IFAC

Meta-model of an irrigation district distributed-parameter model

Stefano Galelli [*] Francesca Pianosi [**] Rodolfo Soncini-Sessa [***]

[*] *Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Milan, Italy (Tel: +39-2-2399.9630; e-mail: galelli@elet.polimi.it).*
[**] *Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Milan, Italy (Tel: +39-2-2399.9630; e-mail: pianosi@elet.polimi.it).*
[***] *Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Milan, Italy (Tel: +39-2-2399.3551; e-mail: soncini@elet.polimi.it).*

Abstract: The release policy for a reservoir's network can be designed by solving a multi-objective control problem. However, the use of resolution algorithms based on stochastic dynamic programming, with their high computational requests, imposes strong simplification in modelling the water system. In particular, downstream irrigation districts are often described as static sub-systems, i.e. assuming an a priori trajectory of the water demand, while the use of the available simulation models is precluded due to their complexity. To overcome this problem, this study proposes the identification of a meta-model, i.e. a model of a simulation model, for describing an irrigation district. The meta-model computes the overall water deficit in the district, based on meteorological inputs and the water supply from an irrigation canal. Even if the meta-model is very simple, it can be given a physical interpretation in relation to the simulation model behavior. Results obtained with the original model and the meta-model over a real world case study, an irrigation district in the Padana plain, Italy, are presented.

1. INTRODUCTION

Water resources management is a quite intricate problem, because of the simultaneous presence of multiple interests, often conflicting; the existence of physical constraints, due to the scarcity of the resource; and the high randomness that affect the system. As for the management of multipurpose reservoir's network, an adequate approach consists in using, as a release policy, one of the Pareto-efficient policies that can be designed through the resolution of a multi-objective control problem (see, for example, Soncini-Sessa et al. (2007a)). A multi-objective control problem can be traced back to a family of single-objective optimal control problems, each of which can be solved through algorithms based on the numerical resolution of the Bellman equation (see for example Yeh (1985) for a review of first applications of stochastic dynamic programming to water resources management and Soncini-Sessa et al. (2007a) for recent improvements). The main limit of this approach is that the computing time increases exponentially with the state dimension and consequently only few state variables can be used to model the water system.

Among the water system components, irrigation districts constitute a typical example of sub-systems whose description is strongly simplified in force of the above considerations. In fact, on the one hand the literature shows a vast variety of distributed parameter conceptual models that can compute the water demand of an irrigation district with good precision; but they use a huge number (hundreds and often thousands) of state variables and thus are incompatible with stochastic dynamic programming. On the other hand, reproducing the district's dynamics with few state variables is not easy: the identification of em-

pirical input-output models or grey-box models is usually precluded because of the lack of data. As a consequence, it is common practice to abandon the dynamic description of irrigation districts, and assume an a priori (usually periodic) trajectory for the water demand (see for example Soncini-Sessa et al. (2007b)).

Our proposal to approach this problem is to resort to a meta-model of the irrigation district. In the literature, the term *meta-model* usually denotes a simplified model that has been identified on data produced by a complex, computationally expensive model, and that can be used to run less time consuming simulations. In the last decades, meta-models have found successful applications for the resolution of optimization problems, where they are coupled with algorithms based on iterative search, which require performing many simulations. Recently, applications of this approach have appeared also in the context of water systems management, e.g., for the optimization of a water distribution system (Broad et al. (2005)) and for optimizing the management policy of a reservoir network with fixed-class control laws (Neelakantan and Pundarukanthan (2000)). Other authors (e.g. Young and Ratto (2007)) use the term *Dominant Mode* model to refer to a reduced order model that mimics a complex simulation model for a particular set of the simulation model parameters; while reserves the term *meta-model* for a reduced order model that mimics the complex simulation model for *all* its possible parameter values, thus allowing for Sensitivity Analysis.

In the present study, the meta-modelling approach is applied to the simulation model of an irrigation district. The parameters of the simulation model are given for fixed, and the scope of the meta-modelling exercise is neither to

perform sensitivity analysis or to reduce the computing time in simulation but rather to reduce the dimension of the state vector. This is in fact the very requirement for the meta-model to be usable to solve an optimal control problem with stochastic dynamic programming.

The distributed parameter model, simply called simulation model in the following, subdivides the irrigation district in 10^4 cells and for each one computes 2 state variables; thus it has $2 \cdot 10^4$ state variables. On the other hand, the meta-model reproduces the overall daily deficit in the irrigation district with only 2 states variables; it has a nonlinear, input-output form and belongs to the class of Data Based Mechanistic (DBM) models (Young (1998)). An application is presented on a large irrigation district located in the Padana Plain, Italy.

The paper is organized as follows. In the next section, the principal methodological issues of the meta-modelling technique will be discussed. Section 3 presents a brief description of the irrigation district under study and of the simulation model. Section 4 is devoted to the meta-model, describing its inputs and output, the data used for its identification, its structure and discussing the results of estimation and validation. Further discussion and concluding remarks are presented in Section 5.

2. META-MODELS

A meta-model (or surrogate or approximation model) is a faster representation of a computationally expensive model, usually denoted as simulation model (or original model). A meta-model is identified based on data produced by simulation with the original model. The choice of the type of meta-model and its identification procedure depend on the goal of the meta-modelling exercise. Kleijnen and Sargent (2000) identifies four goals: understanding of the original model, prediction of future outputs, validation of the original model and optimization, which is the most common goal of meta-models. As for the type of meta-models, the literature shows a vast variety of applications, ranging from linear regression to splines or Neural Network (for a survey, see for example Simpson et al. (2001)).
Kleijnen and Sargent (2000) present a complete methodology for fitting and validating meta-models. For the meta-model discussed in this paper, the following procedure has been used.

(1) The original model is critically analyzed and its principal characteristics studied. This first step is of crucial importance since it constitutes the basis of the whole procedure, creating the knowledge base that will permeate the whole meta-model identification procedure.
(2) The output that best represents the goal of the meta-modelling exercise and its relevant inputs are chosen.
(3) The input-output data that will be used for the meta-model identification are produced via simulation of the original model. According to the practice in meta-models literature, this simulation experiment will be denoted in the following as Experimental Design (ED). This step is also a crucial one, because the data-set produced in the ED directly affects the meta-model range of validity.
(4) The class of models (DBM, Neural Networks, etc.) to which the meta-model will belong is chosen.

(5) The meta-model is fitted on the estimation data-set and it is validated. If this step provides satisfactory results, the procedure ends, otherwise it is necessary to go back to the previous step and select another type of meta-model.

3. SIMULATION MODEL

The simulation model presented in this paper refers to the Muzza-Bassa Lodigiana district, located in the Padana Plain, Italy, south-east from the city of Milan. The irrigation district has an area of about 700 Km^2 and it can be considered representative of agricultural and irrigation practice in a wide portion of the Padana Plain. The cultivated area covers the 85% of the whole district; major crops are cereals (especially maize) and permanent grass. The source of irrigation supply is the Muzza canal, which originates from the Adda River, an effluent from the Lake of Como. The irrigation supply is thus controlled by the release from the lake, which is regulated through the Olginate dam.

The simulation model is a Soil-Vegetation-Atmosphere-Transfer model of the flow in the upper part of the vadose zone, based on previous works of Facchi et al. (2005). It considers the space variability of soils, crops, meteorological and irrigation inputs by subdividing its spatial domain with a regular mesh, thus being characterized by a total of 10^4 cells. Each cell has a side length of 250 m and it is divided into two layers: the first one (evaporative) represents the upper part of the soil profile and it is characterized by a constant depth of 15 cm, while the second one (transpirative), representing the root zone, has a time-varying depth. The two layers are modeled as two nonlinear reservoirs in cascade.
The hydrological balance in the upper layer of the i-th cell is given by

$$\theta_{1,t+1}^{(i)} = \theta_{1,t}^{(i)} + R_{t+1}^{(i)} - I_{t+1}^{(i)} + Q_{i,t+1}^{(i)} + $$
$$- Q_{r,t+1}^{(i)} - E_{t+1}^{(i)} - Q_{u,t+1}^{(i)} \quad \text{(1a)}$$

where $\theta_{1,t}^{(i)}$ is the water content at (discrete) time t, $R_{t+1}^{(i)}$ is the rainfall in the time interval $[t, t+1)$, $I_{t+1}^{(i)}$ the canopy interception, $Q_{i,t+1}^{(i)}$ the irrigation supply, $Q_{r,t+1}^{(i)}$ the surface runoff, $E_{t+1}^{(i)}$ the evaporation and $Q_{u,t+1}^{(i)}$ the percolation to the second layer (upper percolation).
The hydrological balance in the lower layer, instead, is given by

$$\theta_{2,t+1}^{(i)} = \theta_{2,t}^{(i)} + Q_{u,t+1}^{(i)} - T_{t+1}^{(i)} - Q_{d,t+1}^{(i)} \quad \text{(1b)}$$

where $\theta_{2,t}^{(i)}$ is the water content at time t, $Q_{u,t+1}^{(i)}$ is the percolation from the upper layer in $[t, t+1)$, $T_{t+1}^{(i)}$ the transpiration and $Q_{d,t+1}^{(i)}$ the deep percolation. The water contents $\theta_{1,t}^{(i)}$ and $\theta_{2,t}^{(i)}$ are expressed in millimeter (assuming constant value over the cell) and all other flux variables are reported to the same unit of measure. The canopy interception $I_{t+1}^{(i)}$ is computed as a function of the leaf area index, the cover fraction and the volume capacity per unit foliage area. The computation of the surface runoff $Q_{r,t+1}^{(i)}$ is based on the Curve Number method (see Mockus (1972)), while the upper and deep percolation ($Q_{u,t+1}^{(i)}$

and $Q^{(i)}_{d,t+1}$ respectively) are modeled as a Darcian-type gravity flow in the saturated soil (see Brooks and Corey (1964)). The evaporation $E^{(i)}_{t+1}$ and the transpiration $T^{(i)}_{t+1}$ are derived based on the reference crop evapotranspiration $ET0_{t+1}$, which depends on meteorological data (air temperature, solar radiation, air moisture and wind speed). More precisely, evaporation is obtained by multiplying $ET0_{t+1}$ by the evaporative coefficient $K^{(i)}_{e,t}$, i.e.

$$E^{(i)}_{t+1} = K^{(i)}_{e,t}(\theta^{(i)}_{1,t}) \cdot ET0_{t+1} \qquad (2a)$$

where $K^{(i)}_{e,t}$ is a function of the water content $\theta^{(i)}_{1,t}$. Transpiration is obtained by multiplying $ET0_{t+1}$ by the basal coefficient $K^{(i)}_{cb}$ and by the water stress coefficient $K^{(i)}_{s,t}$

$$T^{(i)}_{t+1} = K^{(i)}_{s,t}(\theta^{(i)}_{2,t}) \cdot K^{(i)}_{cb} \cdot ET0_{t+1} \qquad (2b)$$

The product $K^{(i)}_{cb} \cdot ET0_{t+1}$, named potential transpiration and denoted with $T^{(i)}_{c,t+1}$, represents the crop transpiration in conditions of optimal water content. From the above relation, it follows that $T^{(i)}_{t+1} = K^{(i)}_{s,t}(\theta^{(i)}_{2,t}) \cdot T^{(i)}_{c,t+1}$, i.e. actual and potential transpiration are related by the coefficient $K^{(i)}_{s,t}$, which expresses the effect of limited water availability in the soil and depends on the water content $\theta^{(i)}_{2,t}$.

The irrigation supply $Q^{(i)}_{i,t+1}$ is either zero or equal to 150 mm. This value represents the amount of water necessary to completely flood the cell surface, thus reproducing the effect of the flooding irrigation method, which is basically the only irrigation method adopted in the district. The i-th cell receives this amount of water if two conditions are satisfied: first, the water content $\theta^{(i)}_{2,t}$ in the root zone is lower than a threshold value $\theta^{(i)}_r$ (mm), which correspond to a condition where crop growth is at risk; and second, the cell has the right to be irrigated according to the irrigation scheduling, which is computed by considering the currently available water and past supply. Note that if only the second condition is verified, the irrigation supply is refused, i.e. $Q^{(i)}_{i,t+1}$ is set to zero and the corresponding amount of water is withdrawn to the Adda River.

4. META-MODEL IDENTIFICATION

The simulation model, using two state variables for each cell, presents a total number of state variables of the order of $2 \cdot 10^4$, which is far from being tractable in an optimization algorithm based on stochastic dynamic programming, even when adopting advanced methods (e.g., neuro-dynamic programming). In the present section we will discuss the possibility of resorting to an approximate model of the irrigation district, i.e. a meta-model.

4.1 Input-Output selection

In order to synthetically assess the condition of the i-th cell of the irrigation district, the following deficit variable $d^{(i)}_t$ has been introduced

$$d^{(i)}_t = \begin{cases} \theta^{(i)}_{fc,2} - \theta^{(i)}_{2,t} & \text{if } \theta^{(i)}_{2,t} < \theta^{(i)}_r \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

where $\theta^{(i)}_{fc,2}$ (mm) is the water content in the lower layer corresponding to the field capacity. The deficit $d^{(i)}_t$ thus represents the amount of water necessary to re-establish the optimal water content in the i-th cell, when the latter is in stress condition. The cell deficits $d^{(i)}_t$ are then aggregated over a spatial domain that includes all the cells cultivated with maize and permanent grass (the two most common and economically significant crops of the district) to form the total deficit d_t. For the sake of clarity, d_t is expressed in terms of volume (m^3), i.e. $d_t = 250^2 \cdot (\Sigma^i d^{(i)}_t \cdot 10^{-3})$. This scalar variable represents the overall quantity of water lacking in the cells in stress condition, and it can be assumed to be representative for the water demand for irrigation; as such, it is chosen as the output of the meta-model.

The definition of the total deficit d_t deserves some further comment. It must be noted, in fact, that the presence of the threshold $\theta^{(i)}_r$, which distinguishes the stress condition from the no-stress condition, introduces a random effect into the dynamics of the total deficit: if at time t the number of stressed cells is lower than the total number of cells, the future value of d_{t+1} will depend both on the evolution of the stressed cells and on the number of cells that will cross the threshold in the time interval $[t, t+1)$, the last being unpredictable at time t when knowing d_t only. It follows that the accuracy of the prediction of the total deficit increases with the number of stressed cells. We shall return to this issue in Section 4.4.

The selection of the inputs is even more keen and requires some more considerations on the structure of the simulation model. For this, we will first consider the dynamics of one cell (say the i-th) and develop the reasoning as if the meta-model were to approximate one cell. Before proceeding, it must be noted that it is not possible to use inputs that depend on the two state variables $\theta^{(i)}_{1,t}$ and $\theta^{(i)}_{2,t}$, since this would require to introduce other dynamical meta-models to describe the evolution of these variables. Now, back to the definition of the cell deficit $d^{(i)}_t$, we see from (3) that it is a function of the water content $\theta^{(i)}_{2,t}$, whose dynamics, according to (1b), is driven by the transpiration $T^{(i)}_{t+1}$, the percolation $Q^{(i)}_{u,t+1}$ from the upper layer and the deep percolation $Q^{(i)}_{d,t+1}$. The transpiration $T^{(i)}_{t+1}$ can not be used, because, according to (2b), is related to the stress coefficient $K^{(i)}_{s,t}$, which is a function of the water content $\theta^{(i)}_{2,t}$. However the transpiration $T^{(i)}_{t+1}$ can be approximated with the potential transpiration $T^{(i)}_{c,t+1}$, which is therefore selected as input. The percolation $Q^{(i)}_{u,t+1}$ from the upper layer depends on the net rainfall $(R-I)^{(i)}_{t+1}$, the evaporation $E^{(i)}_{t+1}$, the surface runoff $Q^{(i)}_{r,t+1}$ and the irrigation supply $Q^{(i)}_{i,t+1}$. The variables $E^{(i)}_{t+1}$ and $Q^{(i)}_{r,t+1}$ can be neglected because their contribution is scarce, as well as $(R-I)^{(i)}_{t+1}$, because $T^{(i)}_{c,t+1}$, being a function of the meteorological variables, holds the information about the rain. Therefore, the percolation $Q^{(i)}_{u,t+1}$ can be approximated by the irrigation supply $Q^{(i)}_{i,t+1}$, which is selected

as a second input. Finally, the deep percolation $Q_{d,t+1}^{(i)}$ is ignored, because it does not bring more information.

In conclusion, assuming that the deficit of the i-th cell can be described by considering $T_{c,t+1}^{(i)}$ and $Q_{i,t+1}^{(i)}$ as inputs, the dynamics of the total deficit will be driven by the total potential transpiration $T_{c,t+1}$ (m^3) and the district irrigation supply $Q_{i,t+1}$ (m^3), obtained by aggregating the cell potential transpiration and the cell irrigation supply as done for the output variable of the meta-model.

As discussed in the previous section, the water supply to the irrigation district is regulated by a sheduling system, which foresees the possibility for a given cell of refusing allocated water. Therefore the total irrigation supply $Q_{i,t+1}$ can be lower than the total water volume diverted from the Adda River. As a consequence, it is necessary to identify another meta-model for estimating $Q_{i,t+1}$. To this end it is necessary to know the available water, i.e. the volume Q_{t+1} diverted from the river, and the total water demand W_{t+1}, which influences the probability of supply refusals and can be computed as

$$W_{t+1} = T_{c,t+1} + d_t \qquad (4)$$

We will name the first meta-model *deficit-meta-model*, and the second *supply-meta-model*.

4.2 Experimental Design

In order to run the Experimental Design, it is necessary to specify a trajectory for the inputs of the simulation model: the meteorological data and the volume diverted from the Adda River into the Muzza canal. As for the former, time series are available of daily records of air temperature, solar radiation, precipitation, air moisture and wind speed at the Sant'Angelo Lodigiano meteo station, which is located in the western side of the irrigation district; as for the latter, the time series of daily volume at the Muzza canal diversion dam were used. The simulation was run over a period of 7 years, from 1993 to 1999, thus producing a data-set of input and output data. The first 5 years are used for the meta-model calibration, while the last 2 years are used for the validation.

4.3 Choice of the type of meta-model

Two major difficulties must be faced when identifying the deficit-meta-model: the nonlinearity of the output dynamics and the randomness induced by definition (3) of the cell-deficit, from which the output is derived. These difficulties suggests to resort to a Data Based Mechanistic (DBM) model, a class of nonlinear input-output model introduced in the literature by Young (for a survey see Young (1998) and references therein). With this choice, the model takes the form of a transfer function model with State Dependent Parameters (SDP) that, in our application, must account for both the system's nonlinearity and the random behavior of the output. Moreover, in accordance with the DBM modelling philosophy, the meta-model will be given a physically meaningful interpretation, thus maintaining a relation with the physical-based simulation model that would be precluded if other classes of models were chosen, e.g. Neural Networks or NARMAX models.

For the supply-meta-model an algebraic relation will turn out to be sufficient.

4.4 Meta-model calibration and validation

The DBM meta-model that describes the dynamics of the total deficit has the form of a Transfer Function (TF) model with State Dependent Parameters (SDP)

$$d_{t+1} = \frac{\beta_1(d_t)}{1-\alpha(d_t)q^{-1}}T_{c,t+1} + \frac{\beta_2(d_t)}{1-\alpha(d_t)q^{-1}}Q_{i,t+1} + e_{t+1}$$
$$(5)$$

where q^{-1} is the backshift operator (i.e. $q^{-1}d_{t+1} = d_t$), $\alpha(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are state dependent parameters, i.e. coefficients that, instead of being constant as in conventional linear TF model, vary with the value of the 'state' variable d_t. The last term e_{t+1} accounts for process error. The shapes of $\alpha(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ are derived directly from the data by means of a suitable parameter estimation procedure (Young et al. (2002)). The nonparametric estimates of the SDPs so obtained are shown in Fig. 1. Let us first analyze the potential transpiration parameter $\beta_1(\cdot)$ (Fig. 1b): it is positive (the transpiration indeed contributes to increase the deficit) and increasing with the deficit d_t. This last characteristic can be interpreted as follows: when the deficit is low, the number of cells with $\theta_{2,t} > \theta_r$ (not-stressed cells) is likely to be high and thus the future deficit value d_{t+1} will be strongly influenced by the number of cells that will cross the threshold in $[t, t+1)$, but this number is unpredictable, at time t, knowing d_t only. Therefore the deficit dynamics exhibits a strong uncertainty and the relation between input $T_{c,t+1}$ and output d_{t+1} is weakened: this implies a low value for $\beta_1(\cdot)$. On the other hand, as the deficit increases, the number of stressed cells also increases and the deficit dynamics becomes more and more deterministic, thus increasing the value of $\beta_1(\cdot)$. Analogous considerations are valid for the irrigation supply parameter $\beta_2(\cdot)$, shown in Fig.1c; the only difference is that here the parameter sign is negative, which is consistent with the fact that irrigation tends to set down the deficit. Fig.1a finally shows the nonparametric estimate of the state dependent parameter $\alpha(\cdot)$: it appears to be positive (thus guaranteing the positivity of the deficit) and increasing with d_t. More precisely, for low values of d_t it is lower than 1, which implies that the dynamics of d_t is stable. This expresses the fact that low deficits, typical of winter condition, are canceled by rainfall, which is not among the inputs of the meta-model. As the deficit increases, the value of α approaches 1 and exceeds it when d_t equals $6.8 \cdot 10^6$ m^3, thus making the model unstable. This expresses the fact that when more than 30% of the district [1] is in stress condition, it is likely that many cells are just below the threshold θ_r and will reach stress condition in the next step. This is unexplained by the input values and is therefore expressed through instability.

In order to use the model outside of the estimation dataset, the nonparametric estimates of the SDPs reported in Fig. 1 need to be parameterized through suitable functions $f_i(\cdot)$. The following exponential law

$$f_i(d_t) = c_{1,i} \cdot \exp(c_{2,i} \cdot d_t) + c_{3,i} \cdot \exp(c_{4,i} \cdot d_t) \qquad (6)$$

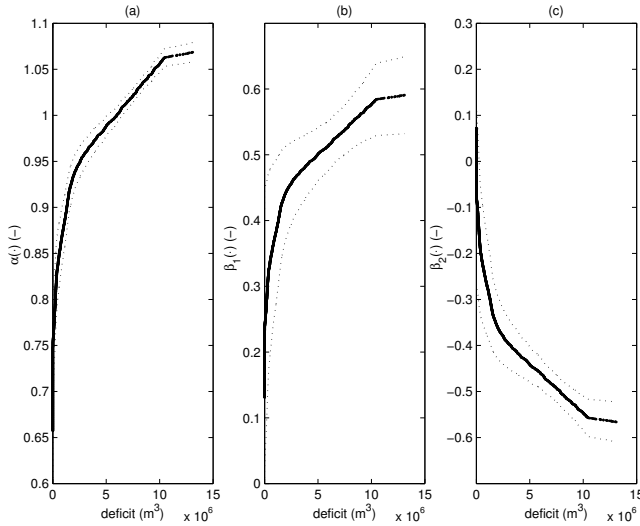[1] When all the cells in the district reach stress condition the value of d_t is about $23 \cdot 10^6$ m^3.

Fig. 1. Nonparametric estimates of the State Dependent Parameters $\alpha(\cdot)$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$. Dotted lines are standard error bounds.

is well suited for this purpose and the parameters $c_{1,i}$, ..., $c_{4,i}$ for $i = 1, 2, 3$ can be estimated by interpolating the nonparametric estimates in Fig.1. Model (5) can then be rewritten as

$$d_{t+1} = ad_t^e + b_1 T_{c,t+1}^e + b_2 Q_{i,t+1}^e + e_{t+1} \qquad (7)$$

where $d_t^e = f_1(d_t) \cdot d_t$ is the effective deficit, and $T_{c,t+1}^e = f_2(d_t) \cdot T_{c,t+1}$ and $Q_{i,t+1}^e = f_3(d_t) \cdot Q_{i,t+1}$ are the so called effective inputs.

Model (7) is linear in the effective inputs, thus its parameters a, b_1 and b_2 can be obtained through a Refined Instrumental Variable method. Finally, the estimates of all the parameters a, b_1, b_2, $c_{1,i}$, ..., $c_{4,i}$ (with $i = 1, 2, 3$) are refined through a global nonlinear optimization procedure. The residual e_{t+1} of model (7) is not white. This is not surprising considering the number of approximation introduced in the meta-modelling exercise, e.g. the aggregation of the spatially distributed deficit into just one value. However the residual can be described as a 1st order autoregressive process: $e_{t+1} = \gamma \cdot e_t + \xi_{t+1}$, with ξ_{t+1} zero-mean white noise.

In Fig. 2, the trajectory of the total deficit predicted by the DBM model is compared with the one produced by the simulation model. A synthetic measure of the meta-model prediction ability is provided by the coefficient of determination, which is equal to 0.99 on both the estimation and validation data-set.

For identifying the supply-meta-model structure, first of all observe that for each cell of the irrigation district (say the i-th), the simulation model estimates the irrigation supply $Q_{i,t+1}^{(i)}$ on the basis of the water demand of the cell (which is in turn a function of the deficit $d_t^{(i)}$ and the transpiration $T_{t+1}^{(i)}$), the available resource (i.e. the volume Q_{t+1} diverted from the Adda River into the Muzza canal times the average efficiency η of the water distribution network) and the irrigation scheduling. To define the supply-meta-model let us imagine for a while that there was no irrigation scheduling: in that case the irrigation supply would be given by
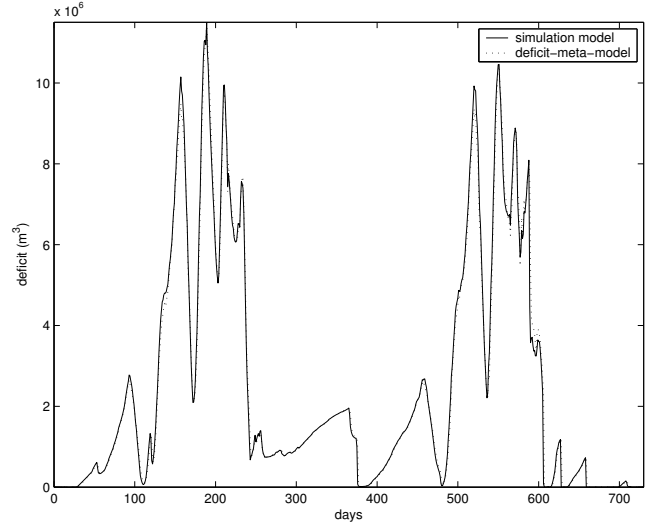

Fig. 2. Trajectories of the total deficit computed by the simulation model (solid line) and predicted with the deficit-meta-model (dashed line) for the two years of the validation data-set.

$$Q_{i,t+1} = \min\{\eta \cdot Q_{t+1}, W_{t+1}\} \qquad (8)$$

where W_{t+1} is the water demand of the irrigation district defined by (4). When scheduling is present, equation (8) does not express the irrigation supply, but only an upper bound of it, since not all the stressed cells can be served in a single round. However we do not know which cells can be served, since this is a spatially distributed information which is not available within a meta-model. Then, we can only try to estimate this lost information by taking into account the past story of the system. To do this, we approximate the fraction of the water demand W_{t+1} that is actually served during the round with the following estimator

$$\tilde{W}_{t+1} = w_{0,1} \cdot T_{c,t+1} + ... + w_{n_1,1} \cdot T_{c,t+1-n_1} +$$
$$+ w_{1,2} \cdot d_t + ... + w_{n_2,2} \cdot d_{t-n_2} \qquad (9)$$

Then, the irrigation supply is given by equation (8), where W_{t+1} is replaced by $\tilde{W}_{t+1}$.

The order $[n_1, n_2]$ of model (9) is chosen by trial-and-error, while the parameters are estimated with a Refined Instrumental Variable method, as in model (7). Remember that, in the perspective of using the meta-model with dynamic programming, it is necessary to keep the model order as low as possible. Unlike the deficit-meta-model, where the model with the lowest possible order was found to have satisfactory performances, here the minimization of the number of regressors implies a reduction of the meta-model performances, as shown in Table 1. The adoption of a $[1, 1]$ model implies a performances reduction of about 10% with respect to the $[2, 2]$ model. Fig. 3 compares the trajectories produced by the $[1, 1]$ model and the simulation one.

Model (9) is very simple and its estimation ability might be judged not completely satisfactory. However, the performance of this model should not be judged per se, but considering the effect on the prediction of the total deficit, which is the meta-model output we are interested in. The combination of the two meta-models produces satisfactory results: the coefficient of determination of model (7), fed

Table 1. Coefficient of determination R^2 over the estimation and validation data-set for different orders of the model (9).

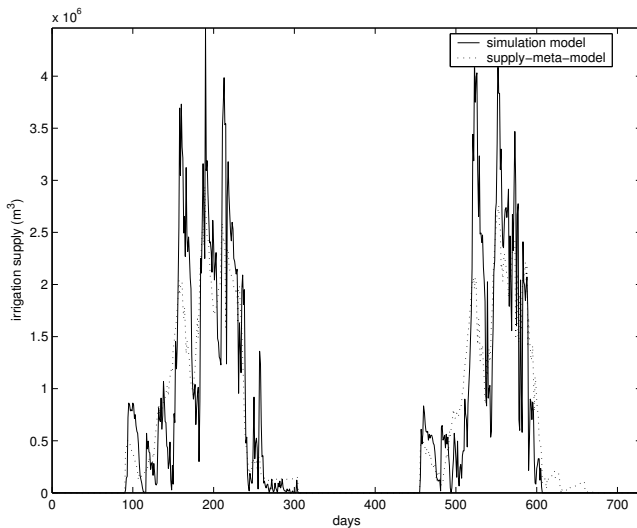| n_{b1}	n_{b2}	R^2 (tar)	R^2 (val)
1	1	0.64	0.67
2	1	0.67	0.72
1	2	0.70	0.72
2	2	0.73	0.78
4	4	0.74	0.80

Fig. 3. Trajectories of the irrigation supply computed by the simulation model (solid line) and estimated by the supply-meta-model (dashed line) for the two years of the validation data-set.

with an irrigation supply $Q_{i,t+1}$ computed with model (9), is equal to 0.98 both on the estimation and validation data-set. This confirms that a simple model is sufficient for our purpose.

5. CONCLUSION

The paper shows the development of a meta-model to approximate the output of a complex, distributed parameter simulation model of an irrigation district in northern Italy. Notwithstanding its simple structure, the meta-model provides satisfactory results in one-step-head prediction, as well as a physical interpretation of the simulation model dynamics. Future research will concentrate on the integration of this meta-model into the global model of the Lario basin water system (which already encompasses the Lake of Como and its catchment) in order to design efficient release policies for the Lake of Como, with control laws depending on both the lake storage and the total deficit in the irrigation district. This will be possible thanks to the low order of the identified meta-model, which guarantees tractability of the optimal control problem with stochastic dynamic programming, while it would have been precluded with the simulation model, which employs thousands of state variables.

REFERENCES

D.R. Broad, G.C. Dandy, and H.R. Maier. Water distribution system optimization using metamodels. *Journal of Water Resources Planning and Managment*, 131(3): 172–180, 2005.

R.H. Brooks and A.T. Corey. Hydraulic properties of porous media. Technical Report Hydrology paper n.3, Colorado State University, Fort Collins, Co., 1964.

A. Facchi, C. Gandolfi, B. Ortuani, and D. Maggi. Simulation supported scenario analysis for water resources planning: a case study in northern italy. *Water science and technology*, 51(3-4):11–18, 2005.

J.P.C. Kleijnen and R.G. Sargent. A methodology for fitting and validating metamodels in simulation. *European journal of the operational research*, 120(1):14–29, 2000.

V. Mockus. Estimation of direct runoff from storm rainfall. National engineering handbook, section 4, hydrology, USDA-SCS (U.S. Department of Agriculture - Soil Conservation Service), Washington, D.C., 1972.

T.R. Neelakantan and N.V. Pundarukanthan. Neural network-based simulation-optimization model for reservoir operation. *Journal of Water Resources Planning and Managment*, 126(2):57–64, 2000.

T.W. Simpson, J.D. Peplinski, P.N. Koch, and J.K. Allen. Metamodels for computer based engineering design: survey and recommendations. *Engineering with computers*, 17:129–150, 2001.

R. Soncini-Sessa, A. Castelletti, and E. Weber. *Integrated and participatory water resources management. Theory.* Elsevier, Amsterdam, NL, 2007a.

R. Soncini-Sessa, F. Cellina, F. Pianosi, and E. Weber. *Integrated and participatory water resources management. Practice.* Elsevier, Amsterdam, NL, 2007b.

W. Yeh. Reservoir management and operations models: a state of the art review. *Water Resources Research*, 21 (12):1797–1818, 1985.

P.C. Young. Data-based mechanistich modelling of environmental, ecological, economic and engineering systems. *Environmental modeling and software*, 13:105–122, 1998.

P.C. Young and M. Ratto. From dominant mode analysis to dynamic meta-modelling. In *5th International Conference on Sensitivity of Model Output (SAMO)*, Budapest, H, June 18-22 2007.

P.C. Young, P. McKenna, and J. Bruun. Identification of nonlinear stochastic systems by state dependent parameter estimation. *International journal of control*, 74: 1837–1857, 2002.