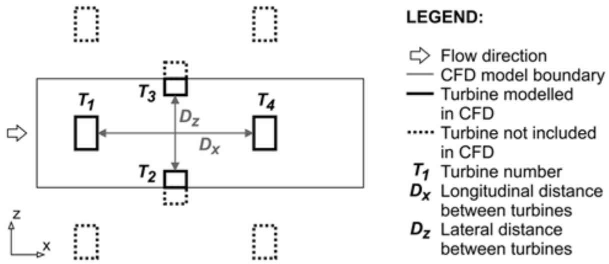

Fig. 2. Layout of tidal farm CFD modelling domain.

3. Optimisation problem

3.1. Decision variables

The optimisation problem includes five continuous decision variables, F_b (N) for $b \in \{1, 2, 3\}$, D_x (m) and D_z (m). The variables F_b are the body forces on the turbines in the b^{th} row. The body force is the defining parameter of a pseudo element in the CFD model which represents the resistance of each turbine on the flow. The turbines T_2 and T_3 are on the same row and therefore have identical body forces: F_2. D_x and D_z are distances between the turbines, in the longitudinal and lateral direction, respectively (see Fig. 2).

3.2. Constraints

There are two classes of constraints. The first class of constraint is defined by fluid dynamics laws such as conservation of mass of flow (Eq. 1) and conservation of momentum (Eq. 2). These constraints are controlled by the 3D CFD model. The second class represents constraints on the decision variables (Eq. 3-5). These limit the domain of the solution space according to lower and upper bounds based on previous work [11, 12]

$$\frac{\partial \bar{u}_i}{\partial x_i} = 0 \tag{1}$$

$$\frac{\partial \bar{u}_i}{\partial t} + \frac{\partial}{\partial x_j}\left(\bar{u}_i \bar{u}_j\right) = -\frac{1}{\rho}\left(\frac{\partial \bar{p}}{\partial x_i} + \delta_{i1}\frac{\partial \langle P \rangle}{\partial x_1}\right) + 2\nu \frac{\partial}{\partial x_j}\bar{S}_{ij} - \frac{\partial \tau_{ij}}{\partial x_j} + \rho g_i + \bar{F}_b \tag{2}$$

$$5\,N \le F_b \le 45\,N; \quad F_b \in \mathbb{R} \tag{3}$$

$$2\,m \le D_x \le 4.6\,m; \quad D_x \in \mathbb{R} \tag{4}$$

$$0.45\,m \le D_z \le 0.9\,m; \quad D_z \in \mathbb{R} \tag{5}$$

where the bar $\overline{(.)}$ defines the resolved scales; $\bar{u}$ is the filtered velocity; ρ is the fluid density; $\bar{p}$ is the filtered pressure; δ_{i1} is the Kronecker-delta; $\partial\langle P \rangle / \partial x_1$ is a constant streamwise pressure gradient; ν is a kinematic viscosity; $\bar{S}_{ij}$ is the strain rate of the resolved scales; τ_{ij} is the sub-grid scale Reynolds stress; g is the acceleration due to gravity; $\bar{F}_b$ is the body force; D_x and D_z are the longitudinal and lateral distances between turbines, respectively.

3.3. Objective function

The objective, which is to be maximised, represents a total power P_T (W) extracted by all turbines from water flow, and is written as:

$$F(F_b, D_x, D_z) = P_T = P_{T1} + P_{T2} + P_{T3} + P_{T4} \tag{6}$$

where P_{T1}, P_{T4} and P_{T2}, P_{T3} *(W)* is the power extracted by turbines T_1, T_4 and half turbines T_2, T_3, respectively.

3.4. Formulation of optimisation problem

The final single-objective optimisation problem is formulated as maximisation of a total power extracted by all four turbines from water flow as follows:

$$\begin{array}{c}\max\{F(F_b, D_x, D_z)\}\\ subject\ to\ (1)-(5).\end{array} \tag{7}$$

4. Solution methodology

The methodology to solve an optimisation problem (Eq. 7) is schematised in Fig. 3. The optimisation algorithm calls on the surrogate model to evaluate the decision variables and approximate the state variables (power output) which is fed to the objective function. The novelty of this method is that it relies solely on the surrogate model within the optimisation loop. This approach is in contrast to previous studies [8, 13], which use the original simulation model to continually update the surrogate model as optimisation progresses. Therefore, our surrogate models 'stand alone' and are not retrained according to the original 3D CFD model during optimisation. The benefit of this approach is enormously increased speed of optimisation.

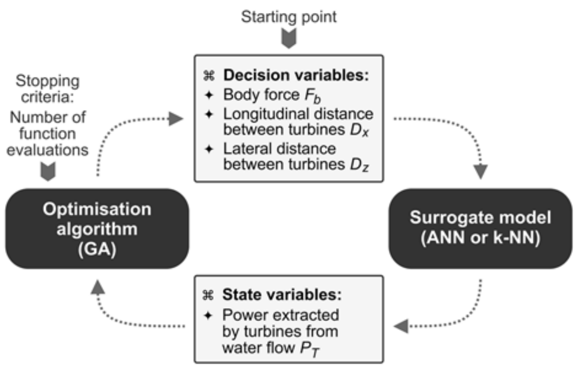

Fig. 3. Solution scheme.

4.1. Genetic algorithm (GA)

GAs are a class of widely applied [14] stochastic optimisation techniques which find an optimal solution by simulating the process evolution. They use a population of solutions, a selection criteria based upon fitness and random variation to create a survival of the fittest process over multiple iterations. After tuning, the parameters were selected. Ultimately, the GA was run with population size of 200, probability of two point crossover of 0.8, and probability of mutation of 0.1. The mutation operator used was Gaussian variation with a mean of zero and standard deviation of 0.1, and was limited by the constraints in decision variables (Eq. 3-5). The algorithm was set to run for 60 iterations, but the optimisation often converged long before that limit. The solution space was technically infinitely large, as the decision variables were continuous. However, changes in the variables with magnitudes of order 10^{-2} do not affect the outcome of the simulation drastically. Finally, the optimisation was run 30 times for each surrogate in order to reduce the impact of the initial population on the optimal solution (the same random seeds were used per run for each surrogate). The GA was implemented in Python using the DEAP library [15].

4.2. Surrogate models

Two surrogate models, ANN and k-NN, are used to approximate results of the 3D CFD model of a tidal farm. We use the CFD model built previously by [11, 12] in OpenFOAM to generate the data needed to construct the surrogate models. The data was bounded and feature scaled using the constraints on the decision variables given by equations (3) to (5). The surrogate models were designed to produce one output value as defined by the objective function, which is a total power P_T (W) extracted by all turbines from water flow.

Running the 3D CFD model is computationally costly, with each simulation taking more than three days on facility hardware, so the final data set consists of only 329 data samples. For the purpose of surrogate model development, this data set was subsequently divided into training and validation (containing 300 data samples), and testing (containing 29 data samples) subsets. Part of the data set used was generated for previous research [13] and the majority was generated according to Latin Hypercube Sampling. However, due to difficulty of working with the model, most inputs were discrete values. A necessary consideration when dealing with surrogate models is that the original model is not perfect, and that optimisation might push the search into areas for which the model was not calibrated.

4.2.1. Artificial neural network (ANN)

ANNs have been proved to be able to successfully approximate computationally intensive simulation models in a variety of water resource applications [16]. They approximate functions based upon a principle of real neurological structures and can be represented as directed graphs including an input layer, a number of hidden layers and an output layer. The ANN surrogate was implemented as a feedforward multilayer perceptron in Python using the Keras library [17]. After systematic search, it was decided to use one hidden layer with seven neurons with the Tanh activation function.

ANN training is essentially an optimisation problem to find the optimal weights in order to reduce the error in its approximation of a CFD model. Therefore, mean square error (MSE) was minimised in training. In this paper, full batch training for 10,000 epochs and a gradient-based optimisation algorithm ADAM [18] were used. ADAM outperformed other popular algorithms, such as stochastic gradient descent due to its ability to manage noisy data.

4.2.2. K-nearest neighbors (k-NN)

K-NN regression is a simple non-parametric approach to regression [19]. K-NN considers the closest k neighbours in the training set to a point whose value is to be approximated. The point is evaluated to be the sum of the values of the k-nearest points. A Python implementation in Scikit-Learn [20] was used. The value for k was set at 2, training data and input values were sphered to decorrelate the variables and the k-closest points were weighted by distance in order to improve approximation.

4.3. Comparison of surrogate models

The surrogate models are compared across four criteria: their (i) accuracy, (ii) efficiency, (iii) robustness and (iv) performance within an optimisation algorithm. These criteria are specified by the following metrics:

- Accuracy is a measure of the error in approximations of the test data by the surrogate model. MSE and the coefficient of determination (R^2) are used as metrics to compare accuracy of surrogate models.
- Efficiency is a measure of the ability of the surrogate model approximate state variables quickly. The average evaluation time for 1,000 evaluations and the average optimisation time are used as metrics to compare the efficiency.
- Robustness reflects the ability of a surrogate model to perform in a stable manner under various circumstances (such as different initial populations) within an optimisation algorithm. The standard deviation in the power output of the optimal solution over 30 optimisation runs is used to compare the robustness of the surrogate models.
- Performance within an optimisation algorithm measures the ability of a surrogate model to reach an optimal solution. This criteria compares the best solution found by each surrogate within optimisation.

5. Results

The results for comparing the surrogate models are contained in Table 2. With respect to the first criteria, accuracy, the ANN surrogate outperforms k-NN for approximating the results of the CFD model in both metrics (i) MSE (lower is better) and (ii) R^2 (closer to 1 is better). These results are supported by Fig. 4, which compares the power output simulated by the CFD model and the power output approximated by the surrogates. In this figure, points closer to the diagonal line demonstrate the ability of the surrogate to represent the CFD simulation more accurately. Fig. 4(a) and Fig. 4(c) show the approximations on the training data for ANN and k-NN, respectively, when performing 10-fold cross validation. Fig. 4(b) and Fig. 4(d) show the approximations on the test solutions by the ANN and k-NN surrogates, respectively.

Similarly for efficiency, ANN outperforms k-NN as both metrics (i) time of 1,000 approximations and (ii) time to find an optimal solution have lower values for the ANN than the k-NN surrogate model. Advantageously, both times for ANN and k-NN are many orders of magnitude faster than the CFD simulation run times, which dramatically reduces the computational expense of the optimisation. It is worth noting that the k-NN surrogate is faster to train than the ANN surrogate. Because both models can be trained in a matter of seconds, this point should not overly influence a preference for one surrogate over the other.

Concerning robustness, standard deviation shows a flaw in the k-NN surrogate model. Once a solution reaches a certain area of the solution space, it is assigned the fitness of the closest point. This means that the optimisation algorithm is unable to discriminate between multiple solutions within the same region. Although robustness is compared according to "lower is better", it is evident that the k-NN surrogate is less robust because it is unable to converge on an identical solution, unlike ANN.

With regard to performance, while the k-NN model predicts a higher power output for its optimal solution, the power output is identical to the best solution in the training set. That means, in conjunction with the superior accuracy, that the optimal solution found by ANN is much more reliable. Hence, without further validation, the ANN surrogate is better in performance than the k-NN surrogate.

In summary, ANN is a superior surrogate to k-NN in every metric and criteria.

Table 2. Comparison of surrogate models.

Criteria →	Accuracy		Efficiency		Robustness	Performance
Metrics →	MSE	R^2	Time of 1,000 approximations (s)*	Time to find an optimal solution (s)**	Standard deviation	Optimal solution found P_T (W)***
Artificial neural network (ANN)	3.465	0.928	0.002	0.236	0.185	36.846
k-nearest neighbours (k-NN)	9.236	0.839	0.192	0.590	0.065	37.465

Note: *Average over 100 times. **Average over 30 optimisation runs. ***The best solution found in 30 optimisation runs.

6. Discussion

We have demonstrated that optimisation relying solely on surrogate models (in place of CFD simulations) is a viable approach. The benefits are that after the initial training of a surrogate model, optimisation is not gated by the simulation time of a CFD, thus the optimisation itself is trivially fast compared to one CFD simulation. Furthermore, because the training data can be produced independently of itself, the training simulations can be run completely in parallel. This is in contrast to an optimisation process requiring validation within each generation, because each validation is dependent on previous generations, so it is strictly sequential with optimisation time growing as a product of CFD simulation time and the number of generations. Another advantage is that unlike in previous studies [13] the optimisation process used is completely automated, which is not possible when using the CFD model for validation as it requires manual validation.

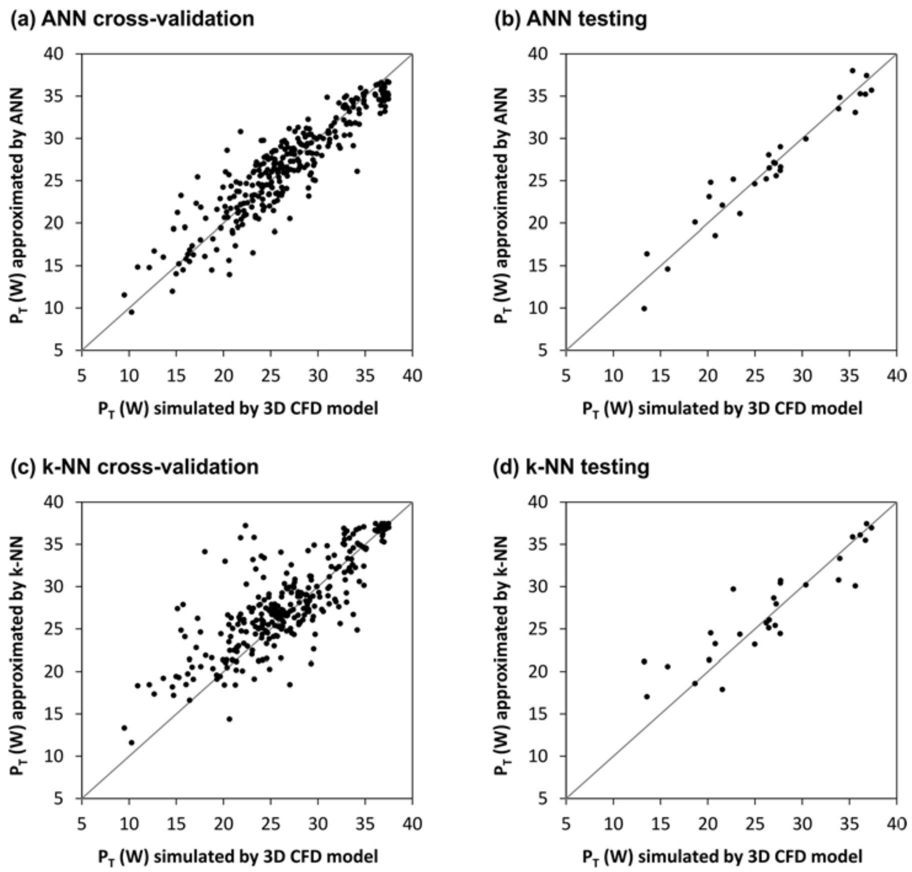

Fig. 4. Approximation of total power extracted P_T (W) by surrogate models.

The drawbacks of this approach are mostly related to the accuracy of the surrogate models. By definition, the GA is not finding the optima in the CFD simulation, but in the ANN or k-NN model. Correspondingly, accuracy has a direct impact on the ability of the surrogates to perform well.

The ANN surrogate model was identified as superior in every criteria to the k-NN surrogate model. Most importantly, the optimal solution found by ANN was the most reliable. The disadvantage of k-NN is that it is unable to discriminate effectively between different solutions. In contrast, the ability of the ANN surrogate to do so allows the optimisation to converge towards an optimal solution, as each genetic variation, through mutation or crossover, creates a change in fitness.

Further research could include an application to a more realistic case study. However, a new CFD model would be required. Currently, we use a model which represents a cell of turbines with mirrored boundary conditions in the lateral direction. It also has fixed water velocity and constant depth. An advanced CFD model could include velocities matching the tidal range, real estuary bathymetry and the capability to represent the whole farm. Then the ability of the proposed methodology using surrogate models could be evaluated in a more industry applicable situation. The authors also recommend future studies analyse the use multiple objectives to diversify the selection and produce more varied solutions.

7. Conclusion

This paper compared two surrogate models, ANN and k-NN, by their ability to replace 3D CFD simulations in optimisation of tidal stream farm layout. The optimisation problem maximised the total power extracted by turbines from water flow. The solution methodology integrated an optimisation algorithm (GA) with a surrogate model, either ANN or k-NN, which were subsequently compared. Due to computational expense of a 3D CFD model, the surrogate models were developed using a limited amount of data generated a priori by the CFD model.

Two surrogate models were compared by four criteria: their accuracy, efficiency, robustness and performance within an optimisation algorithm. This comparison revealed that the ANN surrogate model is superior in each criteria to the k-NN surrogate model, because it is able to approximate a 3D CFD model more accurately, efficiently and robustly. Furthermore, the ANN surrogate consistently converged on an optimal solution while k-NN was unable to discriminate between similar solutions. Ultimately, we demonstrated that optimisation relying solely on surrogate models is a viable approach, with dramatically reduced computational expense per optimisation iteration.

Recommendations for future research include an application of the proposed methodology using surrogate models (in place of CFD simulations) to a more realistic case study and implementation of multiple objectives into an optimisation model.

Acknowledgements

Authors acknowledge Engineering and Physical Sciences Research Council (EPSRC) in the UK, which provided funding for this work under project number EP/J010138/1.

References

[1] Funke, S. W., Farrell, P. E. and Piggott, M. Tidal Turbine Array Optimisation Using the Adjoint Approach, Renewable Energy. 63 (2014) 658-673.
[2] Divett, T., Vennell, R. and Stevens, C. Optimization of Multiple Turbine Arrays in a Channel with Tidally Reversing Flow by Numerical Modelling with Adaptive Mesh. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. 371 (2013) 20120251.
[3] Lee, S. H., Lee, S. H., Jang, K., Lee, J. and Hur, N. A Numerical Study for the Optimal Arrangement of Ocean Current Turbine Generators in the Ocean Current Power Parks, Current Applied Physics. 10 (2010) S137-S141.
[4] Bahaj, A., Myers, L., Thomson, M. and Jorge, N. Characterising the Wake of Horizontal Axis Marine Current Turbines, 7th European Wave and Tidal Energy Conference (EWTEC), Porto, Portugal, 2007.
[5] Roc, T., Conley, D. C. and Greaves, D. Methodology for Tidal Turbine Representation in Ocean Circulation Model, Renewable Energy. 51 (2013) 448-464.
[6] Vennell, R. Tuning Tidal Turbines in-concert to Maximise Farm Efficiency, Journal of Fluid Mechanics. 671 (2011) 587-604.
[7] Blanning, R. W. The Construction and Implementation of Metamodels, Simulation. 24 (1975) 177-184.
[8] Broad, D. R., Maier, H. R. and Dandy, G. C. Optimal Operation of Complex Water Distribution Systems Using Metamodels, Journal of Water Resources Planning and Management, ASCE. 136 (2010) 433-443.
[9] Ostfeld, A. and Salomons, S. A Hybrid Genetic-Instance Based Learning Algorithm for CE-QUAL-W2 Calibration, Journal of Hydrology. 310 (2005) 122-142.
[10] Janssen, A. Fluid Turbine. Patent US20100278647 A1, http://www.google.com/patents/US20100278647 (accessed on 9/10/2015), 2010.
[11] Gebreslassie, M. G., Tabor, G. R. and Belmont, M. R. Numerical Simulation of a New Type of Cross Flow Tidal Turbine Using OpenFOAM– Part II: Investigation of Turbine-to-Turbine Interaction, Renewable Energy. 50 (2013) 1005-1013.
[12] Gebreslassie, M. G., Tabor, G. R. and Belmont, M. R. Numerical Simulation of a New Type of Cross Flow Tidal Turbine Using OpenFOAM– Part I: Calibration of Energy Extraction, Renewable Energy. 50 (2013) 994-1004.
[13] Guidolin, M., Gebreslassie, M. G., Williamson, D., Tabor, G. R., Belmont, M. R. and Savic, D. A. Toward the Optimal Design of a Tidal Farm Layout Using a CFD Model. Renewable Energy (Submitted in December 2015), 2016.
[14] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. and Zechman, E. State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management, Journal of Water Resources Planning and Management, ASCE. 136 (2010) 412-432.
[15] Fortin, F.-A., De Rainville, F.-M., Gardner, M.-A., Parizeau, M. and Gagne, C. DEAP: Evolutionary Algorithms Made Easy, Journal of Machine Learning Research. 13 (2012) 2171-2175.
[16] Maier, H. R. and Dandy, G. C. Neural Networks for the Prediction and Forecasting of Water Resources Variables: A Review of Modelling Issues and Applications, Environmental Modelling & Software. 15 (2000) 101-124.
[17] Chollet, F. Keras. GitHub Repository, GitHub, https://github.com/fchollet/keras (accessed on 28/02/2016), 2015.
[18] Kingma, D. and Ba, J. Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations (ICLR 2015), San Diego, Computing Research Repository, arXiv:1412.6980, http://arxiv.org/pdf/1412.6980.pdf (accessed on 16/02/2016), 2015.
[19] Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, The American Statistician. 46 (1992) 175-185.
[20] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825-2830.