Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 927 – 936

Procedia Engineering

www.elsevier.com/locate/procedia

13th Computer Control for Water Industry Conference, CCWI 2015

Optimal sensors placement for flood forecasting modelling

Grazia Fattoruso[a]*, Annalisa Agresta[b], Guido Guarnieri[a], Bruno Lanza[a], Antonio Buonanno[a], Mario Molinara[c], Claudio Marrocco[c], Saverio De Vito[a], Francesco Tortorella[c] and Girolamo Di Francia[a]

[a]*UTTP/Basic Materials and Devices Lab., ENEA Research Center Portici, P.le E. Fermi, 1 80055 Portici Naples, (Italy)*
[b]*Civil, Architectural and Environmental Engineering Dept., University of Naples,Via Claudio, 21 Naples (Italy)*
[c]*Electromagnetism, Information Engineering and Mathematics Dept., University of Cassino and Southern Lazio, Via G. di Biasio, 43 Cassino, (Italy)*

Abstract

Numerical models are instrumental to more effective flood forecasting and management services though they suffer from numerous uncertainty sources. An effective model calibration is hence essential. In this research work, a methodology of optimal sampling design has been investigated and developed for water drainage networks. Optimal hydrometer sensors locations along the Amato River (South Italy) have been defined by optimizing a two-objective function that maximizes the calibrated model accuracy and minimizes the total metering cost. This problem has been solved by using an enumerative search solution, run on the ENEA/CRESCO HPC infrastructure, evaluating the exact Pareto-front by efficient computational time.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Sampling design, model calibration, hydro-rain gauges, hydrological modelling, flooding forecasting

1. Introduction

Robust forecasts are vital in providing comprehensive flood warning and management services to communities at

* Corresponding author. Tel.: +39-081-7723266; fax:+39-081-7723344.
E-mail address: grazia.fattoruso@enea.it

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.971

risk from flooding.

For flood forecasting, rainfall–runoff, flow routing and hydraulic models are often combined into model cascades and are run automatically in early warning and safety management systems. Hydro-meteorological data are also gathered in real time by distributed sensor networks. Generally, rain-gauges and water level stations networks distributed along the drainage systems and into surrounding area are linked to control centres by telemetry for allowing operators to monitor the situation, giving warnings against indicator or trigger rainfall and water levels, as well as providing inputs into forecasting models, particularly rainfall–runoff and hydraulic models.

However, it is widely recognized that the accuracy of flood forecasts can be influenced by a number of factors, such as the accuracy of input data, and the model structure, parameters and state (initial conditions). The numerous modelling uncertainty sources become a key concern for modeling practical viability and, indeed, the need of efficient calibration processes are commonly highlighted by researchers and practitioners in the field [1, 2].

A calibration process modifies the models uncertain parameters (e.g. pipe roughness, rate of evaporation, rate of permeability) until model predictions match closely with measured values on the real system, under a range of operating conditions within acceptable error bounds. Thus, a calibration process needs a collection of data measured in various points of the real system. In particular, for the selected model variables to be observed, it is necessary to define where in the system observe them, when to observe them (duration and frequency) and what conditions should exist during the observation.

The efficacy of a calibration procedure is directly dependent on the sampling scheme whose performance depends on both the quantity and the spatial distribution of measurements/devices to be optimally located. Both have been shown to affect the resulting model calibration accuracy [1].

Generally, due to the extension of most water drainage systems and inaccessibility of some points, it is practically impossible to install sensors or to take measurements at all of the candidate locations. In order to obtain information for an estimate of the model parameters, data must be collected from a subset of a finite domain of predefined *suitable* locations, subset designed for maximising the performance with respect to specific design criteria. This approach so-called in literature *optimal sampling design*, in the context of water sensor networks, refers to an ideal placement of a limited number of sensor devices along a water drainage network.

In more recent literature, the optimisation of data sampling locations, especially within the context of water distribution systems, has been formulated as a multi-objective optimization problem [3, 4]. The main objectives to be achieved were to maximize the estimated accuracy and minimize the total metering cost. For quantifying and defining measures of effectiveness suitable to assess any given sensor layout on a water distribution network, most of applied approaches use the uncertainty analysis via the First-Order Second-Moment (FOSM) method (e.g. [5, 6, 7, 8, 9, 10, 11]).

In the present research work, a methodology of optimal sampling design has been investigated and developed for water drainage systems aimed at modelling more robust flood forecasts. The optimal sensor locations problem along a drainage network has been formulated as a two-objective optimization problem intended to assess the Pareto front for trade-offs between sensor cost and the resulting model accuracy. This problem has been solved by using an enumerative search algorithm, run on the ENEA/CRESCO HPC (High Performance Computing) infrastructure, evaluating the exact Pareto-front by an efficient computational time.

2. The methodology

2.1. Problem formulation

The problem of the optimal sampling design investigated within this research work is closely related to the problem of model calibration for water drainage networks. More specifically, the objective of the sampling design is here to find a set of optimal water level sensors locations along a drainage network with the aim of calibrating accurately the rainfall–runoff models, generally combined into models cascades for performing flood forecasting scenarios.

A model calibration process aims to determine the selected uncertain model parameters that, when input into a model simulation, produce a reasonable match between predicted and measured value along the real drainage system [5]. Formally, this problem can be defined as follows:

$$Min\ z = y - y^*, \quad y = f(x,\theta)\ \ and\ y^* = f(x,\theta) + \varepsilon \tag{1}$$

Here z = scalar objective function value to be minimized; y = model output values and y^* = measured values; x = input variables and θ = uncertain parameters (e.g. *Curve Number*-CN parameter) and ε = error term. In a calibration process, the selected parameter vector θ must minimize the error term ε.

Taking into account the approaches proposed in literature for water distribution systems [10, 12, 13], the investigated sampling design problem has been formulated and solved as a two-objective optimization problem under parameter uncertainty. The two objectives are the maximization of the calibrated rainfall-runoff model accuracy and the minimization of the sampling design cost.

To quantify the model prediction accuracy, a FOSM (First Order Second Moment) model is used to approximate the relevant parameter and prediction covariance matrices. For our scopes, we have assumed that prediction and measurement variables of interest were the outlet flow rates only. So, if a set of N_l water level sensors with the standard deviation *s* are installed in N_l sub-catchment outlets, the variance of calibrated parameters can be estimated from the diagonal elements of the parameter covariance matrix:

$$Cov_a = \sigma^2 \cdot (J^T \cdot J)^{-1} \tag{2}$$

where σ^2 = error variance in measured quantities; J = Jacobian Matrix of derivatives $dy_i/d\theta_j$ (i=1,….,N_0; j=1,….,N_a); N_0 = number of measurement data in both spatial and temporal domains (e.g. if there are N_t time steps for each of N_l monitoring locations, then $N_0 = N_t \times N_l$), N_a = number of calibration parameters.

The value of the *i*th diagonal element of matrix Cov_a defines the uncertainty of the *i*th calibration parameter.

By evaluating Cov_a can be calculated the model prediction variance-covariance matrix Cov_z (the prediction covariance matrix) as follows [14,15]:

$$Cov_z = J_z \cdot Cov_a \cdot J_z^T \tag{3}$$

were J_z= Jacobian matrix of derivatives $dz_i/d\theta_j$ (i = 1,…., N_{lz}; j = 1,… N_{la}); N_z = number of predicted variables of interest. As in the case of parameter uncertainties, the uncertainty of the *i*th model prediction is estimated as the value of the *i*th diagonal element of matrix Cov_z.

Several methods exist for the calculation of elements of the Jacobian matrices J and J_z [16]. Here, we calculate the elements of the Jacobian matrices through the *finite-difference method*, i.e.:

$$\frac{\partial z_i}{\partial a_k} = \frac{z_i^{\Delta} - z_i}{(a_k + \Delta a) - a_k} \tag{4}$$

where Δa = variation of uncertain parameter a; z_i^{Δ} predicted value in correspondence to $a_k + \Delta a$ value.

Thus, the elements of matrices are evaluated through a step by step procedure: (1) simulate hydrological model using a_k value for k=1; (2) simulate hydrological model using $a_k + \Delta a$ value for k=1; (3) calculate derivatives through the Eq.(4); (4) repeat step (1) for k=2,…, N_0.

So, the first objective intending to maximize the hydrological model calibration accuracy is obtained with the estimation of the prediction uncertainties. It is hence calculated as [12]:

$$Min\ F_1 = \frac{1}{N_z}\sum_{i=1}^{N_z} Cov_{z,ii}^{1/2} \tag{5}$$

Each sampling plan corresponds a different Jacobian matrices J and consequently to different matrices $\mathrm{Cov_a}$, $\mathrm{Cov_z}$ and hence to a different objective function F_1.

The second objective addresses the problem of total metering costs by using a surrogate measure [4, 14], due to the difficulty for defining the capital and operation sensors costs.

The second objective value is hence calculated as:

$$Min\ F_2 = N\ ,\ \ N_{min} \le N \le N_{max} \tag{6}$$

where N= number of measurement devices (i.e. water level sensors); N_{min} = minimum required number of measurement devices; N_{max}=maximum allowed number of measurement devices.

Optimizing the sampling design problem as formulated in the equations (5)–(6) is complex. The problem solution is essentially to identify the following three elements:

$$(n, \phi_n, L_n) \tag{7}$$

where n=1,….., n_{max} is the number of sensors to be located; ϕ_n is the uncertainty related to n sensors and L_n is the sensor locations coordinates vector.

2.2. Enumerative search solution

The problem of optimal sensors locations, formulated by (7) is a complex two-objective optimization problem where the two objectives to be pursued are conflicting and they cannot be adequately combined into a single objective function. In fact, the defined objectives - to minimize the metering costs (i.e. number of sensors to be installed) and to maximize the model calibration accuracy, are directly conflicting being the accuracy a monotonically increasing function of the number of available sensors. In such a problem, therefore, there is not a unique solution that optimizes both the objectives, and thus the optimization procedure aims at finding solutions belonging to the Pareto front [15], i.e., to the set of "Pareto-optimal" solutions.

A solution is Pareto-optimal if there isn't another solution that is superior to one of the two objectives without being inferior to the other one. Focusing on the Pareto front, a tradeoff has to be made between the two objectives instead of considering the full range of every parameter. The optimization procedure must necessarily identify the entire set of solutions so that a decision can be taken according to the specific requirements of the faced problem [16].

In our case, the Pareto front consists of pairs of values (N, υ) which report the best accuracy υ to be achieved by a certain number N of sensors. These pairs form the set of optimal solutions, and a choice can be done on the basis of additional considerations which may be related to the available budget, the possibility of sensor maintenance, the level of accuracy required for the models, and so on.

The identification of the Pareto front can be computationally complex requiring a high computational time, since the evaluation of all possible solutions and the research of the Pareto-optimal ones may lead to an exponential search space (or infinite in the case of continuous objective functions). For this reason, in the literature, several strategies have been proposed based on algorithms that do not guarantee to obtain the exact set of Pareto-optimal solutions, but tend to make more effective the research in the feasible region and to provide appropriate solutions. In other words, these algorithms obtain an adequate approximation of the Pareto front at an acceptable computational cost. Within these methods, an important role is played by the Evolutionary Algorithms (EA) that constitutes the basis of widely-used approaches in sampling design such as MOGA (Multiple Objective Genetic Algorithm) [17] and NSGA -II (Non-dominated Sorting Genetic Algorithm II), [18].

For these reasons, due to the discrete nature of the minimization sensors number problem and the limited range of both the objectives (sensors number and model accuracy value), a viable alternative to the genetic algorithms can be the enumerative search. This is a conceptually simple search strategy to solve multi-objective optimization problems based on evaluating each possible solution in a finite search space that has been properly discretized [19]. A drawback of this approach is its inadequacy to scale when the search space becomes larger.

However, the results obtained through the enumeration are of great interest, because at present it is the only technique that computes the exact Pareto front in a multi-objective problem. Moreover, to overcome the problem of large search space, we can refer to grid computing systems that offer a potentially large amount of computing power. In fact, to estimate the 2^{14} = 16384 possible combinations of our problem for the placement of 14 sensors (as explained in the following section) we have employed the HPC infrastructure - CRESCO, installed at the ENEA

Research Center of Portici (see Appendix A) that has allowed evaluating the exact Pareto front by very efficient computational time.

3. Case study

3.1. Problem description

The investigated optimal sampling design methodology aimed at calibrating accurately the rainfall–runoff models, has been applied to the real drainage network of Amato, a drainage basin located in Calabria (South Italy).

This drainage network is a pilot area within a national R&D project named AQUASYSTEM, proposed within *smart cities and communities and social innovation* Programme Framework. AQUASYSTEM Project aims to investigate and develop innovative technologies and procedures for improving flood forecasting, warning and managing services.

For the case study, the following steps have been carried out: (1) identification of the types of sensors to be installed along the Amato drainage network (among water level sensors, rain gauges etc.); (2) building of the network model; (3) selection of the model uncertain parameters (among roughness canal, rate of permeability, etc.); (4) definition of the variation range of uncertain parameters; (5) implementation of optimization procedure for sampling design.

For the hydrological model calibration, water level and rainfall measurements are needed. As the Amato drainage basin is just equipped by a rain gauges, water level sensors were to be optimally located along the river network.

The configuration of the Amato basin model is given in Fig1. The flow data are used for the rainfall-runoff model calibration and the *Curve Number* (CN) [20] is set as uncertain parameter.

The CN is an empirical parameter used in hydrology for predicting direct runoff or infiltration from rainfall excess; each sub-basin has associated a specified CN value. The major factors that determine CN values are the hydrologic soil group, cover type and hydrologic conditions.

Fig. 1 Amato water drainage network model layout and its indexed sub-catchment areas, built in SWMMM5 software

The Amato drainage network model, building by using EPA-SWMM software [21], counts 27 sub-catchments, indexed as in Fig. 1. The runoff component of SWMM software performs a collection of sub-catchment areas on which rain falls and runoff is generated.

For each sub-catchment area, CN variation ranges have been specified taking into account their soil heterogeneity (Tab. 1).

Table 1. CN Variation Range

ID $_{SWMM5}$	CN interval		ID $_{SWMM}$	CN interval	
	Minimum	Maximum		Minimum	Maximum
40	25	83	64	45	98
41	25	83	67	45	98
42	62	91	68	62	91
46	70	91	69	30	91
47	70	91	71	30	91
49	45	83	72	30	91
55	25	77	73	50	81
57	45	83	74	50	81
58	45	83	75	50	81
59	45	83	76	25	90
60	30	83	77	30	91
61	70	91	78	30	91
62	30	83	79	30	91
63	30	83			

As potential water level sensor locations, a subset of 14 sub-catchment outlet points have been selected on the basis of practical considerations, by the outlet points shown in Fig. 2 (i.e. $ID_{outlet} = 26, 23, 21, 17, 16, 15, 13, 12, 11, 8, 5, 4, 3, 2$).

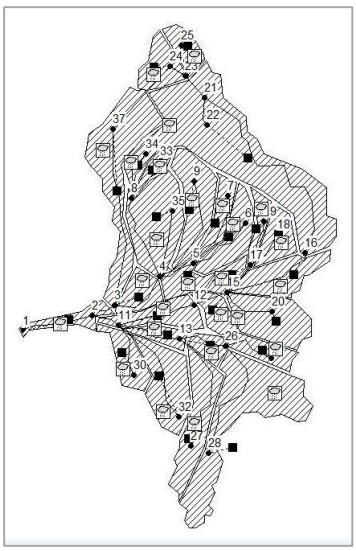

Fig. 2 Indexed sub-catchments outlet points of the Amato drainage network model built in SWMM5 software

3.2. Method application

The developed method allows us to evaluate the flow rate values into the 14 selected outlet points, starting by a set of assigned CN values and available rain data series associated to each delineated sub-catchment. In particular,

rain data to be run by the SWMM5 rainfall-rainoff component have been selected at1-hour time step intervals on 3-days.

In order to apply the FOSM method (see (2)-(6) equations), it needs to specify the following measures:

- collected measures: y_i $(i = 1, \dots, N_o)$ where $N_O = N_s \cdot N_t$ are the measures on N_s locations in N_t time intervals;
- predicted measures: z_i $(i = 1, \dots, N_Z)$ performed by SWMM5 core engine into specified locations;
- calibration parameters: a_k with $k = 1, \dots, N_a$ i.e. the model uncertain parameters

According to the problem description, the calibration parameters a_k are the *CN* values for $N_a = 27$ defined sub-catchments. The total number of the real and predicted flow measures are $N_s + N_z = 14$, i.e. the selected sub-basin outlet points.

For example, if hydrometers to be installed are $N_s = 3$, then the model prediction values are $N_z = 11$. Thus, assuming that $N_t = 71$, given by 1-hour flow rate data for 3 days, the collected measures are $N_O = 3 \cdot 71 = 213$.

In the first step, a set of values within the variability ranges (see Tab. 1) are selected for the calibration parameters $a_k = a_k^0$, $k = 1, \dots, 27$. These values are input to SWMM5 software in order to evaluate 71 different flow rate values for each potential sensor location. The measures corresponding to N_s locations with y_i $i = 1, \dots, 213$ and the measures corresponding to N_Z predicted locations with z_i $i = 1, \dots, 11$ are selected by the 71 values. Note that for each location, the average of the 71-hours data is considered.

In the second step, varying the calibration parameters, the calculations are repeated, obtaining a new set of $a_k = a_k^0 + \Delta a$, $k = 1, \dots, 27$. By the calculated values $y_i^{\Delta}, i = 1, \dots, 213$ e $z_i^{\Delta}, i = 1, \dots, 11$ through the SWMM5 software, it is possible to evaluate the covariance matrices Cov_a and Cov_z, depending by J that is the Jacobian matrix of the derivatives (see equation (2) and (3)).

The evaluated uncertainty values have been normalized respect to the ideal configuration Ψ where $N_S = 14$ and $N_Z = 0$. The aforesaid two steps has been run 100 times, on a set of parameters $a_k = a_k^j$ $k = 1, \dots, 27$ and $j = 1, \dots, 100$, extracted through the *latin hypercube* [22, 23], obtaining a set of φ_j, $j = 1, \dots, 100$. The calibration accuracy value has been calculated by the average of φ_j:

$$\bar{\varphi} = \frac{1}{100} \sum_{r=1}^{100} \varphi_j \tag{8}$$

where $\bar{\varphi}$ is the estimation of the calibration accuracy for the N_S selected locations.

Thus, the described procedure has been applied for evaluating the Pareto front. The algorithm steps are basically:

1. for each value $N_S = 1, \dots, 14$, consider $\binom{N_S}{14} = \frac{14!}{(14-N_S)! \cdot N_S!}$ possible locations schemes
2. for each $\binom{N_S}{14}$ possible location scheme, named L_{Ns}, evaluate the related calibration accuracy
3. evaluate $\bar{\varphi}_{N_s}^*$ = max $\bar{\varphi}_{Ns}$

The algorithm output is a table of elements $(N_S, \bar{\varphi}_{N_s}^*, L_{N_S}^*)$, where each $L_{N_S}^*$ locations scheme represents the optimal solution with $\bar{\varphi}_{N_s}^*$ associated calibration accuracy value. Note that the optimal locations schemes $L_{N_S}^*$ are shown as sequence of 1 and 0 where 1 means that the outlet point should be monitored; 0 means not monitored.

By the Pareto front graph that plots the calculated $\bar{\varphi}_{N_s}^*$ values vs the $N_S = 1, \dots, 14$ values, it is possible to visualize the relationship among calibration accuracy values and the number of hydrometers to be installed.

4. Results and discussion

In Fig. 3 is showed the obtained Pareto optimal Front obtained by the proposed optimal sampling design method applied to the study case of Amato basin. In this graph, on the x-axis are plotted the number of N_S hydrometers to be installed, and on the y-axis the $\bar{\varphi}_{N_s}^*$ values i.e. the minimum parameter uncertainty values corresponding to the optimal locations schemes obtained on N_S sensors.

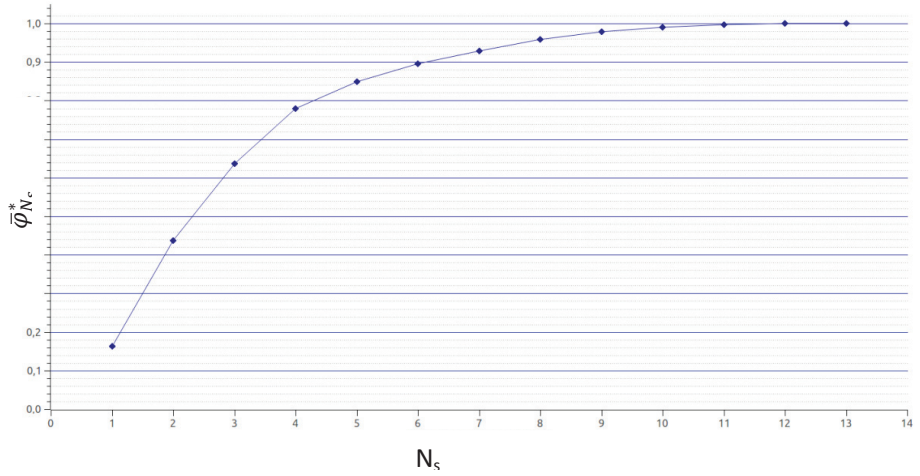

Fig. 3 Pareto optimal front obtained by the proposed optimal sampling design method based on the enumerative search

By analyzing the graph of $\bar{\varphi}^*_{Ns}$ vs N_s in Fig. 3, it is possible to observe that the slope of the curve corresponding to the first four N_s values is greater than the slope corresponding to the successive N_s values. This means that the model calibration accuracy, by installing $N_s > 4$ water level sensors along the Amato river network, does not increases significantly though the number of monitoring points is greatest.

In the Tab. 2, for each Pareto optimal solution (N_s $\bar{\Phi}^*_{N_s}$), the related optimal hydrometers locations scheme $L^*_{N_S}$, as sequence of 0,1 where 0 means sub-catchment outlet point to be not monitored and 1 to be monitored, is specified. So the operator knows exactly where the N_S hydrometers have to be installed along the Amato river network and the level (%) of accuracy of the simulated rainfall-runoff process.

Table 1. Pareto optimal solutions.

N_S	$\bar{\Phi}^*_{N_s}$	$L^*_{N_S}$ – id outlets													
		26	**23**	**21**	**17**	**16**	**15**	**13**	**12**	**11**	**8**	**5**	**4**	**3**	**2**
1	0,1629	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2	0,4377	1	0	0	0	0	0	0	0	0	0	0	0	1	0
3	0,6363	1	0	0	0	0	0	0	0	0	1	0	0	1	0
4	0,7796	1	0	0	0	0	1	0	0	0	1	0	0	1	0
5	0,8484	1	0	1	0	0	1	0	0	0	1	0	0	1	0
6	0,8961	1	0	1	0	0	1	0	0	0	1	0	1	1	0
7	0,9294	1	1	1	0	0	1	0	0	0	1	0	1	1	0
8	0,9587	1	1	1	0	0	1	0	0	1	1	0	1	1	0
9	0,9783	1	1	1	1	0	1	0	0	1	1	0	1	1	0
10	0,9898	1	1	1	1	0	1	0	0	1	1	1	1	1	0
11	0,9962	1	1	1	1	1	1	0	0	1	1	1	1	1	0
12	0,9996	1	1	1	1	1	1	1	0	1	1	1	1	1	0
13	0,9999	1	1	1	1	1	1	1	1	1	1	1	1	1	0
14	1,0000	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Analyzing the sequences shown in Tab.2, we can notice that the optimal solutions, when N_s increases, are conservative. In fact, the 1-optimal solution includes the outlet point indexed 3. The 2-optimal solution adds the only location indexed 26; the 3-optimal solution includes the sub-catchment outlets to be monitored {3, 26, 11} and so

on. In the Fig. 4, the map visualizes the spatial distribution of the optimal solutions for 1-5 hydrometers to be installed.

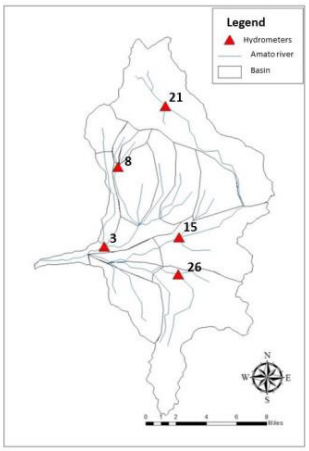

Fig. 4 Map of the 5- optimal hydrometers locations scheme

We can notice that in the case of 1 hydrometer to be installed, it is optimally located in proximity of the Amato watershed outlet point. Adding the second hydrometer, it is to be located in the south-east area of the Amato basin. The third hydrometer is to be installed in the north-west area and finally the fourth hydrometer covers the central area of the Amato basin. Adding the fifth hydrometer, it covers the north area of the Amato basin though the calibration model improves slightly (see Fig. 3).

Acknowledgements

This research work has been funded by PON R&C 2007-2013 Smart Cities and Communities and Social Innovation/ABSIDE-AQUASYSTEM Project. The authors thanks the local Protection Civil Authority, Multi-risks Functional Center of Calabria (South Italy) that supported this research works by providing all datasets related the Amato basin, a pilot area of the AQUASYSTEM Project.

Appendix A. CRESCO HPC Infrastructure

The CRESCO HPC (High Performance Computing) system installed in the ENEA Research Center of Portici (NA) consists of 3 Linux Clusters. CRESCO is integrated in ENEA GRID, a large infrastructure for cloud computing, which includes all the ENEA computing resources installed at the various research centres in Italy

The users can use computing resources x86-64 Linux systems (the Cresco HPC gather ~ 10000 cores) and dedicated systems (e.g. GPU systems). This computing resources are spread essentially across three ENEA research centres. In particular, the three research centres are ENEA-Portici with CRESCO2 (2,720 cores), CRESCO3 (2,016 cores), CRESCO 4 (4,864 cores), ENEA-Frascati with CRESCOF (480 cores); and ENEA-Casaccia with CRESCOC (192 cores). ENEA Portici is connected to the Internet through the PoP GARR of Napoli–Monte S. Angelo by means of two 1 Gbps links (GARR is the Italian Research & Education Network, planning and operating the national high-speed telecommunication network for University and Scientific Research).

CRESCO4 is the newest cluster of the data center; it is available since 2014 and consists of 304 Intel Sandy Bridge computational nodes. For each node there are 16 cores and a 64 GB RAM. The interconnection among the nodes of each cluster is based on Infiniband technology and each cluster mounts 2 kinds of shared file systems AFS (Andrew File System) and GPFS (General Parallel File System).

CRESCO 4 is used for several application in various research areas, such as Material Science, Technologies for

Energy and Industry, Environmental Modeling and Nuclear Fusion.

CRESCO 4 has been the principal infrastructure used for the implementation of the research work here presented.

As part of this work, we created a specific project area on the AFS file system with the aim to make available to all authors easy way to share data and codes. In particular, in this space we installed the codes EPA-Swmm5 and Scilab-5.4.1. In order to find the optimal sensor locations schemes for improving the simulation of flood forecasting scenarios, a *brute force* approach was used which chains Scilab and SWMM5 processes. We submitted multiple instances of the implemented algorithm on Cresco HPC system using the job array method. In this way, simultaneous instances of the *brute force* processes using different data sets were run. For submitting this type of jobs on the cluster CRESCO4, it needed to access the front-end nodes via *ssh* protocol and submit the tasks with the syntax of LSF (Load Sharing Facility, ENEA Grid scheduler).

References

[1] Walski, T. M. (1983). Technique for calibrating network models. Journal of Water Resources Planning and Management, 109(4), 360-372.
[2] Ormsbee, L. E. (1989). Implicit network calibration. Journal of Water Resources Planning and Management, 115(2), 243-257.
[3] Bush, C. A., & Uber, J. G. (1998). Sampling design methods for water distribution model calibration. Journal of Water Resources Planning and Management, 124(6), 334-344.
[4] De Schaetzen, W. B. F., Walters, G. A., & Savic, D. A. (2000). Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms. Urban Water, 2(2), 141-152.
[5] Pillar, V. D. (1999). On the identification of optimal plant functional types.Journal of Vegetation Science, 10(5), 631-640.
[6] Ahmed, S., Elsheikh, M., Stratton, I. M., Page, R. C. L., Adams, C. B. T., & Wass, J. A. H. (1999). Outcome of transphenoidal surgery for acromegaly and its relationship to surgical experience. CLINICAL ENDOCRINOLOGY-OXFORD-, 50, 561-568.
[7] Nagar, A. K., & Powell, R. S. (2000). Observability analysis of water distribution systems under parametric and measurement uncertainty. InBuilding Partnerships (pp. 1-10). ASCE.
[8] Meier, R. W., & Barkdoll, B. D. (2000). Sampling design for network model calibration using genetic algorithms. Journal of Water Resources Planning and Management, 126(4), 245-250.
[9] Lansley, P. (2001). The promise and challenge of providing assistive technology to older people. Age and ageing, 30(6), 439-440.
[10] Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2003). A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks. Journal of Hydraulic Research, 41(5), 481-492.
[11] Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2005). Multiobjective design of water distribution systems under uncertainty. Water Resources Research,41(11).
[12] Behzadian, K., Kapelan, Z., Savic, D., & Ardeshir, A. (2009). Stochastic sampling design using a multi-objective genetic algorithm and adaptive neural networks. Environmental Modelling & Software, 24(4), 530-541.
[13] Preis, A., Whittle, A., & Ostfeld, A. (2009). Multi-objective Sensor Network Placement Model for Integrated Monitoring of Hydraulic and Water Quality Parameters. In World City Water Forum, WCWF.
[14] Yu, G., Powell, R. S., & Sterling, M. J. H. (1994). Optimized pump scheduling in water distribution systems. Journal of optimization theory and applications,83(3), 463-488.
[15] Van Veldhuizen, D. A., & Lamont, G. B. (1998). Multiobjective evolutionary algorithm research: A history and analysis. Technical Report TR-98-03, Department of Electrical and Computer Engineering, Graduate School of Engineering, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
[16] Abbass, H. A. (2001). MBO: Marriage in honey bees optimization-A haplometrosis polygynous swarming approach. In Evolutionary Computation, 2001. Proceedings of the 2001 Congress on (Vol. 1, pp. 207-214). IEEE.
[17] Murata, T., & Ishibuchi, H. (1995, November). MOGA: Multi-objective genetic algorithms. In Evolutionary Computation, 1995., IEEE International Conference on (Vol. 1, p. 289). IEEE.
[18] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation, IEEE Transactions on, 6(2), 182-197.
[19] Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., & Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic acids research, 38(suppl 2), W214-W220.
[20] Hjelmfelt, A., Weinberger, E. D., & Ross, J. (1991). Chemical implementation of neural networks and Turing machines. Proceedings of the National Academy of Sciences, 88(24), 10983-10987.
[21] http://www.epa.gov/athens/wwqtsc/html/swmm.html
[22] McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 239-245.
[23] Helton, J. C., & Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety, 81(1), 23-69.