Agricultural Water Management 97 (2010) 187–200

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Building a metamodel of an irrigation district distributed-parameter model

S. Galelli [a,*], C. Gandolfi [b], R. Soncini-Sessa [a], D. Agostani [b]

[a] Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza L. da Vinci, 32, I-20133 Milano, Italy
[b] Dipartimento di Ingegneria Agraria, Università degli Studi di Milano, Via Celoria, 2, I-20133 Milano, Italy

ARTICLE INFO

Article history:
Received 16 April 2009
Accepted 5 September 2009
Available online 17 October 2009

Keywords:
Flood irrigation water management
Distributed-parameter models
Lumped-parameter models
Metamodels
Green water
Blue water

ABSTRACT

Complex decision-making problems, related to planning and management of irrigation water resources, generally preclude the use of large, distributed-parameter models, which are then commonly substituted by lumped-parameter models. This paper, with the aim of improving the quality of these latter, introduces a new approach for their design. This approach is based on metamodelling, which proposes to identify a simple, lumped-parameter model on the basis of the data produced via simulation with a distributed-parameter model. The approach proposed is tested on a real-world case study, namely the identification of a metamodel describing the water demand of the Muzza–Bassa Lodigiana irrigation district (Italy). The metamodel, which inherits the physical description of the original distributed-parameter model, is sufficiently simple to permit the resolution of an optimal control problem, i.e. the design, via stochastic dynamic programming, of the release policy of Lake Como, serving the Muzza irrigation district.

1. Introduction

Fresh water scarcity is presently affecting arid and semi-arid countries (i.e. about 2 billion people) and it is reasonably expected that in the next few decades a larger part of the world's population will experience this problem, because of the combination of climate change, industrial development and population growth (Vörösmarty et al., 2000). The agricultural sector, which accounts for about the 70% of the actual global water withdrawal (see Kijne, 2003), may thus face frequent water shortages, even in the mid latitudes. Efficient planning and management of agricultural water resources is thus an unavoidable task, which requires, in turn, an accurate modelling of the water stored in the root zone of crops and, thus, of the irrigation water demand. The former is the so-called 'green water' (Falkenmark, 1991), as opposed to the 'blue water' of lakes, rivers and groundwater stores.

The artificial process of water distribution and application to agricultural fields is intrinsically distributed in nature, and distributed hydrological modelling approaches, despite their huge data demand, are generally employed for the simulation of the soil water dynamics in the vadose zone. A number of distributed-parameter models have become available for this purpose, either physically based (see, for instance, Ahrends et al., 2008; van Dam et al., 2008; Sudicky et al., 2006; Voss and Provost, 2002; Singh et al., 1999), or conceptual (Niswonger et al., 2006; Arnold and Fohrer, 2005; Neitsch et al., 2002; Bergström, 1995; Refsgaard and Storm, 1995; Young et al., 1989). Despite the large improvements in processor speed, both types of models are still characterized by a severe computational effort. The number of simulations that is possible to produce is in practice limited, which is one of the reasons why decision makers ask for scenario (or what-if) analysis even when the formulation of an optimal decision-making problem would be more appropriate. In fact, when dealing with planning or management problems (e.g. the design of a reservoir release policy serving a cultivated area or the optimal allocation of water among different crops), these models are unusable, since their state dimensions are such that optimization algorithms would require billions of years to determine the optimal solution. This statement may appear to overstate the case, while in fact it is understates it. To verify it, consider the case of a single reservoir that supplies an irrigation district. When the system is modelled by discretizing the reservoir storage with 10 values and representing the district with an a priori given trajectory of the water demand, the design of a release policy requires about 1 min on a good desktop computer. When the irrigation system is described by a distributed-parameter model with 10,000 state variables (each one discretized with 10 values), the computing time for the same design turns out to be $10^{10,000}$ min, since the computing time grows exponentially with the state dimensions (this is the so-called 'curse of dimensionality', see Section 2.1). To feel what $10^{10,000}$ min are, note that from the Big Bang until now only 10^{16} min have elapsed.

In order to make optimal decision-making problems actually tractable, it is therefore necessary to resort to computationally efficient models, such as lumped-parameter models. This kind of

* Corresponding author. Tel.: +39 02 2399 3670; fax: +39 02 2399 9611.
E-mail address: galelli@elet.polimi.it (S. Galelli).

0378-3774/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.agwat.2009.09.007

model can easily be obtained for the dynamics of the storage in a reservoir or of the inflow to a reservoir from the contributing catchment: the first is well represented by a mass balance equation and the second by empirical models, e.g. an Auto Regressive Moving Average eXogenous (ARMAX) model or a Data Based Mechanistic (DBM) model (see Soncini-Sessa et al., 2007), since long time series of measured inflows are commonly available. The water demand of an irrigation district cannot be treated with this approach, because it is intrinsically distributed in space and measurement is still impractical (Schiermeier, 2008); therefore the identification of empirical models is generally precluded for cultivated areas. It is then common practice to employ lumped-parameter, conceptual models. Vedula and Mujumdar (1992); Vedula and Kumar (1996) and Ghahraman and Sepaskhah (2002) used, for example, a simple soil water balance for the design, via Stochastic Dynamic Programming (SDP), of a reservoir release policy under a multiple crops scenario. Shangguan et al. (2002) employed a similar balance equation for optimizing the allocation of irrigation water resources on a regional scale. The main drawback of this modelling approach consists in the absence of any calibration and in the large number of simplifications it requires.

In order to overcome these limitations, an attractive approach is the combined use of distributed-parameter and lumped-parameter models: granted that the former cannot be directly employed in optimization problems, it seems reasonable to employ the detailed outputs they provide for the improvement of the lumped-parameter models' quality. This can be achieved with techniques based on metamodelling. A metamodel is a simple, computationally efficient model that is identified on a data-set produced by simulating a large, distributed-parameter, physically based or conceptual model. Metamodelling, first introduced by Blanning (1975), is a well-known technique in mechanical and aerospace engineering (see Queipo et al., 2005, for a recent review), where it is widely employed to make optimization studies feasible. In relation to agricultural systems, the only field where metamodelling has already been applied and reported in the literature is the analysis of diffuse pollution in arable soils. One of the first applications was by Bouzaher et al. (1993), who developed a metamodel to predict the groundwater concentration of herbicides for a scenario analysis. Børgesen et al. (2001) and Haberlandt et al. (2002) employed a linear regression metamodel and a fuzzy-rule-based metamodel for the estimation of nitrogen leaching on a regional scale. To the authors' knowledge, apart from a recent work by van Walsum and Groenendijk (2008) concerning simulation efficiency, there are no reported applications of metamodelling to the estimation of the dynamics of irrigation water demand in cultivated areas.

The purpose of this paper is to introduce metamodelling in the field of agricultural water management and to test the design of a metamodel of a large, distributed-parameter, conceptual model (simply distributed-parameter model in the following), which describes the irrigation water demand of the Muzza district, located south-east of Milan (Italy) and fed by the River Adda (the Lake Como effluent). The goal of the metamodelling exercise is to strongly reduce the state dimensionality of the distributed-parameter model, i.e. to replace the distributed-parameter, two-layer, bucket model with a lumped-parameter, one-layer, bucket model, which can be subsequently employed to design, via SDP, the Lake Como release policy.

Even if the case we consider concerns the real-time control (management) of a reservoir, the approach we propose can be used for planning problems as well. For this reason, in the next section we describe how to use metamodelling within a planning/management problem and we set out the procedure we adopted to identify the metamodel. Section 3 presents the distributed-parameter model. Section 4 concerns the design of the metamodel. In Section 5 we briefly present the results that can be achieved by controlling the system with the designed release policy and we subsequently re-validate the metamodel on the trajectories obtained. In Section 6 we draw some conclusions on this work.

2. Metamodelling within planning/management problems

In the first part of this section, we introduce a procedure for using metamodelling within a planning/management problem by describing the management problem (the design of a reservoir release policy) that we will actually solve in the following sections. Then we will immediately see how the procedure can be straightforwardly extended to planning problems. In the second part, we describe the procedure that we employed for the identification of the metamodel. The reader interested in a survey of metamodelling techniques may usefully consider Simpson et al. (2001) or Queipo et al. (2005).

2.1. Design procedure

The optimal management of a water system, composed of a reservoir serving an irrigation district, requires us to design a release policy that (daily) suggests the amount of water to be released, given the reservoir storage and the irrigation water demand. The most efficient and consolidated approach for such design is based on the formulation and resolution of an Optimal Control Problem (OCP). Among the different methods that can be adopted for solving this kind of OCP, SDP is by far the most used (see Soncini-Sessa et al., 2007; or Castelletti et al., 2008b, for recent reviews of this topic). The main limit of SDP is that its computational requirements grow exponentially with the dimension of the state vector, so that, even when adopting the most advanced approaches (see, for example, Castelletti et al., 2008a), only few state variables can be used to model the water system.

In order to make the OCP practically solvable, the dynamics of all the water system components must thus be described with lumped-parameter models. As pointed out in Section 1, the reservoir storage and the inflow from the catchment can be easily and effectively described with this kind of model. As far as the dynamics of the irrigation water demand is concerned, the common approach (see Ghahraman and Sepaskhah, 2002, and references therein) consists in employing a lumped-parameter, conceptual model (i.e. a simple soil–water balance). Our proposal is to resort to a metamodel, which is a lumped-parameter, empirical model, identified on the data produced via simulation with a distributed-parameter model. This metamodel must present a small dimension of the state vector, so that the design of a reservoir release policy can be performed through the resolution of the OCP via SDP.

The *design procedure* we propose is shown in Fig. 1. In Phase A the metamodel is identified and validated on an Input–Output (I–O) data-set suitably designed via simulation with the distributed-parameter model (we will analyze this phase in greater detail in Section 2.2). Then (Phase B) the release policy is designed by solving an OCP, where the irrigation district dynamics is described with the metamodel. In Phase C the behavior of the water system controlled by the designed policy is simulated. In so doing, the irrigation district dynamics is described with the distributed-parameter model, and not with the metamodel, in order to obtain trajectories of its input and output over the simulation horizon, on the basis of which the metamodel can be validated for the second time (Phase D). If the performance measures are low, it means that the metamodel explanatory capacity is weak in the region of the I–O space where the system is driven by the designed policy. It is thus necessary to re-calibrate the metamodel by adding these trajectories to the calibration data-set considered in Phase A. On the contrary, if the validation is satisfactory, the procedure ends.

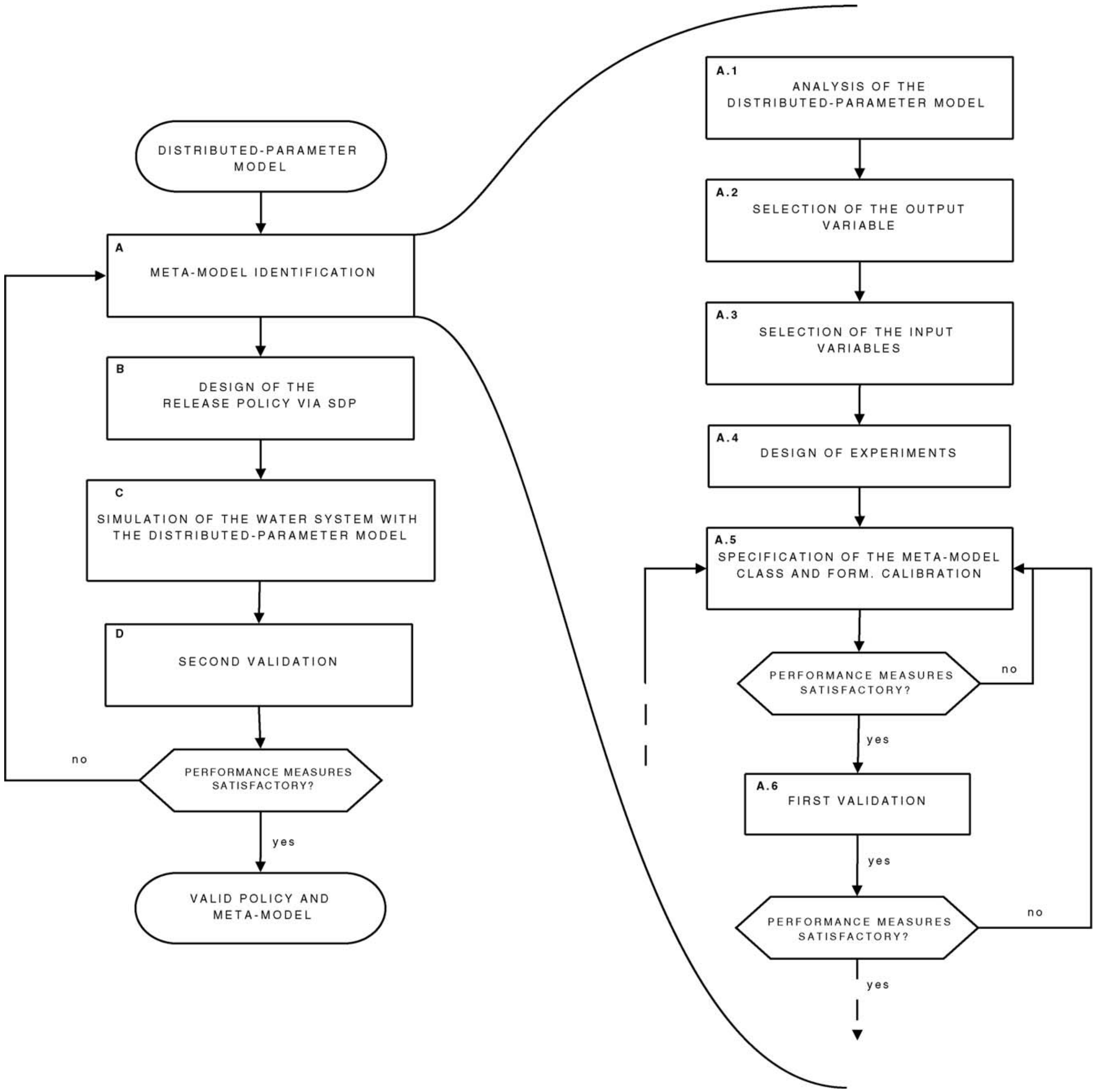

Fig. 1. The procedure for the design of a release policy using metamodelling and SDP. On the right-hand side an explosion of Phase A concerning the identification of the metamodel.

The reader can easily note that the resolution of the management problem (policy design) in Phase B can be straightforwardly substituted by the resolution of a planning problem (e.g. the optimal allocation of water among different crops), without affecting the rest of the procedure. Both management and planning problems, indeed, base their resolution on the use of optimization algorithms, which, in their turn, can actually be applied only when the state dimension of the model is low, i.e. a lumped-parameter model is used to describe the dynamics of the system we want to manage or plan.

2.2. Metamodel identification procedure

In order to identify the metamodel, we adopted the procedure shown on the right-hand side of Fig. 1. It is partially based on the methodology presented by Kleijnen and Sargent (2000). As is current practice in metamodelling, the procedure is valid for a Multi Input–Single Output (MISO) metamodel, because it permits us to build a metamodel emulating the dynamics of the variable we are interested in. When more variables are to be emulated (let's say n), i.e. when a Multi Input–Multi Output (MIMO) metamodel must be identified, the procedure can be applied n times, one for each MISO metamodel. Let us present the contents of each step.

A.1 The distributed-parameter model is analyzed, with the aim of gaining knowledge of the physics of the system it describes and expressing this knowledge with a causal network, which will be used in the following steps.

A.2 The metamodel output, namely the distributed-parameter model variable that we want to emulate, is defined. The definition is obviously based on the metamodel goal: in the present application, for example, in order to design the Lake

Como release policy, we need to forecast the one-step-ahead irrigation water demand of the Muzza district. Operationally, the definition consists in establishing the spatial and time domain on which some of the distributed-parameter model variables (eventually) have to be aggregated into one single variable, namely the metamodel output. The choice of the spatial domain can be based on physical considerations or on clustering techniques (see Jain et al., 1999, for a review of the principal cluster analysis methods), while the temporal domain depends on the relative size of the decision time-step and of the simulation time-step adopted in the distributed-parameter model. Finally, the range of variability of the metamodel output can be derived from physical considerations (as in our case) or through simulation (step A.4). This last information will turn out to be useful in the release policy design (Phase C).

A.3 One has to identify, through the causal network, those input variables that affect the metamodel output. When they turn out to be too many, one has to identify the most important one through physical considerations or data-driven techniques (e.g. correlation analysis). If the selected input variables are spatially and temporally distributed, they must be aggregated with the same spatial and temporal resolution adopted for the metamodel output, so that consistency is maintained between the definition of the output and input variables. The spatial and temporal aggregation, moreover, permits us to strongly reduce the number of input variables.

A.4 This step is known in the metamodelling literature as Experimental Design (ED) or Design of Experiments (DOE) and its concern is to produce, via simulation with the distributed-parameter model, an I–O data-set. This set will be split into two sub-sets to be used for calibrating and validating the metamodel in the next step. The DOE aims at exploring the I–O space while considering the computational burden induced by each simulation with the distributed-parameter model. It can possibly be based on statistical techniques (for a recent review, see Kleijnen et al., 2005).

A.5 The class of models (e.g. autoregressive-exogenous or data-based mechanistic models, splines or neural networks, etc.) to which the metamodel will belong is first chosen. Then, the metamodel form (e.g. the number of neurons and hidden layers of a neural network) within the selected class is specified. Finally, the metamodel parameters are estimated with a suitable algorithm. If the performance measures (e.g. the coefficient of determination R^2) on the calibration data-set are satisfactory, one proceeds to the next phase; otherwise, the form or the class of the metamodel is changed and the step repeated.

A.6 If the metamodel provides satisfactory performance measures also on the validation data-set, it can be considered valid on the designed I–O space. The metamodel can then be used in the OCP for the design of the reservoir release policy. If the performance measures are not satisfactory, it is necessary to return to the previous step, considering a different form or class for the metamodel.

3. The Muzza distributed-parameter model

The Muzza–Bassa Lodigiana district is located in the Padana Plain (Italy), south-east of the city of Milan. It has an area of about 700 km^2 and irrigation is practised there with the border method (or free-surface flooding). The cultivated area covers 85% of the whole district; major crops are cereals (especially maize) and permanent grass. The source of the irrigation supply is the Muzza main canal, which originates from the River Adda, the Lake Como effluent. The irrigation supply is thus controlled by the release from the lake, which is regulated through the Olginate dam. A schematic view of the whole system is given in Fig. 2.

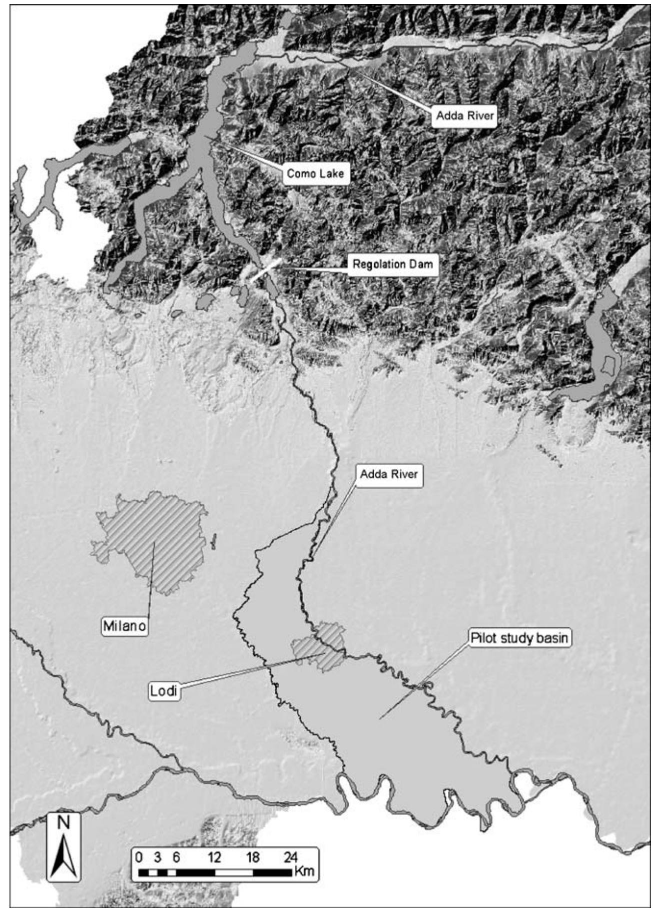

Fig. 2. Schematic view of Lake Como, its effluent (the River Adda) and the pilot study basin (the Muzza district).

A series of significant droughts in the last ten years has increased interest in water savings and attention has been paid to the water losses produced when the flow diverted into the Muzza main canal is higher than the irrigation water demand (and the net diverted flow is thus higher than the irrigation supply), as, for instance, happens during precipitation events. These losses could be decreased by designing a lake release policy that depends not only on the lake storage, but also on the actual irrigation demand. However, the Muzza district's gravity flow canals are characterized by high inertia, which does not permit us to change their flow rapidly in order to follow the rapid fluctuation of the irrigation water demand. As a consequence, the implementation of a release policy based on-demand requires a modification of the canals, such that they can be regulated on-demand (it is for this reason that the lake regulator presently tends to track a target demand: an a priori given trajectory of the water demand).

This is a strong issue in canal management, and some recent papers (notably Mareels et al., 2005) showed that the rehabilitation of open canal conveyance and distribution networks of large irrigation districts can be such that they can be practically regulated on-demand. This is obtained by combining the classical use of regulation structures (*gates*), in order to transform the gravity flow canals into a cascade of *pools*, with sensors and control algorithms to operate automatically all the structures. The routing and scheduling of water is then realized through an automatic adjustment of the gates. With this modification, within the limit of a fraction of the pool capacity, a water volume appears to be shifted along the canals with a velocity that is not the water velocity (about 2 km/h in the

Muzza district case), but the celerity of the backward surface perturbations,[1] generated by the gates' movement. The inertia of the canals is thus strongly reduced and, within limits imposed by the volume of the smallest pool, the irrigation demand can be satisfied almost in real time.

Since our policy design problem needs to be combined with on-demand regulation of canals, two assumptions were made in building the distributed-parameter model. First, it was assumed that the modification of the canals is such that their inertia becomes negligible, which means that their flow can be considered of the plug-flow type. This does not mean that the canals' capacity is unlimited, but that their flow can be varied almost instantaneously and thus quickly adapted to the fast-varying irrigation water demand. Second, it was assumed that the transport time is null, since, considering the Muzza district's size, the decision time-step of the management problem (1 day) is definitely larger than the time required for a released volume to reach the most distant field in the district. These two assumptions together imply that the volume released from the lake is delivered to the crops in a time that is shorter than the decision time step, which means that the canal dynamic does not have to be considered. Therefore, the modelling effort was concentrated on the hydrological water balance in the district.

3.1. The model

The model implemented for the Muzza district is a large, distributed-parameter, conceptual model, which allows the simulation of the irrigation water distribution and the computation of the hydrologic balance in the root zone in the whole district on a daily basis. The model includes a number of modules, devoted to specific tasks: water sources, conveyance and distribution, soil–crop water balance.

The water balance module (Facchi et al., 2004) accounts for the space variability of soils and crops, as well as of meteorological and irrigation inputs, by subdividing the irrigation district with a regular mesh: soil and crop characteristics as well as meteorological inputs and irrigation supply are homogeneous in each cell of the mesh but may vary from cell to cell. Each individual cell identifies a soil volume which extends from the soil surface to the lower limit of the root zone, and a one-dimensional representation of the hydrological processes is adopted within it. The soil volume of each cell is subdivided into two layers: the upper one (evaporative layer) represents the upper 15 cm of the soil; the bottom one (transpirative layer) represents the root zone and has a time-varying depth $Z_{r,t}^{(i)}$. The two layers are modelled as two non-linear reservoirs in cascade (see Fig. 3). The water percolating out of the bottom layer constitutes the recharge to the groundwater system.

The dynamics of the water content $U_{1,t}^{(i)}$ [mm] in the evaporative (first) layer of the i-th cell is governed by the following balance equation:

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

where $R_{t+1}^{(i)}$ [mm] is the rainfall, $I_{t+1}^{(i)}$ [mm] is the canopy interception, $Q_{r,t+1}^{(i)}$ [mm] is the net runoff from the cell, $E_{t+1}^{(i)}$ [mm] is the evaporation, $Q_{u,t+1}^{(i)}$ [mm] is the outflow to the transpirative layer and $Q_{i,t+1}^{(i)}$ [mm] is the irrigation supply, all in the time interval $[t, t+1)$.[2]

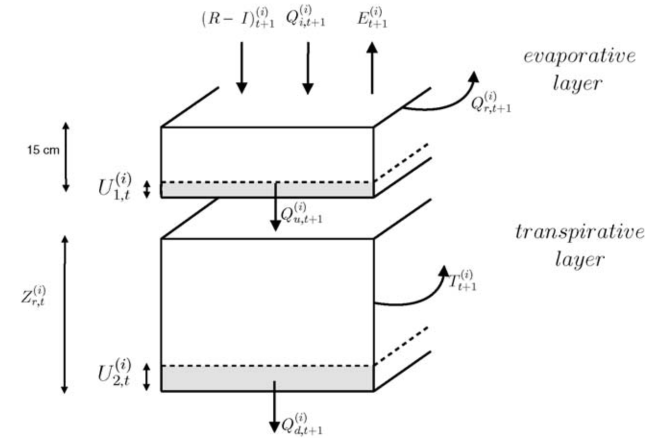

Fig. 3. Sketch of the processes being modelled in the ith cell by the distributed-parameter model. The significance of the symbols and of the arrows is explained within Section 3.1.

A similar equation holds for the dynamics of the water content $U_{2,t}^{(i)}$ [mm] (the so-called green water) in the transpirative (second) layer

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

where $T_{t+1}^{(i)}$ [mm] is the transpiration and $Q_{d,t+1}^{(i)}$ [mm] is the outflow from the root zone to the groundwater system, both in the time interval $[t, t+1)$.

The evaporation $E_{t+1}^{(i)}$ and the transpiration $T_{t+1}^{(i)}$, in Eqs. (1) and (2) respectively, are computed using the FAO-56 dual crop coefficient method (Allen et al., 1998): the evaporation $E_{t+1}^{(i)}$ is determined by multiplying the reference crop evapotranspiration $ET0_{t+1}^{(i)}$ [mm] (computed with the FAO-Penman–Monteith equation) by the evaporative coefficient $K_{e,t}^{(i)}$ [−] (which depends on $U_{1,t}^{(i)}$). The transpiration $T_{t+1}^{(i)}$ is obtained by multiplying $ET0_{t+1}^{(i)}$ by two coefficients. The first is the water stress coefficient $K_{s,t}^{(i)}$ [−] (whose value depends on $U_{2,t}^{(i)}$), which expresses the effect of limited soil water availability; the second is the basal coefficient $K_{cb}^{(i)}$ [−], which accounts for the differences in the crop canopy and aerodynamic resistances relative to the reference crop. It follows that

$$T_{t+1}^{(i)} = K_{s,t}^{(i)}(U_{2,t}^{(i)}) \cdot K_{cb}^{(i)} \cdot ET0_{t+1}^{(i)} \quad (3)$$

By observing that the product $K_{cb}^{(i)} \cdot ET0_{t+1}^{(i)}$ is the potential transpiration $T_{c,t+1}^{(i)}$ [mm] (i.e. the crop transpiration in the condition of optimal water content) we can rewrite Eq. (3) as follows:

$$T_{t+1}^{(i)} = K_{s,t}^{(i)}(U_{2,t}^{(i)}) \cdot T_{c,t+1}^{(i)} \quad (4)$$

Drainage discharges $Q_{u,t+1}^{(i)}$ and $Q_{d,t+1}^{(i)}$ are determined using a simplified scheme, similar to those used in other models (e.g. ANSWERS2000, Bouraoui et al., 1997; EPIC, Williams et al., 1984), which consider a Darcian-type gravity flow in the unsaturated soil (Gandolfi et al., 2006).

The runoff rate $Q_{r,t+1}^{(i)}$ in Eq. (1) is calculated by using the SCS-Curve Number method (USDA-SCS, 1972, 1986), which is used for agricultural areas by models like AGNPS (Young et al., 1989) and AQUACROP (Steduto et al., 2009). The canopy interception $I_{t+1}^{(i)}$ is evaluated by the Braden formula (Braden, 1985), as a function of the leaf area index, the cover fraction and the volume capacity per unit foliage area, which are variable according to the crop type and the growing stage. Storm runoff, as well as irrigation tail-waters, are collected by the drainage network and may eventually complement the supply $Q_{i,t+1}^{(i)}$ to cells of downstream units.

[1] This celerity is equal to $(\sqrt{g \cdot h} - v)$, where g is the gravitational acceleration, h is the water depth in the canal and v is the flow velocity. Considering a depth h of 1.5–3.0 m, for instance, the celerity is about 12–18 km/h.

[2] The subscript in the symbol of a state variables, as $U_{1,t}^{(i)}$, denotes the time at which its value is considered, while the subscript of a flow variable, as $R_{t+1}^{(i)}$, denotes the termination time of the period in which the flow occurs.

During the irrigation season (1 April to 31 October) the water flow q_{t+1}^{Muzza} [m^3/s] diverted from the River Adda into the Muzza main canal, which is the only water source, is delivered to the district. The water available for irrigation (simply net diverted flow in the following) is equal to the diverted flow q_{t+1}^{Muzza} multiplied by the conveyance and distribution efficiency η. Due to the hypothesis that gates will be placed in the canals, in order to regulate them on-demand, propagation processes within the network were not considered in the model. Regulation on-demand implies that an irrigation application occurs (i.e. the daily irrigation supply $Q_{i,t+1}^{(i)}$ is non-zero) only when two conditions are satisfied.

First, the soil moisture deficit $D_{2,t}^{(i)}$ [mm] (defined as the field capacity $U_{2,fc}^{(i)}$ [mm] minus the soil water content $U_{2,t}^{(i)}$) is higher than a threshold value

$$D_{2,t}^{(i)} > \alpha \cdot RAW_t^{(i)} \tag{5}$$

where α is a coefficient lower than 1 and $RAW_t^{(i)}$ [mm] is the soil water content readily available to the crop (Allen et al., 1998). Eq. (5) can be rewritten as

$$U_{2,t}^{(i)} < U_{2,fc}^{(i)} - \alpha \cdot RAW_t^{(i)} \tag{6}$$

and therefore the first condition requires that the soil water content $U_{2,t}^{(i)}$ in the transpirative layer fall below the threshold $U_{r,t}^{(i)} = U_{2,fc}^{(i)} - \alpha \cdot RAW_t^{(i)}$ [mm].

The second condition is that the irrigation system can provide water to irrigate the cell. When $Q_{i,t+1}^{(i)}$ is non-zero, its value was assumed to be 180 mm/day for each cell in the district (the quantity averagely used for the border method).

The description of the water conveyance and distribution reflects the typical structure of the irrigation network in most districts of the Padana plain. The district is subdivided into a number of units that receive irrigation supply from one or more sources. During the irrigation season, water derived from the sources is conveyed to the units and each of them receives a fixed share of the flow that is diverted from the sources supplying that specific unit through the irrigation network. In practice an incidence matrix is used to represent the links between the sources and the units, and two numerical values are associated with each active link: the fraction of the diverted discharge which is conveyed to the unit and the conveyance efficiency. At the moment, the module does not fully account for flow propagation processes within the channels of the irrigation network, as only time delays between diversion and delivery can be taken into account. Once water is delivered to a unit, distribution within the unit takes place either on a demand basis or on a rotation basis; in the latter case, in each day a fixed number of cells is explored to check if irrigation is required and a cell is actually irrigated only if the soil water content, provided by the soil volume balance model, is below a given threshold. The number of explored cells is a function rotation period (turn), which is a characteristic of each unit and may vary within the district depending mainly on soil and crop types. Irrigation tail-waters from a unit are collected by the drainage network and may complement the water supply to downstream units.

3.2. Model inputs and calibration

The model inputs are the time series of the average daily flow diverted by the water sources and of the following meteorological variables: rainfall, maximum and minimum temperature, wind speed, maximum and minimum relative humidity, solar radiation (simply meteorological inputs in the following). Meteorological inputs must be available for a number of stations that is sufficient to account for the spatial variability of the variables within the district. Spatial interpolation of the daily values in the different stations is carried out inside the model.

Given the land use and the input series, the model computes the corresponding daily time series of the water balance terms for each cell of the basin. The irrigation demand of the district is computed in each day as the sum of the irrigation demands of all the cells where soil water content is below the threshold defined in Eq. (6). Model implementation for the Muzza district was based on a careful collection and collation of all the available physiographic, meteorological and hydrometric data, as well as on a thorough investigation of the irrigation management rules and criteria. The Muzza canal is the only source of irrigation supply and, through a network of smaller channels, conveys water to 65 irrigation units. The diverted flow is distributed among the units, each receiving a fixed share of the total flow. Time series of the Muzza flow and of the meteorological variables in six agro-meteorological stations were available for the period 1993-2004. Land use maps were obtained from the Agricultural Agency of the Lombardia Region, which publishes yearly digital maps of the main crops, derived from the farmers' land use statements for CAP payments. The crop parameters (crop coefficient, leaf area index, root depth, and crop height) were obtained by combining different literature references: Po River Authority (PRA, 1998), Baldoni and Giardini (1982), Huygen et al. (1997), Allen et al. (1998), and USDA-SCS (1972, 1986). The duration of the phenological phases for each crop was adapted from (Donatelli, 1995; Stockle and Nelson, 1996), using local observations. Soil parameters were derived from the 1:250,000 pedological map of the district, which includes 60 main pedological units, combined with field observations of the physico-chemical characteristics of the soil profiles which are representative of each unit. The parameter values were computed by using so-called Pedo-Transfer Functions (PTFs), i.e. empirical relationships that relate the values of the main soil parameters (e.g., water content at field capacity and wilting point, saturated hydraulic conductivity) to the values of selected physico-chemical variables (e.g., bulk density, porosity, organic matter). The PTFs of Rawls and Brakensiek (1989) were selected among the different types of functions that are available in the literature (see Wösten et al., 2001; Gijsman et al., 2003 for a review), as they proved to give the best performances for the soils of the area (Ungaro and Calzolari, 2001; Baroni et al., 2009). A mesh size of 250×250 m^2 was selected as a good compromise between discretization accuracy and detail afforded by the available information. In order to check the mesh adequacy, we conducted a pilot study on a smaller area (17 km^2), where a 1:50,000 pedological map was available, using also a mesh size of 25×25 m^2; the results obtained shown that the differences in simulated irrigation demand become very small when aggregated values over areas larger than a few square km are considered.

The unavailability of distributed, continuous measurements of soil water content, which is a typical limitation when dealing with large irrigation systems modeling, hindered the calibration of the model. The only observation data that could be effectively used for calibration is the number of irrigation applications during the crop season in a number of irrigation units, which were provided by the irrigation Consortium for a few years. The measured values of the pumping rates in a few units, adjacent to the Muzza district, where irrigation supply is provided by gauged pumping wells, were used to support calibration. For obvious reasons, calibration had to be limited to a minimum number of parameters. We kept the soil hydraulic parameters fixed to the values that were obtained by applying the pedo-transfer functions of Rawls and Brakensiek (1989) and the conveyance, distribution and irrigation efficiencies fixed to the values that were provided by the Muzza district. Only the parameter that pertains to the application of irrigation to the individual cells (namely the soil water deficit threshold for irrigation) was calibrated. Reasonable values of this parameter

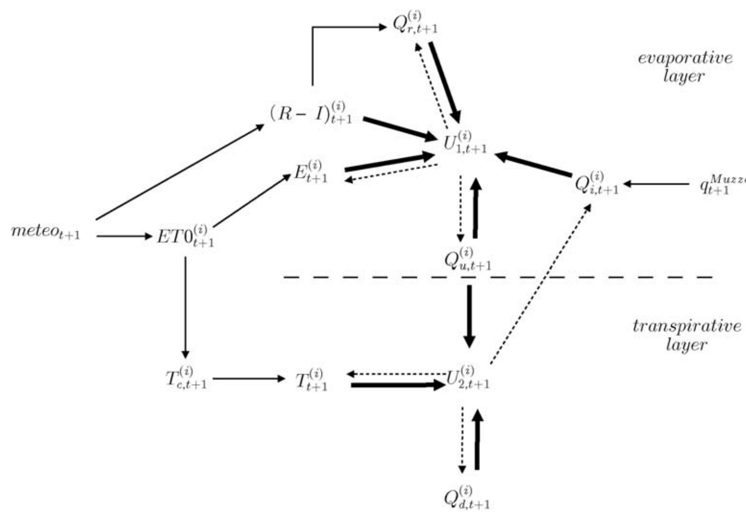

Fig. 4. Causal network of the processes modelled in the *i*th cell of the distributed-parameter model. Symbols are defined in Section 3; the influences represented by dotted lines have effect with one-step delay.

should fall in the range 0.6–0.9, reflecting the precautionary point of view of farmers that require irrigation before the stress condition is reached, in order to prevent damages if the irrigation is actually available only a few days later than when the demand is expressed. Indeed, a value of 0.8 gave a satisfactory agreement between the simulated and observed values of the number of irrigations. A qualitative model validation was based on the comparison of the registered and simulated deficit episodes in critical years (e.g. the 2003 drought).

A simplified version of the Muzza model was used as a basis for the metamodel identification: following the assumptions on the rehabilitation of the conveyance and distribution network, illustrated in the previous section, the whole district was simulated as a unique, large irrigation unit, where water allocation to the cells occurs on-demand, with a priority given to the cells which express their demand first, in the case that the water demand in a given day exceeds the Muzza flow in the same day. Irrigation supply to a cell occurs, provided that enough water is available, with a time delay in the network that is smaller than the decision time step (one day) of the water release from Lake Como.

4. Identification of the Muzza metamodel

The distributed-parameter model has a spatial domain D of something like 10^4 cells. Since the model employs two state variables ($U^{(i)}_{1,t}$ and $U^{(i)}_{2,t}$) for each cell, the dimension of its state vector I_t is of the order of 2×10^4: a figure which is totally incompatible with algorithms based on SDP. In order to design the Lake Como release policy, while describing dynamically the state of the Muzza district, it is therefore necessary to resort to a metamodel. The metamodel is characterized by a daily time-step, which corresponds to the decision time-step of the management problem. This section describes its identification, based on the procedure shown on the right-hand side of Fig. 1.

4.1. Step A.1: analysis of the distributed-parameter model

In order to drive the metamodel identification, we derived the causal-network of the processes occurring in the *i*th cell (see Fig. 4). Its construction is carried out in three steps:

- From the balance Eqs. (1) and (2), we derived the direct influence of each term on the state variables $U^{(i)}_{1,t}$ and $U^{(i)}_{2,t}$. The daily irrigation supply $Q^{(i)}_{i,t+1}$, for example, has a direct influence on $U^{(i)}_{1,t+1}$. Direct influences are represented by bold arrows in Fig. 4.
- We accounted for feedbacks between the state variables and the other equation terms. The daily irrigation supply $Q^{(i)}_{i,t+1}$, for example, depends on $U^{(i)}_{2,t}$. Feedback influences are represented by dotted lines.
- Finally, we considered the relations between terms (the net rainfall $(R-I)^{(i)}_{t+1}$, for instance, affects the runoff $Q^{(i)}_{r,t+1}$) and the relations between terms and model inputs (the daily irrigation supply $Q^{(i)}_{i,t+1}$, for example, depends on the diverted flow q^{Muzza}_{t+1}).

4.2. Step A.2: selection of the output variable

As stated, the metamodelling exercise is undertaken in order to design, via SDP, a release policy that (daily) suggests the amount of water to be released from Lake Como, given the current reservoir storage s_t [m^3] and the irrigation water demand W_t [m^3/s] of the Muzza district (defined at the main canal intake). As a consequence, it follows that the metamodel must supply the one-step ahead forecast w_{t+1} of the irrigation water demand, given W_t. In the metamodel identification, the value of W_t is obtained by aggregating the daily water demand of each cell of the distributed-parameter model on its spatial domain D; while, in the subsequent application of the policy, its value will be supplied to the Lake regulator by the Muzza district regulator, who collects the individual estimates made by each farmer on his/her own field. In their turn, farmers can estimate their irrigation water demands either on the basis of their experience alone, or by means of networks of soil moisture sensors.

Since in the Muzza district the irrigation is practised with the border method, we can assume that the same volume of water is applied to each cell in deficit condition, independently of the actual size of its water deficit (granted that $U^{(i)}_{2,t} < U^{(i)}_{r,t}$). Then

$$W_t = \frac{V}{\eta \cdot \Delta} \cdot N_t \tag{7a}$$

where V [m^3] is the volume[3] supplied to an irrigated cell; η (equal to 0.65) a parameter accounting for both conveyance and distribution efficiency in the district (remember that W_t is defined at the main canal intake); Δ [s] the length of the decision time-step (one day, i.e. 86,400 s) and N_t the number of cells in deficit

[3] $V = 180 \times 10^{-3} \times S^{cell}$ m^3, where $S^{cell} = 62500$ m^2 is the cell surface.

condition at time t in the domain D, namely

$$N_t = \sum_{i=1}^{N^{tot}} \nu_t^{(i)} \quad \text{with} \quad \nu_t^{(i)} = \begin{cases} 1 & \text{if } U_{2,t}^{(i)} < U_{r,t}^{(i)} \\ 0 & \text{otherwise} \end{cases} \tag{7b}$$

where N^{tot} is the number of cells of D.

From Eq. (7a), it appears natural to compute the forecast w_{t+1} with the following equation:

$$w_{t+1} = \frac{V}{\eta \cdot \Delta} \cdot n_{t+1} \tag{8}$$

where n_{t+1} is the forecast of the number N_{t+1} of cells in deficit condition at time $t+1$. As a consequence, we decided to adopt, as state of the metamodel, the estimate n_t of the number N_t of cells in deficit condition, whose values can be

- estimated at time t by inverting Eq. (7a), given the value supplied by the Muzza regulator (or by the distributed-parameter model), namely

$$n_t = \frac{\eta \cdot \Delta}{V} \cdot W_t \tag{9}$$

- forecasted at time $t+1$ through the recursive equation we are going to define in step A.5. According to this position, the metamodel output is the forecasted irrigation water demand w_{t+1}, computed through Eq. (8).

The value of the water demand can vary between 0 m^3/s (when all the cells are in non-deficit condition) and about 1100 m^3/s, which is reached when all the N^{tot} cells in D are in deficit condition.

A comment is essential on the adoption of the estimate n_t of cells in deficit condition as state of the metamodel: the presence of the threshold $U_{r,t}^{(i)}$ in Eq. (7b) introduces a stochastic effect into the dynamics of n_t. If at time t the value of N_t is lower than the total number N^{tot} of cells, its value at time $t+1$ will depend not only on the evolution of the conditions of those cells that were in deficit condition at time t (and therefore on N_t), but also on the number of cells that will cross the threshold $U_{r,t}^{(i)}$ in the time interval $[t, t+1)$. This last value is unpredictable at time t, when only N_t is known. It follows that we should expect that the accuracy of the estimate n_{t+1} of N_{t+1} increases with the number N_t of cells in deficit condition at time t. In Section 4.5 we will show how to deal with this stochastic effect.

4.3. Step A.3: identification of the input variables

With respect to the previous step, the input variables selection is a more sophisticated process, which requires further physical considerations. We will first consider the dynamics of one single cell (referring to the causal network of Fig. 4) and develop the reasoning as if the metamodel were to approximate one cell. The extension to the domain D will just require us to consider N^{tot} cells, instead of one only. Before proceeding, it must be stressed that the selected inputs should not depend on the distributed-parameter model state variables $U_{1,t}^{(i)}$ and $U_{2,t}^{(i)}$ (belonging to the state vector I_t), since they are not components of the metamodel state (and they cannot be, since they are spatially distributed).

According to Eq. (7), the irrigation water demand W_t depends on the number N_t of cells in deficit condition. In the ith cell this condition is related to the water content $U_{2,t}^{(i)}$, whose dynamics (see the causal network) is driven by the transpiration $T_{t+1}^{(i)}$, the percolation $Q_{u,t+1}^{(i)}$ from the upper layer and the deep percolation $Q_{d,t+1}^{(i)}$. These variables are then potential candidates to be metamodel inputs.

As for the transpiration $T_{t+1}^{(i)}$, the causal network shows that, in its turn, it is a function of $U_{2,t}^{(i)}$, and thus it cannot be an input. However, it obviously depends on the potential transpiration $T_{c,t+1}^{(i)}$, which, in turn, depends on the meteorological variables only. Therefore, the total potential transpiration $T_{c,t+1}$ [m^3/s] on the domain D

$$T_{c,t+1} = \frac{\sum_{i=1}^{N^{tot}} T_{c,t+1}^{(i)} \cdot S^{cell}}{10^3 \cdot \Delta} \tag{10}$$

is a candidate metamodel input variable.

As far as the percolation $Q_{u,t+1}^{(i)}$ is concerned, it depends, through $U_{1,t}^{(i)}$, on the net rainfall $(R-I)_{t+1}^{(i)}$, the evaporation $E_{t+1}^{(i)}$, the surface runoff $Q_{r,t+1}^{(i)}$ and the irrigation supply $Q_{i,t+1}^{(i)}$. Compared to the other two variables, the contribution of $E_{t+1}^{(i)}$ and $Q_{r,t+1}^{(i)}$ is scarce, so that they can be disregarded. Evaporation indeed becomes small, compared to transpiration, when crops reach the full growth, due to the soil coverage by the crop canopy. Runoff from the cells may be due to rainfall or irrigation: storm runoff is negligible owing to the limited slope and to the generally high retention capacity of agricultural soil and vegetation; irrigation tail-water may occur, but its amount is small compared to the gross irrigation demand. The contribution of the net rainfall $(R-I)_{t+1}^{(i)}$ is conspicuous and it could be a second candidate input variable. However, it can be observed that the potential transpiration is influenced by cloud coverage and air humidity and that its values drop to a minimum when rainfall events occur. The potential transpiration thus holds some information on rainfall occurrence and, in order to maintain the number of input variables as small as possible, it is therefore reasonable to leave out $(R-I)_{t+1}^{(i)}$.

Only the influence of the irrigation supply $Q_{i,t+1}^{(i)}$ on $Q_{u,t+1}^{(i)}$ remains to be explained, and then it would appear that the total irrigation supply $Q_{i,t+1}$ [m^3/s]

$$Q_{i,t+1} = \frac{\sum_{i=1}^{N^{tot}} Q_{i,t+1}^{(i)} \cdot S^{cell}}{10^3 \cdot \Delta} \tag{11}$$

is the third candidate for metamodel input. However, the causal network shows that $Q_{i,t+1}^{(i)}$ does depend on the state $U_{2,t}^{(i)}$, since the ith cell is not irrigated when $U_{2,t}^{(i)} > U_{r,t}^{(i)}$. As a consequence, the total irrigation supply $Q_{i,t+1}$ depends on the state of the distributed-parameter model and thus it cannot be a metamodel input. To overcome this difficulty, observe that the value of $Q_{i,t+1}$ is driven by $\eta \cdot q_{t+1}^{Muzza}$, where q_{t+1}^{Muzza} is the flow diverted from the River Adda into the Muzza main canal. In other words, q_{t+1}^{Muzza} can be considered a proxy of $Q_{i,t+1}$, and as such we opted for adopting it as the second metamodel input. Notice that the approximation will be the better, the lower are the losses of water (i.e. when $Q_{i,t+1}$ nearly equals $\eta \cdot q_{t+1}^{Muzza}$). We should expect that also the metamodel performances will follow the same behavior.

In conclusion, we will consider as inputs of the metamodel the total potential transpiration $T_{c,t+1}$ and the diverted flow q_{t+1}^{Muzza} and as output the estimated irrigation water demand w_{t+1}.

4.4. Step A.4: design of experiments

We now need to generate, via simulation with the distributed-parameter model, the data-sets that will be used for calibrating and validating the metamodel. Since these data-sets directly affect the metamodel range of validity, we need to determine carefully the values of the two input variables ($T_{c,t}$ and q_t^{Muzza}) for which the metamodel must be valid.

As far as the potential transpiration $T_{c,t}$ is concerned, we have the historical measurements of the meteorological variables over the period 1993–2004. We judged that the values assumed by $T_{c,t}$ over this period cover a sufficiently large range, and we thus decided to employ just this meteorological scenario, varying only, as subsequently shown, the diverted flow q_t^{Muzza}. This decision is based on the following consideration: $T_{c,t}$ is just a function of the meteorological variables and, as such, it cannot be controlled. From the control point of view it is a disturbance. On the contrary, the diverted flow q_t^{Muzza} is directly influenced by the release decision from Lake Como, through which the condition of the irrigation district can be controlled. Moreover, since the metamodel will be used within an SDP algorithm, we need to make it valid for a large range of possible diverted flows. Therefore, we need to explore the behavior of the system for different trajectories of q_t^{Muzza}, while considering that each simulation with the distributed-parameter model is strongly time-consuming. Moreover, consider that such trajectories must be physically meaningful: it does not make sense, for example, to use values of q_t^{Muzza} causing the soil moisture to be lower than the wilting point, since it would lead to the crops withering. This means that we cannot generate purely random values of q_t^{Muzza}, but we need to define a priori a set of trajectories that we expect to keep the soil moisture between the wilting point and the field capacity. For defining these trajectories it is important to remember that q_t^{Muzza} can vary between 0 and 110 m^3/s (the capacity of the Muzza main canal).

Given this premise, we decided to run, over the period 1993–2004, four simulations with the following trajectories (*scenarios*) of the diverted flow q_t^{Muzza}:

Scenario 1: historical flow;
Scenario 2: constant flow equal to 70% of the capacity, i.e. 77 m^3/s;
Scenario 3: constant flow equal to the Muzza main canal capacity, i.e. 110 m^3/s;
Scenario 4: a piece-wise constant function, obtained by maintaining constant, for short periods of the order of few weeks, a value of q_t^{Muzza} randomly selected between 0 and 110 m^3/s.

Via simulation with the distributed-parameter model we thus obtained four data-sets, plus a fifth one composed of the aggregation of the first four.

To give the reader a more physical feeling, we report in Fig. 5 the trajectories of the metamodel I–O variables in the first year of Scenario 1 (historical flow). Panel (a) shows that the demand W_t is null during the winter period, while it starts to increase only a few weeks before the beginning of the irrigation season, because of the permanent grass requirements (see its potential transpiration in panel (c)). Once the irrigation season starts (panel (b)), the total irrigation supply $Q_{i,t}$ follows the water demand, subject to the water availability provided by the net diverted flow $\eta \cdot q_t^{Muzza}$. Notice how this last variable is often larger than $Q_{i,t}$, since, as explained in Section 1, the inertia of the canal prevents the Muzza district regulator from following the rapidly varying water demand.

Fig. 6 gives a 3-D representation of the values of $T_{c,t}, Q_{i,t}$ and W_t that we obtained for each scenario. Scenario 1, based on historical values of q_t^{Muzza}, is characterized by values of W_t between 0 and 400 m^3/s. This last value is generally in correspondence with large values of transpiration and (then) irrigation supply. On average, in Scenarios 2 and 3, the irrigation water demand is lower, because of the larger value of the diverted flow (77 and 110 m^3/s respectively) and, thus, of the total irrigation supply. Finally, in the last scenario the points are distributed more uniformly in the I–O space. For all scenarios, note that most of the points are concentrated around the origin: this is due to the winter period, when the potential transpiration is so low that, even if irrigation is not active, the corresponding water demand is null.

The first seven years (1993–1999) of each data-set were used for the metamodel calibration, while the remaining five (2000–2004) were used for its validation.

4.5. Step A.5: metamodel class and form, calibration

The selection of the metamodel class requires us to consider two different aspects of this metamodelling exercise. On one hand, as introduced in Section 4.2, we need to treat the stochastic effect in the dynamics of n_t; on the other hand, we aim at creating a metamodel that has a physically meaningful interpretation, in order to maintain a relation with the distributed-parameter, conceptual model. For this second reason, we discarded black-box models (e.g. neural networks) and we adopted the class of state-dependent parameter models (Young, 1998).

For the metamodel transition function, we adopted a mass-balance equation applied to the estimated number n_t of cells in deficit condition

$$n_{t+1} = n_t - n_{t+1}^{irr} + \chi_{t+1} + \varepsilon_{t+1} \tag{12a}$$

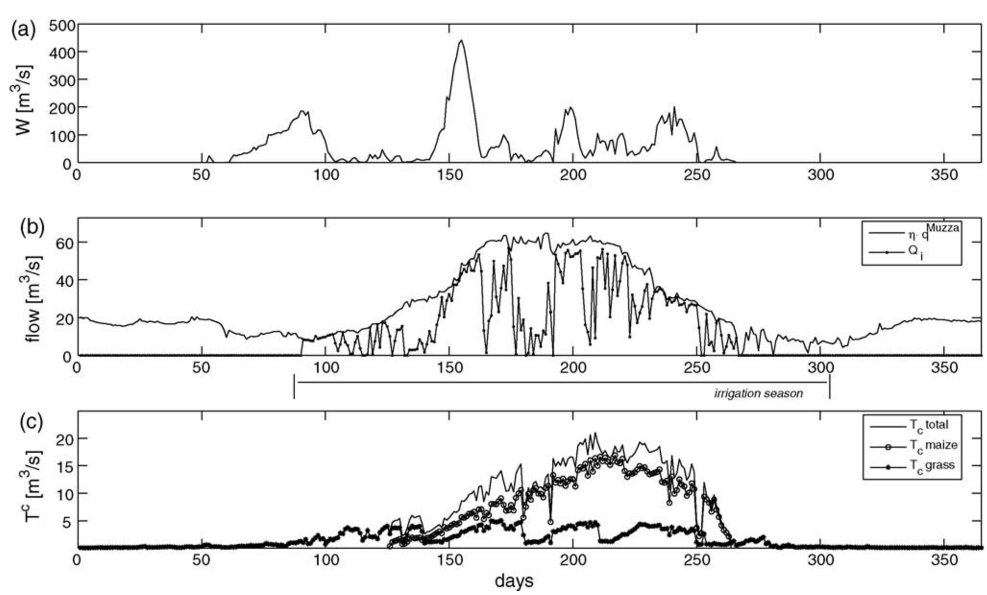

Fig. 5. Trajectories, over the first year of scenario 1, of the irrigation water demand W_t (panel (a)); the net diverted flow $\eta \cdot q_t^{Muzza}$ and the total irrigation supply $Q_{i,t}$ (panel (b)); the total potential transpiration $T_{c,t}$ and those of maize and permanent grass (in panel (c)).

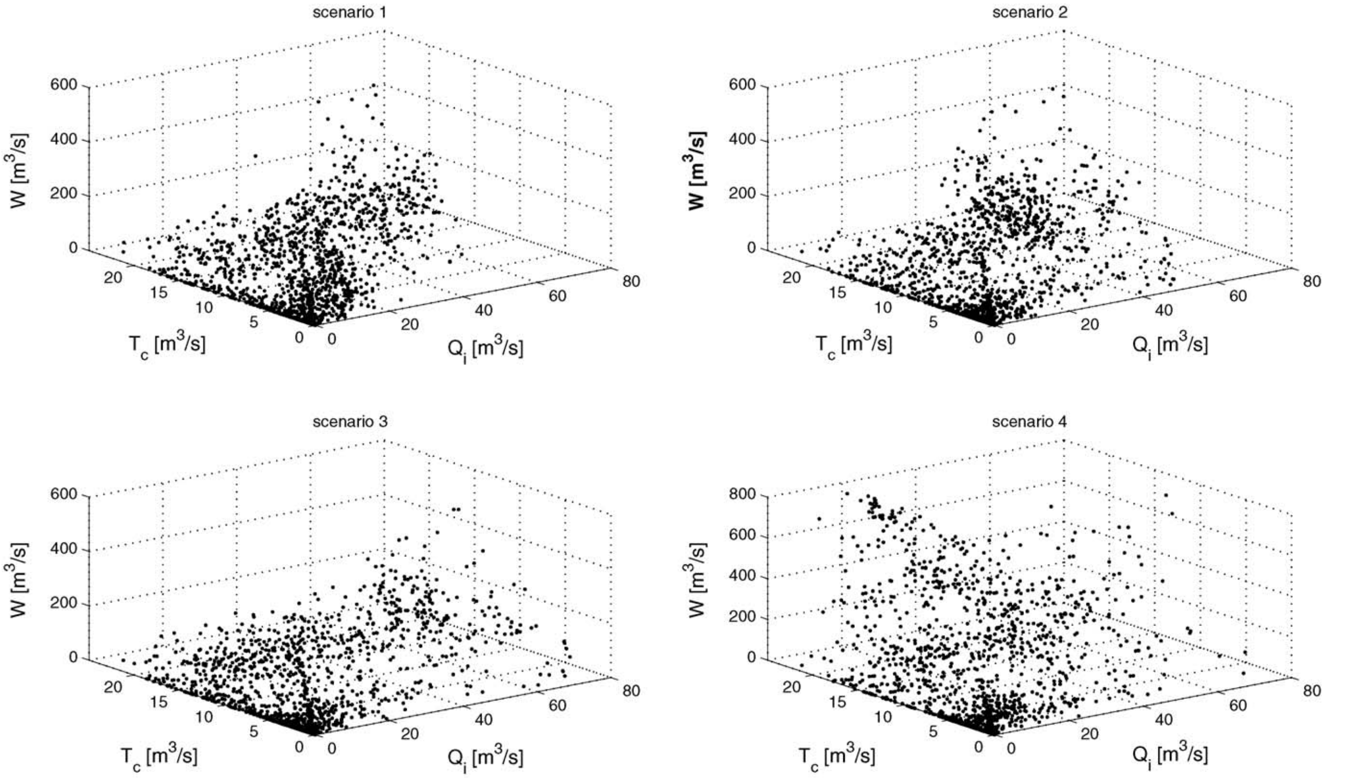

Fig. 6. Values of the irrigation water demand W_t, as a function of the potential transpiration $T_{c,t}$ and the irrigation supply $Q_{i,t}$, for the four different scenarios over the period 1993–1999.

where n_{t+1}^{irr} is the number of cells that can be irrigated in the time interval $[t, t+1)$ given the diverted flow q_{t+1}^{Muzza}, namely

$$n_{t+1}^{irr} = \frac{\eta \cdot \Delta}{V} \cdot q_{t+1}^{Muzza} \tag{12b}$$

and χ_{t+1} is the number of cells that enter (or leave) the deficit condition in the time interval $[t, t+1)$. Finally, ε_{t+1} is the process error.

The output transformation is given by Eq. (8), which is here reported

$$w_{t+1} = \frac{V}{\eta \cdot \Delta} \cdot n_{t+1} \tag{12c}$$

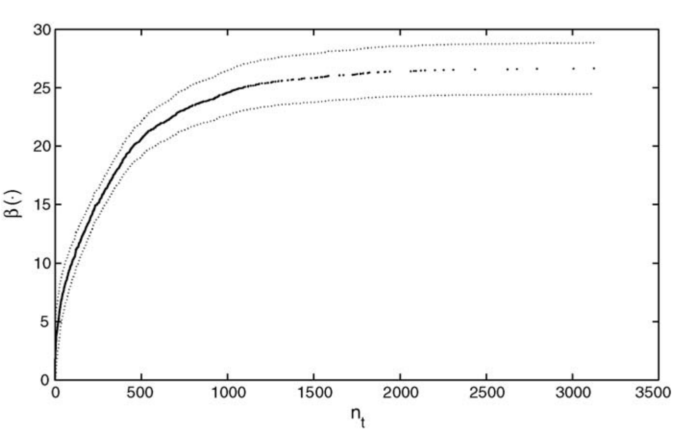

Fig. 7. Non-parametric estimate of the state-dependent parameter $\beta(\cdot)$ for the data-set of scenario 1. Dashed lines are standard error bounds ($\pm 4\sigma$).

To complete the definition of the metamodel form, we posed

$$\chi_{t+1} = a \cdot \beta(n_t) \cdot T_{c,t+1} \tag{13}$$

where $\beta(\cdot)$ is a state-dependent parameter, i.e. a parameter that, instead of being constant as in conventional linear models, is a function of the state variable n_t. By setting $a = 1$, the shape of $\beta(\cdot)$ was identified from the designed data with the procedure proposed by Young et al. (2002). For example, using the calibration data-set obtained from Scenario 1, the form reported in Fig. 7 was obtained. It can be interpreted as follows: when n_t is low, the number of cells with $U_{2,t} < U_{r,t}$ is likely to be low , too , and thus the future value of n_{t+1} will be strongly influenced by the number χ_{t+1} of cells that will enter (or leave) the deficit condition in the time interval $[t, t+1)$. However, this last number is unpredictable at time t, when only n_t is known. As a consequence, the dynamics of n_t shows a strong uncertainty and the relation between the input $T_{c,t+1}$ and the output n_{t+1} is weak, which implies a low value of $\beta(\cdot)$. Conversely, as n_t increases, its dynamics becomes more and more deterministic, and therefore the value of $\beta(\cdot)$ increases. The state-dependent parameters we obtained with the other scenarios present similar shapes and the physical interpretation here discussed is still valid.

In order to employ the metamodel outside the calibration data-set, the non-parametric estimate of $\beta(\cdot)$ was parameterized through the following logarithmic law

$$\beta(n_t) = \log(n_t + b^2) + c \tag{14}$$

and an initial estimate of b and c was obtained by interpolating the curve of Fig. 7.

On the basis of Eq. (14) model (12a), (12b), (13) was rewritten as

$$n_{t+1} = n_t - n_{t+1}^{irr} + a \cdot [\log(n_t + b^2) + c] \cdot T_{c,t+1} + \varepsilon_{t+1} \tag{15a}$$

Table 1
Coefficient of determination R^2 over the five calibration data-sets for the model (15). Remember that data-set 5 is composed of the aggregation of the first four.

Data-set	R^2
1	0.834
2	0.814
3	0.598
4	0.965
5	0.910

Table 2
Coefficient of determination R^2 [–] over the four validation data-sets for the models calibrated on the five data-sets.

Model	Validation data-set			
	R^2 (1)	R^2 (2)	R^2 (3)	R^2 (4)
1	0.887	0.870	0.683	0.977
2	0.887	0.870	0.678	0.977
3	0.879	0.863	0.698	0.969
4	0.882	0.863	0.645	0.980
5	0.887	0.870	0.672	0.979

and the estimates of the parameters a, b and c were refined through a global non-linear optimization procedure.

Finally the disturbance ε_{t+1} was modelled as a Gaussian, cyclostationary, stochastic process, i.e.

$$\varepsilon_{t+1} \sim N(\mu_{t_{modT}}, \sigma^2_{t_{modT}}) \tag{15b}$$

where $\mu_{t_{modT}}$ and $\sigma^2_{t_{modT}}$ are the periodic expected-value and variance, estimated from the model residuals.

Since the policy we are going to design will suggest the release from Lake Como on the basis of the storage s_t and the irrigation water demand W_t, the parameters of model (15) were calibrated by using it in one-step ahead prediction form, i.e. by estimating n_t at each step from the water demand W_t supplied by the distributed-parameter model through Eq. (9).

Since we have five calibration data-sets, we obtained five different parameterizations, and, in the following, we will use the term metamodel i to refer to the metamodel calibrated on the ith data-set.

The coefficient of determination[4] R^2 [–], obtained with the five metamodels, is reported in Table 1. The values of R^2 clearly show that the metamodel performs better on those data-sets which are associated to scenarios characterized by a low diverted flow (as data-set 4) than on the other ones (as data-set 3). There are two reasons for this behavior. First, the precision of the estimate n_t increases with increasing values of n_t (see Section 4.2), and large values of n_t occur when the diverted flow is low. Second, when this latter case occurs, the distance between the irrigation supply $Q_{i,t+1}$ and its estimate $\eta \cdot q^{Muzza}_{t+1}$ is small, and therefore q^{Muzza}_{t+1} is a more significant proxy of $Q_{i,t+1}$ (see Section 4.3).

4.6. Step A.6: first validation

Each of the calibrated metamodels is then tested on the four validation data-sets. From the performances reported in Table 2 it emerges that the five metamodels are characterized by similar performances: even the one calibrated on the fifth data-set does not show a higher performance than the other one. Thus for the design of the Lake Como release policy we can choose any of them.

[4] $R^2 = 1 - cov(W - w)/cov(W)$, where W is the trajectory of the irrigation demand simulated by the distributed-parameter model, while w is the trajectory forecasted by the metamodel.

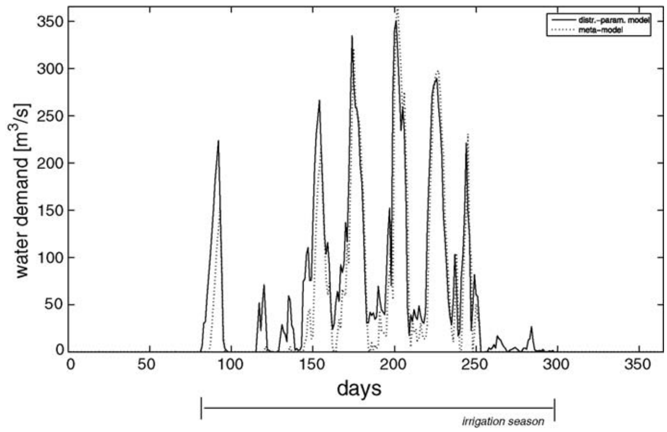

Fig. 8. Trajectories of the water demand computed with the distributed-parameter model (solid line) and metamodel 2 (dashed line) on the fourth year (2003) of the validation data-set 2.

Table 3
Coefficient of determination R^2 [–] over the calibration and validation data-set 2 for metamodel 2 with different numbers of regression terms.

Metamodel 2 with	Calibration data-set 2 R^2	Validation data-set 2 R^2
Eq. (13)	0.814	0.870
Eq. (16a)	0.814	0.870
Eq. (16b)	0.837	0.880

We adopted metamodel 2. A visual comparison of the irrigation water demand computed with the distributed-parameter model and metamodel 2 on the fourth year (2003) of the validation data-set 2 is shown in Fig. 8.

Before using this metamodel for the design of the reservoir release policy, we want to know if an increase of the number of regressors adopted to define χ_{t+1} might perceivably increase its predictive capabilities. Indeed, instead of Eq. (13), we might assume

$$\chi_{t+1} = a_1 \cdot \beta(n_t) \cdot T_{c,t+1} + a_2 \cdot \beta(n_{t-1}) \cdot T_{c,t} \tag{16a}$$

or

$$\chi_{t+1} = a_1 \cdot \beta(n_t) \cdot T_{c,t+1} + a_2 \cdot \beta(n_{t-1}) \cdot T_{c,t} + a_3 \cdot \beta(n_{t-2}) \cdot T_{c,t-1} \tag{16b}$$

or even a larger number of regressors. By passing from Eq. (13) to Eq. (16a) and Eq. (16b), the coefficient of determination R^2 increases only slightly (see Table 3), but, at the same time, the state dimension of the model increases from 1 to 3 and 5, and, with it, the computing time to design the release policy via SDP exponentially increases. By comparing the strong increase in the computing time with the nearly insignificant improvements of the performance measures, we definitively opted for Eq. (13) and thus for metamodel 2.

5. Second validation (Phase D)

As proposed in the design procedure (see Section 2), once the metamodel was successfully calibrated and validated on the I–O data-set generated in Step A.4, it was employed to design, via SDP, the Lake Como release policy (Phase B in Fig. 1; for further details see Galelli and Soncini-Sessa, 2009) with the software tool TwoLe (see Soncini-Sessa et al., 1999, 2003). The objective of this policy is to minimize the water shortages, namely the difference between

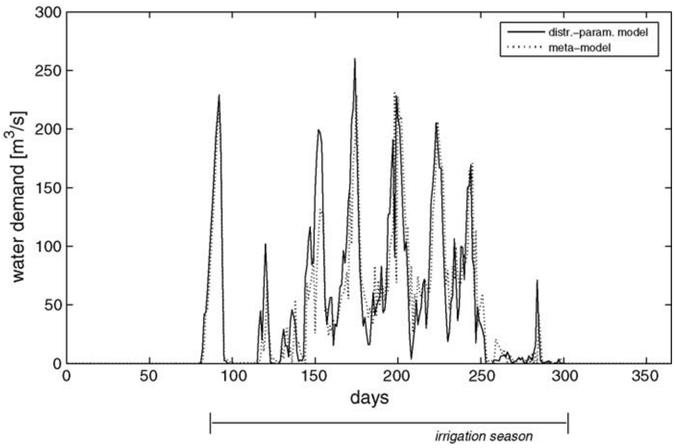

Fig. 9. Trajectories of the water demand computed with the distributed-parameter model (solid line) and metamodel 2 (dashed line) for the year 2003 (second validation).

the irrigation water demand and the diverted flow, when the former is larger than the latter. The water system controlled by the designed policy was then simulated (Phase C) in order to obtain the trajectories of the metamodel input and output variables over the validation period 2000–2004. In the simulation, the distributed-parameter model was employed to represent the irrigation district. On the basis of the trajectories so obtained, the metamodel was validated for the second time (Phase D). The coefficient of determination R^2 is equal to 0.807, a value comparable with the mean value (0.851) obtained in the first validation (see Table 2). This low decrease in the value of R^2 was expected. In fact, the designed lake regulation tends to minimize the number of cells in deficit condition N_t (i.e. the water shortages) and, as pointed out in Section 4.5, the precision of the estimate n_t of N_t supplied by the metamodel increases with increasing values of n_t. A visual comparison of the irrigation water demands computed with the distributed-parameter model and the metamodel (for the year 2003) is given in Fig. 9. Both evaluations prove that the metamodel is a satisfactory representation of the original distributed-parameter conceptual model.

At this point the main objective of this paper is achieved; however, before concluding, we would like to comment on the trajectories obtained in Phase C. A specimen of them is shown in Fig. 10(a) for the year 2003. Note that all through the irrigation season, q_t^{Muzza} follows the water demand, within the limit of the canal capacity (110 m^3/s). Since the total amount of available water is finite, the best allocation is obtained when no water losses occur, i.e. when the diverted flow is never greater than the water demand. This implies that the net diverted flow is never greater than the irrigation supply and that all the available water is actually applied to the fields at the end of the irrigation season, i.e. the reservoir is empty at that time. Figs. 10(a), (b) and 11(a) show that these conditions are met by the policy under test.

The reader might be astonished by the large difference between the water demand and the water supply shown in Fig. 10(a). To understand its value, consider that the year 2003 was extremely dry and remember that border irrigation is adopted in the district (when the soil moisture deficit in a cell overcomes the threshold $\alpha \cdot RAW_t^{(i)}$, see Eq. (5), its water demand suddenly jumps from 0 to 0.13 m^3/s). Under totally uniform conditions of soil, crop and meteorological characteristics, all the cells would reach the threshold at the same time, and the water demand would suddenly jump from 0 to 1100 m^3/s. In practice, due to the partially uniform conditions and to the turned irrigation, the water demand is not a sequence of instantaneous peaks, but rather a sequence of waves (the larger the inhomogeneities, the smoother the resulting demand pattern). However, this large shortage does not necessarily imply that crops incurred water stress. In fact, the irrigation threshold $\alpha \cdot RAW_t^{(i)}$ (with α lower than 1) is such that the irrigation demand is expressed before stress conditions are reached, and hence the crop can tolerate a short delay (a few days) of irrigation application. Indeed, with the value of α calibrated for the Muzza district, an almost perfect overlap of the transpiration T and the potential transpiration T_c occurred (see Fig. 11(b)). This confirms that the water allocation we obtained with the designed policy is optimal also with respect to the crops' condition.

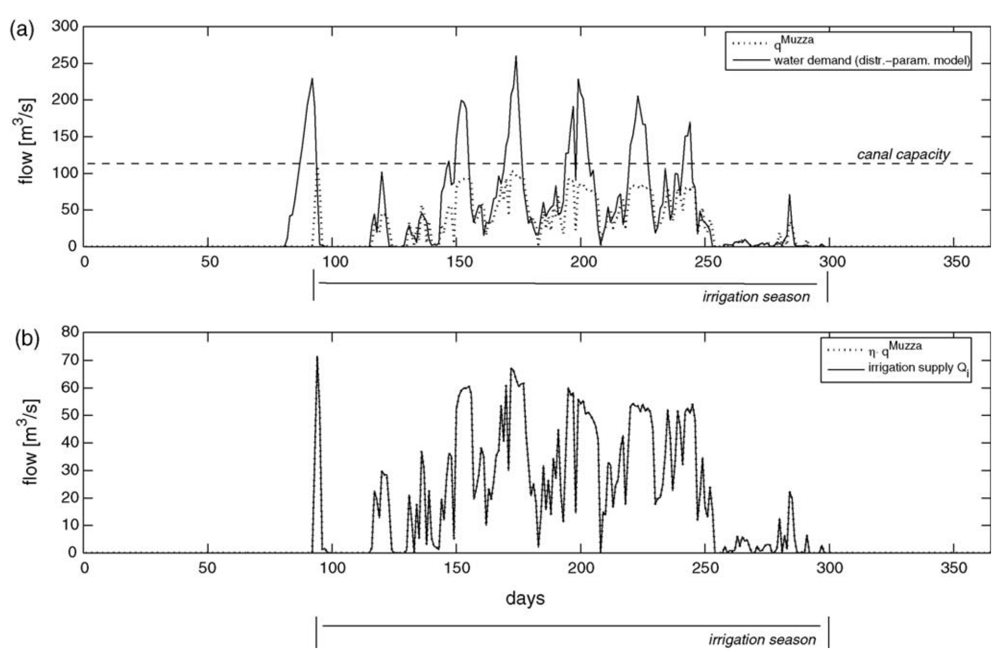

Fig. 10. Trajectories, in 2003, of the diverted flow q_t^{Muzza} (dashed line) vs the water demand W_t (solid line) (panel (a)); the net diverted flow $\eta \cdot q_t^{Muzza}$ (dashed line) vs the irrigation supply $Q_{i,t}$ (solid line) (panel (b)).

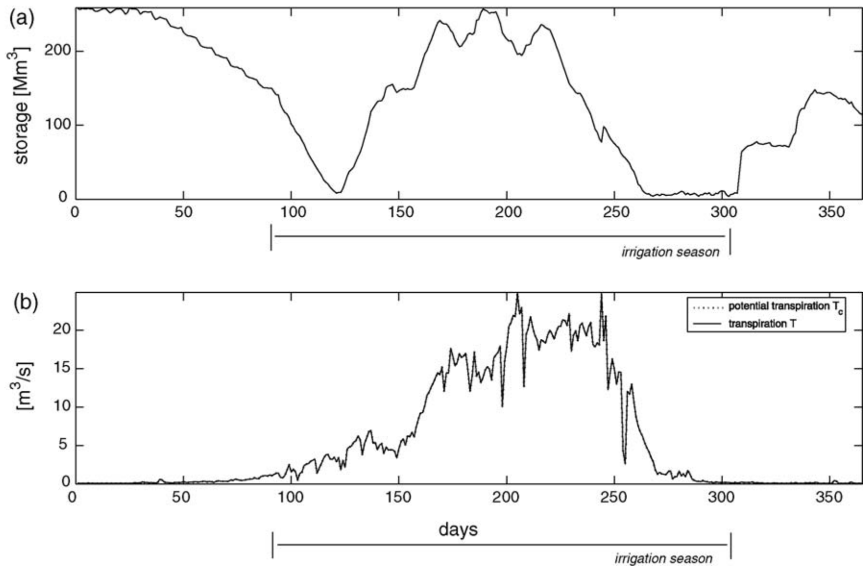

Fig. 11. Trajectories, in 2003, of the Lake Como storage (panel (a)); the potential transpiration T_c (dashed line) and the transpiration T (solid line) (panel (b)).

Finally, it is important to underline that, contrary to what happens when considering a non-dynamic target demand (i.e. an a priori given trajectory), the area between the dynamic water demand W_t and the diverted flow q_t^{Muzza} in Fig. 10(a) does not represent the total supply deficit. In fact W_t is a state variable and, as such, it keeps memory of its previous values. To clarify this concept, consider an n day period in which no irrigation takes place and no variation in the number of cells in deficit conditions occurs. The value W_1, which is present on day one, remains unchanged for the following n days, so that the total supply deficit at the end of the period is $86,400 \cdot n \cdot W_1$. The irrigation volume that is needed in day $n+1$ to satisfy the demand is $86,400 \cdot W_1$, and clearly not $86,400 \cdot n \cdot W_1$. If the same volume had been supplied in day 1, the total supply deficit in day n would have been null. It follows that a good indicator for the policy design is the total supply deficit, which is actually the objective we considered in Phase B.

6. Conclusions

In this paper we presented a procedure for identifying a metamodel of a distributed-parameter, conceptual model of a large irrigation district, to be used for planning or management purposes. The objective was to check the possibility of obtaining a dynamic representation of the irrigation district demand with the metamodel and then to design a reservoir release policy based on this latter. In fact, in this type of problem, the irrigation demand is generally approximated by a fixed static target demand (periodic over the year), owing to the severe limitations on the number of state variables that can be treated by optimization algorithms, in contrast to the large state dimension of distributed-parameter models. The proposed procedure was applied to the pilot study of Lake Como and the downstream Muzza irrigation district (northern Italy). As a first case of application, we needed to start from simplified, yet realistic, conditions for the irrigation district. In practice, we made the following two assumptions: (i) the flow diverted from the River Adda into the Muzza canal can be conveyed to any point in the district within a time period which is shorter than the time step of the release decision from Lake Como (one day); (ii) the actual supply of irrigation to the individual plots takes place on an on-demand basis, with priority given to the plots which expressed their demand first, in the case that the total demand exceeds the Muzza canal flow. Though clearly simplified, this district setup is not unrealistic; indeed, some recent papers (notably Mareels et al., 2005) have showed that the rehabilitation of open canals conveyance and distribution networks can be such that they can be regulated practically on-demand, with very small time delays in the network. The design of the metamodel was based on the simulated outputs of the distributed-parameter, conceptual model of the Muzza district, with a number of state variables of the order of 10^4. The metamodel has only one state variable and, despite its simple structure, lends itself to a physical interpretation of its state-dependent parameter and provides good performances in one-step-ahead prediction of the irrigation demand. Thanks to the minimum state dimensionality of the metamodel, it was possible to embed it within a stochastic dynamic programming algorithm, in order to design a release policy that includes, among its arguments, not only the lake storage, but also the estimate of the irrigation water demand. Thus, the so-called green water (the water stored in the root zone of crops; Falkenmark, 1991) is explicitly considered in the release decision, in addition to blue water (the lake storage). If this policy, together with the irrigation system rehabilitation, had been adopted in the year 1993–2004 for the management of Lake Como, an average saving of 75 Mm^3 per year would have been obtained, at the same time halving the irrigation deficit. 75 Mm^3 is a significant amount: it is one fourth of the operative capacity of Lake Como and this volume can be usefully employed to improve the environmental quality of the River Adda, the effluent from the Lake.

The results obtained encourage further research on the metamodelling of the irrigation demand on the district scale, in order to overcome some of the limitations of the pilot application here presented. Firstly, it must be recognized that the identification of the metamodel is based on some 'intuitive notions' of the irrigation system behavior and of its representation with the distributed-parameter model, which leaves some degree of subjectivity. The replication of the process is thus not straightforward and it is interesting to explore the possibility of automatizing the building of the metamodel by means of machine learning techniques (see, for instance, Das, 2001). Secondly, further research is required to formulate satisfactory guidelines for the Design of Experiments in step A.4, since the quality of the metamodel strongly depends on it.

Acknowledgement

The authors are grateful to Prof. P.C. Young, who provided the Captain Toolbox (http://www.es.lancs.ac.uk/cres/captain) used to develop the metamodel.

References

Ahrends, H., Mast, M., Rodgers, C., Kunstmann, H., 2008. Coupled hydrological–economic modelling for optimised irrigated cultivation in a semi-arid catchment of west Africa. Environmental Modelling and Software 23, 385–395.
Allen, R., Pereira, L., Raes, D., Smith, M., 1998. FAO irrigation and drainage paper 56. Crop evapotranspiration. Guidelines for computing crop water requirements, Food and Agriculture Organization, Rome, I.
Arnold, J., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modeling. Hydrologic Processes 19, 563–572.
Baldoni, R., Giardini, L., 1982. Herbaceous crops. Patron (Ed.), Bologna, I (in Italian).
Baroni, G., Facchi, A., Gandolfi, C., Ortuani, B., Horeschi, D., van Dam, J., 2009. Uncertainty in the determination of soil hydraulic parameters and its influence on the performance of two hydrological models of different complexity. Hydrology and Earth System Sciences Discussions 6 (3), 4065–4105.
Bergström, S., 1995. The HBV model. Computer Models of Watershed Hydrology. Water Resources Publications, V.P. Singh, Highlands Ranch, CO.
Blanning, R., 1975. The construction and implementation of metamodels. Simulation 24 (6), 177–184.
Børgesen, C., Djurhuus, J., Kyllingsbaek, A., 2001. Estimating the effect of legislation and nitrogen leaching by upscaling field simulations. Ecological Modelling 136, 31–48.
Bouraoui, F., Vachaud, G., Haverkamp, R., Normand, B., 1997. A distributed physical approach for surface-subsurface water transport modeling in agricultural watersheds. Journal of Hydrology 203 (1–4), 79–92.
Bouzaher, A., Lakshminarayan, P., Cabe, R., Carriquiry, A., Gassman, P., Shogren, J., 1993. Metamodels and nonpoint pollution policy in agriculture. Water Resources Research 29, 1579–1587.
Braden, H., 1985. Ein energiehaushalts und verdunstungsmodell für wasser und stoffhaushaltsuntersuchungen landwirtschaftlich genutzer einzugsgebiete. Mitteilungen deutsche bodenkundliche gesselschaft 422, 294–299.
Castelletti, A., Pianosi, F., Soncini-Sessa, R., 2008a. Integration, participation and optimal control in water resources planning and management. Applied Mathematics and Computations 206 (1), 21–33.
Castelletti, A., Pianosi, F., Soncini-Sessa, R., 2008b. Water reservoir control under economic, social and environmental constraints. Automatica 44 (6), 1595–1607.
Das, S., June 28 to July 1, 2001. Filters, wrappers and a boosting-based hybrid for feature selection. In: Proceedings of the Eighteenth International Conference on Machine Learning, Williamstown, MA.
Donatelli, M., 1995. Systems for integrated crop management. Special publication of the istituto sperimentale agronomico, ISA - Sezione di Modena, Modena, I (in Italian).
Facchi, A., Ortuani, B., Maggi, D., Gandolfi, C., 2004. Coupled SVAT-groundwater model for water resources simulation in irrigated alluvial plains. Environmental Modeling and Software 19 (11), 1053–1063.
Falkenmark, M., 1991. Environment and development: Urgent need for a water perspective. Water International 16, 229–240.
Galelli, S., Soncini-Sessa, R., 2009. Combining metamodelling and stochastic dynamic programming for the design of reservoir release policies. Environmental Modelling and Software, doi:10.1016/j.envsoft.2009.08.001.
Gandolfi, C., Facchi, A., Maggi, D., 2006. Comparison of 1d models of water floe in unsaturated soils. Environmental Modelling and Software 21, 1759–1764.
Ghahraman, B., Sepaskhah, A., 2002. Optimal allocation of water from a single water reservoir to an irrigation project with pre-determined multiple cropping patterns. Irrigation Science 21, 127–137.
Gijsman, A., Jagtap, S., Jones, J., 2003. Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. European Journal of Agronomy 18, 77–106.
Haberlandt, U., Krysanova, V., Bárdossi, A., 2002. Assessment of nitrogen leaching from arable land in large river basins. Part 2: regionalisation using fuzzy rule based modelling. Ecological Modelling 150, 277–294.
Huygen, J., van Dam, J., Kroes, J., Wesseling, J., 1997. SWAP 2.0: input and output manual. Technical document, WAU and DLO - Staring Centrum, Wageningen, NL.
Jain, A., Murty, M., Flynn, P., 1999. Data clustering: a review. ACM Computing Surveys 31 (3), 264–323.
Kijne, W., 2003. Unlocking the Water Potential of Agriculture. Food and Agriculture Organization, Rome, I.
Kleijnen, J., Sanchez, S., Lucas, T., Cioppa, T., 2005. A user's guide to the brave new world of designing simulation experiments. INFORMS Journal of Computing 17 (3), 263–289.
Kleijnen, J., Sargent, R., 2000. A methodology for fitting and validating metamodels in simulation. European Journal of the Operational Research 120 (1), 14–29.
Mareels, I., Weyer, E., Ooi, S.K., Cantoni, M., Li, Y., Nair, G., 2005. Systems engineering for irrigation systems: successes and challenges. Annual Reviews in Control 29 (2), 191–204.
Neitsch, S., Arnold, J., Kiniry, J., Srinivasan, R., Williams, J., 2002. Soil and water assessment tool user's manual version 2000. GSWRL Report 02-02, BRC report 02–06, Texas Water Resources Institute, College Station, TX.
Niswonger, R., Prudic, D., Regan, R., 2006. Documentation of the unsaturated-zone flow (UZF1) package for modeling unsaturated flow between the land surface and the water table with MODFLOW-2005. Tech. Meth. 6-A19, USGS, Reston, VA.
PRA, 1998. Study and research activities in support of the river basin plan - sub project 4.1: Land use and agriculture. Technical document, Po River Authority, Parma, I (in Italian).
Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, P., 2005. Surrogate-based analysis and optimization. Progress in Aerospace Sciences 41, 1–28.
Rawls, W., Brakensiek, D., 1989. Estimation of soil water retention and hydraulic properties. In: Morel-Seytoux, D. (Ed.), Unsaturated Flow in Hydrological Modeling, Theory and Practice. Kluwer, Dordrecht, NL, pp. 275–300.
Refsgaard, J., Storm, B., 1995. MIKE SHE. Computer Models of Watershed Hydrology. Water Resources Publications, V.P. Singh, Highlands Ranch, CO.
Schiermeier, Q., 2008. Water: a long dry summer. Nature 452, 270–273.
Shangguan, Z., Shao, M., Horton, R., Lei, T., Qin, L., Ma, J., 2002. A model for regional optimal allocation of irrigation water resources under deficit irrigation and its applications. Agricultural Water Management 52 (2), 139–154.
Simpson, T., Peplinski, J., Koch, P., Allen, J., 2001. Metamodels for computer based engineering design: survey and recommendations. Engineering with Computers 17, 129–150.
Singh, R., Subramanian, K., Refsgaard, J., 1999. Hydrological modelling of a small watershed using MIKE SHE for irrigation planning. Agricultural Water Management 41, 149–166.
Soncini-Sessa, R., Castelletti, A., Weber, E., 2003. A DSS for planning and managing water reservoir systems. Environmental Modelling and Software 18 (5), 395–404.
Soncini-Sessa, R., Castelletti, A., Weber, E., 2007. Integrated and Participatory Water Resources Management. Theory. Elsevier, Amsterdam, NL.
Soncini-Sessa, R., Rizzoli, A., Villa, L., Weber, E., 1999. TwoLe: a software tool for planning and management of water reservoir networks. Hydrological Science Journal 44 (4), 619–631.
Steduto, P., Hsiao, T., Raes, D., Fereres, E., 2009. Aquacrop-the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agronomy Journal 101, 426–437.
Stockle, C., Nelson, R., 1996. Cropsyst User's Manual - Version 2.0. Technical Document, Washington State University. Pullman, WA. Departement of Biological Systems Engineering.
Sudicky, E., Park, Y., Unger, A., Jones, J., Brookfield, A., Colautti, D., Therrien, R., Graf, T., October 22–25, 2006. Simulating complex flow and contaminant transport dynamics in an integrated surface-subsurface modelling framework. In: Geological Society of America Abstracts with Programs. Philadelphia, USA.
Ungaro, F., Calzolari, C., 2001. Using existing soil databases for estimating retention properties for soils of the pianura padano-veneta region of north italy. Geoderma 99, 99–121.
USDA-SCS, 1972. National engineering handbook. Part 630 hydrology, section 4, U.S. Department of Agricolture—Soil Conservation Service, U.S. Government Printing Office, Washington, DC, USA.
USDA-SCS, 1986. Hydrology for small watersheds. Technical release no. 55, U.S. Department of Agricolture—Soil Conservation Service, U.S. Government Printing Office, Washington, DC, USA.
van Dam, J., Groenendijk, P., Hendriks, R., Kroes, J., 2008. Advances of modeling water flow in variably saturated soils with SWAP. Vadoze Zone Journal 7, 640–653.
van Walsum, P., Groenendijk, P., 2008. Quasi steady-state simulation of the unsaturated zone in groundwater modeling of lowland regions. Vadoze Zone Journal 7 (2), 769–781.
Vedula, S., Kumar, D.N., 1996. An integrated model for optimal reservoir operation for irrigation of multiple crops. Water Resources Research 32 (4), 1101–1108.
Vedula, S., Mujumdar, P., 1992. Optimal reservoir operation for irrigation of multiple crops. Water Resources Research 28 (1), 1–9.
Vörösmarty, C., Green, P., Salisbury, J., Lammers, R., 2000. Global water resources: vulnerability from climate change and population growth. Science 289, 284–288.
Voss, C., Provost, A., 2002. SUTRA, a model for saturated-unsaturated, variable-density ground-water flow with solute or energy transport. Water-resources Investigations Report 02–4231, USGS, Reston, VA.
Williams, J., Jones, C., Dyke, P., 1984. Modeling approach to determining the relationship between erosion and soil productivity. Transactions of the American Society of Agricultural Engineers 27 (1), 129–144.
Wösten, J., Pachepsky, Y., Rawls, W., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology 251, 123–150.
Young, P., 1998. Data-based mechanistich modelling of environmental, ecological, economic and engineering systems. Environmental Modeling and Software 13, 105–122.
Young, P., McKenna, P., Bruun, J., 2002. Identification of nonlinear stochastic systems by state dependent parameter estimation. International Journal of Control 74, 1837–1857.
Young, R., Onstad, C., Bosch, D., Anderson, W., 1989. AGNPS: a non-point source pollution model for evaluating agricultural watersheds. Journal of Soil and Water Conservation 44, 168–173.