Neural networks as surrogate models for simulating a hybrid water-energy system

Extended abstract

1. Introduction

The dynamically changing energy field, as well as the ever increasing water and energy demand, pose a strong need for developing new methods that are capable of managing them effectively. In the last decades, economy stresses and environmental needs have orientated the energy sector towards the use of renewable energy resources and the reduction of fossil fuel use. In addition, the Directive 2009/28/EC promotes the use of energy from renewable resources making their penetration to the energy mix even more necessary. Due to the unpredictable nature of the stochastic meteorological processes (i.e., solar radiation, wind speed etc.), the renewable energy sources are characterized by inherent uncertainty. This fact results in significant fluctuations of the produced energy, and therefore instability to the energy grid; moreover, it poses an obstacle to the effective penetration of renewable energy resources.

Pumped-storage systems, as components of broader hybrid water - energy systems, are considered as one of the best available techniques for the mitigation of the renewable energy resources uncertainty and the fulfillment of both water and energy demands. These schemes play a key role to the regulation of the energy balance of electric systems, as they provide the means for storing the excess of the produced energy and making it available during the demand peaks. Therefore, most of uncertainties concerning renewable energy production are drastically lessened. What is necessary for the operation of the pumped-hydroelectric-energy-storage is the use of interconnected water storage infrastructures (i.e. tanks, reservoirs) in different altitudes in order to form a scheme aiming at the integrated management of water and energy fluxes, contributing to better meeting the corresponding needs.

The management of pumped – storage systems is implemented through the simulation and optimization of their operational characteristics so that their optimum efficiency is achieved. During the simulation – optimization procedure, synthetic time series accounting for the stochastic regime of the physical system are used, the length of which is long enough to obtain the desirable accuracy for the uncertainty assessments. Given that the meteorological processes as well as energy demand are characterized by significant fluctuations during small time periods, the typical time step of the analysis of hybrid water – energy systems is that of the hour. The complexity of the combined management of water and energy fluxes, the hourly time step of the computations and the long horizon of the time series used introduce a highly demanding and challenging computational burden to the simulation and optimization procedures.

The present work approaches the aforementioned challenge through the use of artificial neural networks as surrogate models (metamodels) in an attempt to reduce the computational time of the simulation procedure. Simulation through surrogate modeling refers to the development and use of models that are faster in computations and surrogate/substitute the slower analytical simulation models by imitating their behavior without being so computationally "expensive" as them. In the literature, many applications of surrogate modeling in water resources can be found, meanwhile a limited number of them refer to reservoir management applications. Generally, in many of them, the role of the surrogate model is held by neural networks. However, there was not found any application of artificial neural networks (ANNs) concerning reservoir management.

In the current work, ANNs are used for the simulation of the response surface of the energy balance as it is reformed after the pump – storage procedure of the system. The results showed that, through the use of neural networks as surrogate models, a significant amount of the computational effort and time can be saved, thus triggering the investigation of their use in further applications.

2. Case study

This methodology is applied to a hypothetical hybrid water-energy system situated at the non-connected Aegean island of Astypalaia in Greece. The present work assumes that the existing infrastructure of the Livadi reservoir, which is used for drinking water supply and irrigation, also produces hydroelectric energy through a hypothetical power station at the end of the abstraction tube. The system is integrated with a hypothetical upstream tank, as well as with photovoltaic panels and wind farms of installed capacity equal to 0.5 and 1.0 MW, respectively. The tank is connected with the reservoir through a penstock with a reversible pump-turbine that can either produce hydroelectric energy or consume energy via pumping. At each time step, only one of the two operational modes is allowed. This allows for regulating the energy balance and enabling the infrastructure to serve multipurpose uses, i.e. water supply for domestic use, irrigation of four basic cultivations (orchards, arable land, vegetables and vineyards) and energy production.

The system is assumed to have the following characteristic values: maximum dam height, which is also the maximum operational height of the reservoir, 32 m, overall and storage capacity of the reservoir 1,050,000 m^3 and 875,000 m^3 respectively, reservoir area 105,000 m^2, flow capacity of the turbine 1,000 m^3/h, performance capacity of the turbine 0.85. Additionally, the tank is situated at 200 m altitude higher than the reservoir and is of 50,000 m^3 storage capacity. The reversible turbine is of 1,500 m^3/h flow capacity and their efficiencies are 0.85 for energy production and 0.80 for pumping.

3. Input Data

As explained next, for the modelling of the hybrid water-energy system we used both historical and synthetic data, of 7 and 100 years length, respectively, provided by recent research studies in the same area.

The synthetic data were produced by employing different stochastic methods, depending on the examined process. In particular, the synthetic rainfall data were produced through the Castalia software that preserves the statistical properties of the parent historic data across three temporal scales (daily, monthly, annual) and also reproduces the long-term persistence (Hurst-Kolmogorov dynamics) at the annual and over-annual scale, the periodicity and the intermittent behavior of the process. The water demand data for the two uses (domestic, agricultural) were obtained by the recent work by Papoulakos *et al.* (2017), while the synthetic energy production from the PV panels and the wind farm were derived from the work of Chalakatevaki *et al.* (2017), based on synthetic solar radiation and the wind speed data, estimated by Moschos *et al.* (2017) and Koudouris *et al.* (2017) respectively. These synthetic time series had been produced by employing the methodology by Dimitriadis and Koutsoyiannis (2015a), which is suitable for processes exhibiting double periodicity, and retains the dependence structure through the empirical climacogram of each process (Dimitriadis and Koutsoyiannis, 2015b). Finally, the electric energy demand data were estimated by Mavroyeorgos *et al.* (2017).

4. Outline of Methodology

The representation of the aforementioned system was employed through two approaches, herein referred to as analytical and surrogate.

The analytical approach involved two detailed simulation models, i.e. a conceptual rainfall-runoff scheme, driven by daily rainfall and mean daily temperature data, for estimating the inflows to the reservoir and the actual evapotranspiration, and a reservoir operation model, running at hourly time step, using as inputs the inflows, and the water and energy demands, for estimating the water and energy balance components (outflows, storages, energy surplus and deficits). A detailed description of the computational procedure is given in next section.

The second approach involved the use of ANNs as surrogates of the two analytical models. We have tested several structures, by changing the inputs, the number of neurons and the outputs. An example is given in Figure 1. The general structure of all neural networks was that of the multi – layer perceptrons with one hidden layer and was determined after several trial and error procedures. Different settings with respect to input and output data were also tested.

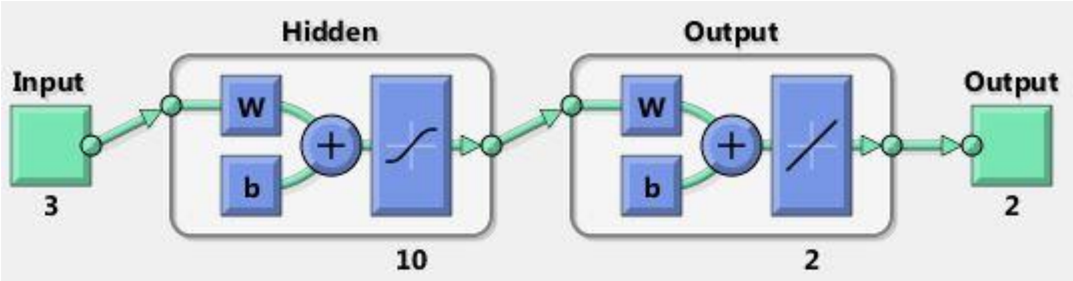

Figure 1: Example of ANN structure, comprising a sigmoid hidden layer and a linear output layer of two neurons.

Key assumption of the methodology was the use of historical data of 7 years length for the construction and training of the artificial neural networks, and synthetic data of 100 years length for their structural validation. The two input data sets (historical and synthetic) were introduced separately to the analytical simulation models, to provide the outputs of interest. Next, the historical input data were used as inputs for the artificial neural networks, while the corresponding output data of the simulation models were the target data according to which the ANNs were trained. Finally, the synthetic data were used to validate the efficiency of the constructed ANNs according to the NSE efficiency coefficient. After the validation procedure, the most suitable neural network structures were further evaluated, in terms of computational performance.

5. Analytical Simulation Models

Two detailed simulation models were used for the current analysis, a conceptual rainfall – runoff model and the model that represents the water and energy fluxes across the hybrid system. The first one uses as inputs daily rainfall, mean daily temperature and daily extraterrestrial radiation data for the estimation of actual evapotranspiration, runoff, and underground losses through a soil moisture accounting scheme. Temperature data were used also for water demand estimation. A more detailed description of this model can be found at Papoulakos *et al.* (2017). The outputs of the rainfall – runoff model were then disaggregated from the daily time step to hourly and served as inputs for the second model.

In Figure 2 we show a schematic diagram of the component and associated fluxes of the reservoir operation model. Initially, it is necessary to make certain assumptions concerning the hierarchy of the different uses as well as for the way in which they are expressed in the simulation routine. Therefore, water for domestic use is ranked first; energy production is second and for irrigation is last amongst priorities. Inputs of the model are the net inflows in the reservoir, the water demand for domestic and agricultural use and the energy balance as it comes for the difference between the energy demand and the energy produced from the renewable resources (solar and wind energy).

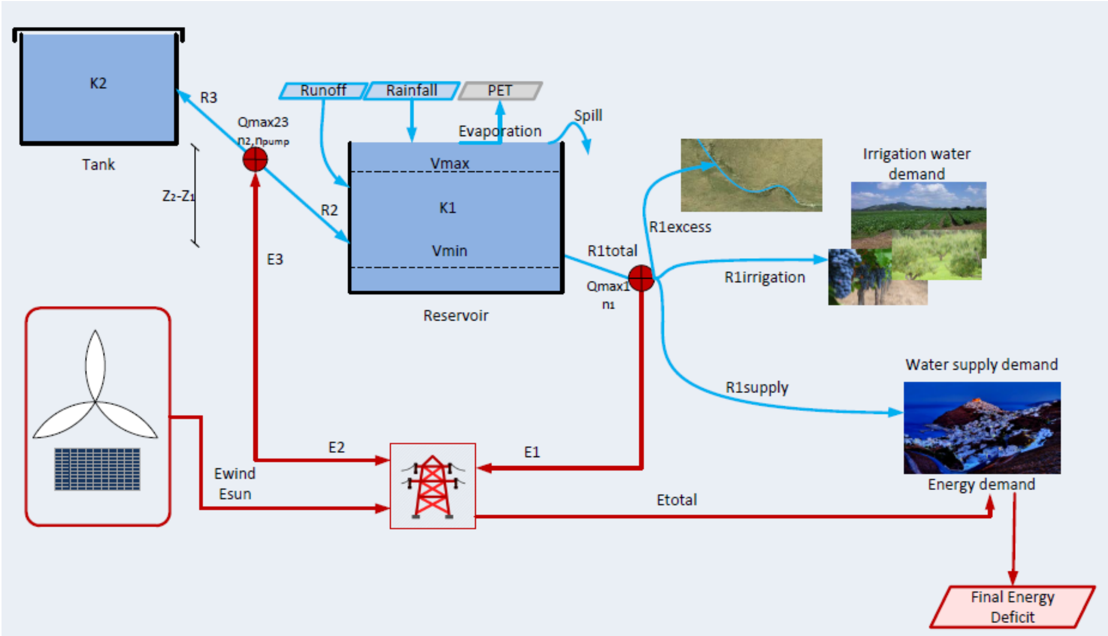

Figure 2: Outline of analytical model for reservoir operation.

The simulation model calculates the water fluxes (outflows and water exchange between the reservoir and the upstream tank) accounting for the current demand for water supply and irrigation, and the need for energy production (in case of energy deficit) or energy storage (in case of energy surplus from the other renewables). Given properties of the system are the minimum and maximum storage of the reservoir, the height and the area of the tank, the altitude difference between the tank and the reservoir, the coefficients of hydraulic losses in each penstock, the performance coefficients of hydroelectric energy production and pumping as well as the flow capacity of the penstocks - tubes.

The initial conditions of the water reserves in the tank and the reservoir are assumed as the half of their storage capacity. At first, the net inflows in the reservoir are calculated as the sum of the runoff, which is derived from the rainfall – runoff model, the actual evaporation from its surface and the rainfall on it (I= Rain + Q - ET) and they are aggregated to the existing reserve. Then, the release off the reservoir to satisfy the domestic use is estimated as the minimum value among the domestic demand value, the available reserve and the responding flow capacity. If the domestic water needs are not satisfied, water from the tank reserve is used in order to fulfill the deficit. The release off the tank is the minimum among the water deficit, the flow capacity of the tube connecting the tank with the reservoir and the tank reserve. At this point, it is important to remark that the flow capacity of the water abstraction tube is not included in the previous restrictions, as it is considered to have been adequately designed for covering the domestic and agricultural water needs and since there is water deficit, it is impossible for this limit to have been reached. Next, water abstraction for irrigational purposes is estimated. This volume is the minimum among the water demand for irrigation, the reserve and the remaining flow capacity. In this case, there is not possibility of covering any deficit from the tank's reserve, as irrigational use is of last priority for the system. Therefore, the dynamic energy potential is preserved in as high level as possible in order to be available in peak demand periods.

When abstractions are completed, we check the current reserve. In case the storage capacity is exceeded, water is pumped upwards to the tank. The pumped volume is the minimum among the excess, the flow capacity and the available storage capacity of the tank. For clarification, if there is an excess of water, there is no need for using the tank's reserve for domestic use and therefore with the expression above, the one – way flow in the tube connecting the tank and the reservoir is ensured, within the same time step. Then, in case that there is remaining excess volume, an amount of it, that is equal to the remaining flow capacity of the abstraction tube, is conducted out of the reservoir in order take full advantage of its energy potential. Further remaining water excess, if any, is driven out through the spillway.

For the simulation procedure, it is assumed that after the completion of the above water fluxes, the energy produced or consumed because of them update the initial energy balance. Further water fluxes are estimated, driven by the current energy balance conditions, deficits or excesses.

In case of energy deficit, there is production of energy by employing the reversible turbine. The volume that outflows off the tank is estimated as the minimum among the equivalent water demand for the production of the energy deficit, the flow capacity, the reserve in the tank and the available storage capacity of the reservoir. Due to this restriction, one – way flow is ensured within the same time step, as if an outflow upwards to the tank had taken place previously in order to avoid spilling, the current available storage capacity of the reservoir would be zero.

In case of energy excess, the water amount needed for pumping is estimated. Practically, the excess of energy is stored as potential energy because of the new position of the water amount. The volume that is finally pumped is the minimum among the equivalent volume estimated, the flow capacity, the available water in the reservoir and the available storage capacity of the tank.

Finally, the energy consumed or produced is estimated and the final revision of the energy balance takes place. Outputs of the model are the values of the energy excess and deficit, the water fluxes and storage in the tank and the reservoir at the end of the time step. The storage values at the end of each previous time step are initial conditions at the beginning of their next one.

6. Results

For the neural network surrogate of the rainfall – runoff model, the trials executed can be classified in four different groups. The **first** group contains the ANNs of the first attempt and served only for determining the structure, the data division (amount of training, testing and validation data) as well as the training algorithm. The **second** one includes neural networks in which the hidden layer size was increasing by one neuron each time for 100 repetitions. The inputs of these networks are the same as these of the rainfall – runoff model and result to a multiple output, a vector containing the estimated values of runoff, actual evapotranspiration and soil moisture. The **third** group is made up from ANNs that have the same inputs as these of the second one, but they are of one single output, estimating separately the runoff, the actual evapotranspiration and the soil moisture. For each one of the three variables 50 different repetitions were made, increasing the hidden layer size by one neuron. In the **fourth** group, in each repetition the hidden layer size is kept the same (15 neurons) but the rainfall data are introduced multiple times; in every repetition the rainfall lag is increased by one phase each time. The trials were made for a multiple output, as in the second group, and for the single output of the runoff as well. The best NSEs of the validation procedure are summarized in Table 1 according to the variable they refer to.

Table 1: NSEs during validation for the rainfall – runoff surrogate.

	Q	ET	S
NSE	0.56	0.64	0.60
Found in group	2	3	4a

As it can be seen from the table above, the NSE for all the three examined variables is quite good for the evapotranspiration and the soil moisture. However, for the runoff, that is going to be further used as input of the hybrid system simulation model, the efficiency coefficient value is not quite satisfactory for this purpose. Therefore, although the neural networks could definitely be a surrogate to substitute the rainfall – runoff model, in this work, they were not finally employed as such models and the original rainfall – runoff model was used to estimate the hybrid system simulation model inputs.

With regard to the neural network surrogate of the water – energy hybrid system simulation model, the trials performed can be classified in three different groups. These of the **first** group were trained with target data the water fluxes in the penstocks, but as the procedure turned out to be highly time demanding and resulted in poor estimations (negative NSE values), the specific approach was set aside and the work focused on simulating the final energy balance time series.

The neural networks of the **second** group simulate the final energy balance either overall or by simulating the deficit and the excess of energy separately. In the first case, the training was performed for 93 repetitions by increasing the hidden layer size by one neuron in each one. In the second one, the number of the hidden neurons was kept the same in every repetition and equal to 10, and the training was repeated for 150 times for the deficits and 143 for the surpluses. None of the trained neural networks in this group estimated well enough the values of the final energy demand. Even though that for the training procedure the best NSE values were quite good, 0.71 in the first case and 0.77 for the deficits in the second case, the highest corresponding values at the validation stage was 0.39 and 0.45 respectively. What is more, the surpluses simulated in the second case do not give good results, as in training the highest NSE value is 0.33 and in validation equal to 0.32. All of the NSE values are lower than the acceptable threshold of 0.50.

Finally, the **third** group of this category's neural networks estimate the final energy balance overall, but they need as extra input variables the stored water in the tank and the reservoir. In this case, two different structures were tried, that of the MLP (50 different repetitions of the training procedure) and that of the NARX (non-linear autoregressive with external (exogenous) input) neural network (one repetition). In this group, the values of the efficiency coefficient are generally higher than these of the previous ones. There are some negative values of the NSE, as in the previous groups, but most of them are higher than 0.90. For the MLP networks, the highest value is equal to 0.96 in training and 0.91 in validation. The final energy balance as estimated by the original simulation model and the MLP surrogate for the validation data set is presented in Figure 3.

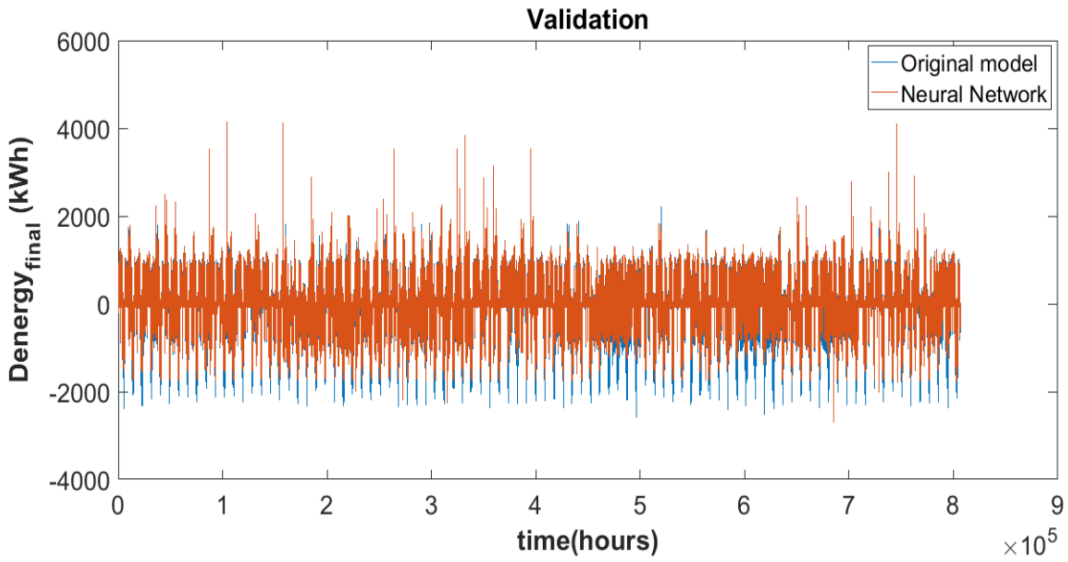

Figure 3: Final energy balance of MLP neural network.

For the NARX neural network the highest NSEs are 0.92 for training and 0.84 for validation. The results of the NARX neural network for the validation data set against the estimation from the original model are presented in Figure 4.

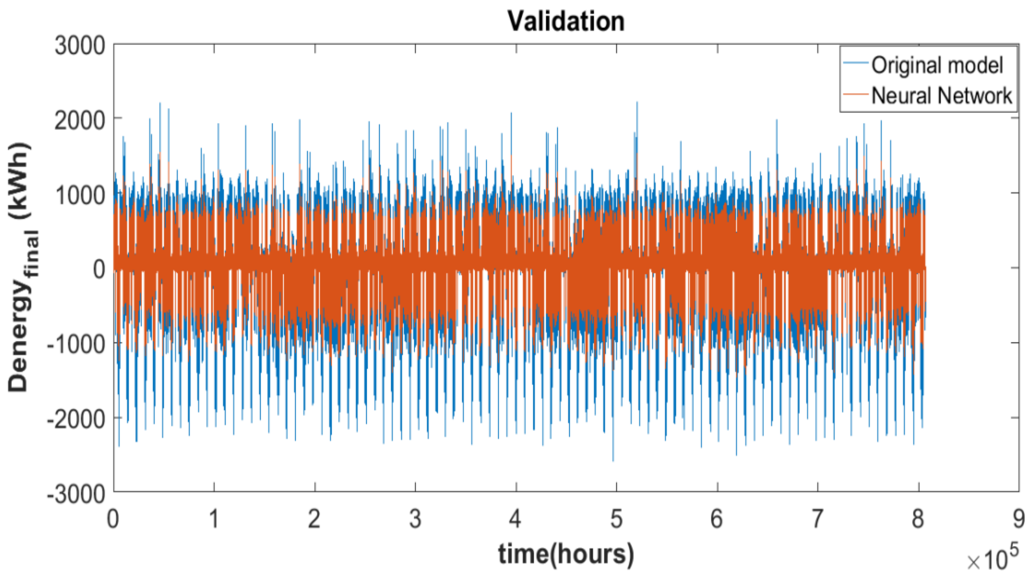

Figure 4: Final energy balance of NARX neural network.

In both cases, these values are highly encouraging. If NARX networks are used, a preprocessing of the data is needed. This procedure could be quite time demanding and therefore, in combination with the fact that in validation the NSE value was lower than the corresponding of the MLP network, does not make the NARX the best choice for the purposes of the current work. As a result, the neural network that was finally selected to surrogate the response surface of the final energy balance was the MLP neural network that presented NSE equal to 0.91 during validation. This network, along with the original rain – runoff model and the simpler models for estimating the irrigation demands and the potential evapotranspiration, consist the final modeling scheme that was used in this work.

The selected neural network and the original simulation model were performed one more time in order to estimate the efficiency in terms of time saving amount. The results showed that the original model needed 14.4 s, while the neural network surrogate model needed only 0.3 s, which means 97.9% of the computational time was saved.

7. Conclusions

First of all, it is important to remark that the neural networks were constructed in the basis of having the same input data as the analytical models. So the periodic function of the solar radiation was taken into account. Its use, along with the use of rain lag, may be one of the reasons for the low NSE values of the corresponding neural networks for runoff estimation. As reported in the literature, the repetition of information may result in over – trained neural networks. However, the neural networks for evapotranspiration and soil moisture present higher NSE values than these for runoff metamodeling, even though they employ the same input variables. Runoff seems to be benefited more, when estimated by the multiple output neural networks, as its best NSE value is found in this group. Although the efficiency coefficient values of the neural networks estimating the rainfall – runoff model outputs are generally acceptable, for the present application are not considered sufficient enough, as they are intended to be used as inputs to the next model, which is of certain efficiency as well.

The water – energy hybrid system simulation model is more demanding, due to the complexity of the system. The regulation of the energy balance that takes place through the system is reflected to the total reserve of water and its distribution to the different storage facilities. Therefore, it is more difficult to have satisfactory enough estimations by using the same input data as for the original model. The use of water storage time series had an important contribution to the increase of the efficiency of the neural networks for the final energy balance estimation. Even when the deficits and surpluses of energy were separately simulated, it was not possible to have so satisfactory results as these of neural networks employing the current storage information. As the system simulated is more complicated, additional information is needed.

For the original model, the simulation had duration of 14s. This run may not seem to be so time demanding to turn to surrogate models. However, this value could be much higher in case of a more extended system with a great number of hydraulic works and power plants or in case of using even longer time series for achieving better accuracy in the system's efficiency estimation. What is more, this time duration would be even more critical when multiple simulation repetitions are needed (i.e. for optimization, sensitivity analysis, uncertainty estimation). Therefore, through validation on time basis, it is clear that the use of artificial neural networks significantly contributes to computational time reduction. As the time saving results seem to be promising through this procedure, a new target is set; that of exploring the possibilities of neural networks as surrogate models for integrated management of water – energy hybrid systems for optimization purposes.