Water demand forecasting for the optimal operation of large-scale drinking water networks: The Barcelona Case Study. ⋆

Ajay Kumar Sampathirao *, **Juan Manuel Grosso** **
Pantelis Sopasakis *, **Carlos Ocampo-Martinez** **,
Alberto Bemporad *, and **Vicenç Puig** **

* *IMT Institute for Advanced Studies Lucca, Piazza San Ponziano 6, Lucca 55100, Italy.*
** *Automatic Control Department, Technical University of Catalonia (UPC). Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Llorens i Artigas, 4-6, 08028 Barcelona, Spain*

Abstract: Drinking Water Networks (DWN) are large-scale multiple-input multiple-output systems with uncertain disturbances (such as the water demand from the consumers) and involve components of linear, non-linear and switching nature. Operating, safety and quality constraints deem it important for the state and the input of such systems to be constrained into a given domain. Moreover, DWNs' operation is driven by time-varying demands and involves an considerable consumption of electric energy and the exploitation of limited water resources. Hence, the management of these networks must be carried out optimally with respect to the use of available resources and infrastructure, whilst satisfying high service levels for the drinking water supply. To accomplish this task, this paper explores various methods for demand forecasting, such as Seasonal ARIMA, BATS and Support Vector Machine, and presents a set of statistically validated time series models. These models, integrated with a Model Predictive Control (MPC) strategy addressed in this paper, allow to account for an accurate on-line forecasting and flow management of a DWN.

Keywords: Demand Forecasting, Model Predictive Control, Drinking Water Networks, Control of Large-Scale Systems

1. INTRODUCTION

1.1 Motivation

Drinking Water Networks (DWN) are large-scale, multiple-input multiple-output systems whose operation is liable to a set of operating, safety and quality-of-service constraints while at the same time their dynamics is affected by disturbances of stochastic nature (see Brdys and Ulanicki [1994], Grosso et al. [2013]). All these characteristics render their control a challenging problem. The optimal management of DWNs is a complex task with outstanding socio-economic and environmental implications and has received considerable attention by the scientific community as Ocampo-Martinez et al. [2012] and Barcelli et al. [2010] have pointed out.

Water usage can vary in both the long-term and the short-term, usually exhibiting time-based patterns for different areas. Hence, a better understanding of the characteristics of the time-series is necessary so as to perform accurate forecasts of water demand and have an optimal closed-loop performance. The overall objective of DWNs managers is to provide a reliable water supply in the cheapest way, guaranteeing availability and continuity of the service with a certain probability and without delay under some operating conditions, specific environments and uncertain events. Accordingly, optimal management of these systems is a complex task and has become an increasingly environmental and socio-economic research subject worldwide.

Therefore, the operation of a DWN is strongly conditioned by the uncertain water demand, which follows a non-stationary dynamics (see Quevedo et al. [2006]). In this paper, three well-established time series modelling methodologies are employed to capture the dynamics of water demand, namely a Seasonal Auto-Regressive Integrated Moving-Average (sARIMA) model, a BATS model developed by De Livera et al. [2011], and a Support Vector Machine model. All these models are statistically validated and are accompanied by an estimation of their prediction error. All these approaches proved to be adequate for the modelling of the demand.

On the other hand, in this paper Model Predictive Control (MPC) (see Rawlings and Mayne [2009], Qin and Badgwell [2003]) is used for computing optimal decisions regarding the operation of a DWN taking into account the management criteria and the various operating, safety and physical constraints of the problem. It is known that MPC is well suited for large-scale MIMO systems and is also

⋆ This work was financially supported by the EU FP7 research project EFFINET "Efficient Integrated Real-time monitoring and Control of Drinking Water Networks," grant agreement no. 318556.

amenable to structural changes of the dynamical model of the considered system. Moreover, a performance index related to water and energy costs is optimised leading to a suitable operation of the water network. The forecasting of demands and their variances are then used to propagate uncertainty in the open-loop prediction within the MPC strategy to assure better handling of constraints and proper fulfilment of the control objectives (management criteria).

Most of the results presented in this paper resulted from the preliminary work performed within the framework of the FP7-funded EU project EFFINET.

1.2 Hydraulic Modelling

The dynamics of DWNs have been studied in depth in the last two decades (see Brdys and Ulanicki [1994], Ocampo-Martinez et al. [2009]). The hydraulic model of the DWN consists of the mass balance equations for the water in every reservoir $i = 1, \ldots, N_r$ and distribution node $j = 1, \ldots, N_n$. Let $V_i(t)$ be the volume of water inside the tank i and let $q_{i,p}^{out}(t)$ for $p = 1, \ldots, M_i^{out}$ and $q_{i,q}^{in}(t)$ for $q = 1, \ldots, M_i^{in}$ be the influx and outflux streams from and to tank i. It then follows that:

$$\frac{\mathrm{d}V_i(t)}{\mathrm{d}t} = \sum_{p=1}^{M_i^{in}} q_{i,p}^{in}(t) - \sum_{l=1}^{M_i^{out}} q_{i,l}^{out}(t). \tag{1}$$

At every tank it should hold that:

$$V_i^{min} \leq V_i \leq V_i^{max}, \tag{2}$$

where V_i^{min} and V_i^{max} are respectively the lower and the upper tank volume capacities. The upper limit is imposed so that the tank does not overflow and should in all cases be treated as a hard constraint. In most cases, the flows $q_{i,p}^{in}$ and $q_{i,l}^{out}$ are not driven by gravity, but are instead controlled by a set of pumps which come with certain technical limitations which give rise to the constraints:

$$0 \leq q_i(t) \leq q_i^{max}. \tag{3}$$

In a DWN, a *node* is a meeting point of three or more pipes. There, the mass balance yields the static equality constraint:

$$\sum_{p=1}^{M_j^{in}} q_{j,p}^{in}(t) = \sum_{l=1}^{M_j^{out}} q_{j,l}^{out}(t) \tag{4}$$

for $j = 1, \ldots, N_n$ where N_n is the number of nodes. It is assumed here, that all flows in the network can be modelled with a set of unidirectional positive flows, which translates to the constraint $q_i(t) \geq 0$, for every flow. Certain outgoing flows q_i are actual demands from the supply network and, as such, they are stochastic variables. Let us first of all put together the aforementioned mass balance equations to arrive at the following expression in discrete-time:

$$x_{k+1} = A_d x_k + B_d u_k + G_d d_k, \tag{5}$$

where $x \in \mathbb{R}^{n_x}$ is the state vector corresponding to the volumes of water in the storage tanks, $u \in \mathbb{R}^{n_u}$ is the vector of manipulated inputs and $d \in \mathbb{R}^{n_d}$ is the vector of uncertain demands. Mass preservation gives rise to the following input-disturbance coupling equation:

$$E u_k + E_d d_k = 0, \tag{6}$$

where E and E_d are matrices of proper dimensions.

In this context, the state and input constraints can be rewritten as:

$$u_k \in \mathcal{U} \triangleq \{u \in \mathbb{R}^{n_u} \mid u^{min} \leq u \leq u^{max}\},\ \forall k \in \mathbb{N} \tag{7a}$$

$$x_k \in \mathcal{X} \triangleq \{x \in \mathbb{R}^{n_x} \mid x^{min} \leq x \leq x^{max}\},\ \forall k \in \mathbb{N} \tag{7b}$$

Both $\mathcal{X}$ and $\mathcal{U}$ are compact polytopes. The flow model that results from this analysis is a Linear Time-Invariant (LTI) discrete-time dynamical model with linear constraints which perfectly fits into the control framework of linear MPC.

1.3 Notation

In this paper, $\mathbb{N}_{[k_1,k_2]}$ denotes the set of natural numbers between k_1. Let k_2, $\mathcal{P} \subset \mathbb{R}^n$ be a polytope and $A \in \mathbb{R}^{m \times n}$ be a matrix; then it is defined $A \cdot \mathcal{P} \triangleq \{y \in \mathbb{R}^m : y = Ax; x \in \mathcal{P}\}$. Let $\mathcal{P}$, $\mathcal{Q}$ be two polytopes in $\mathbb{R}^n$; the Pontryagin difference of these polytopes is the polytope $\mathcal{P} \ominus \mathcal{Q} \triangleq \{z \in \mathbb{R}^n, \exists q \in \mathcal{Q}$, such that $z + q \in \mathcal{P}\}$. The Minkowski sum of $\mathcal{P}$ and $\mathcal{Q}$ is the polytope $\mathcal{P} \oplus \mathcal{Q} \triangleq \{z = p + q : p \in \mathcal{P}, q \in \mathcal{Q}\}$.

2. DEMAND FORECASTING

The reliable modelling and the ability to predict the upcoming water demands from every output node of the network is an essential task for the design of proper controllers for the DWN. The non-stationarity of the demand time series along with the presence of multiple seasonal patterns calls for state-of-the art modelling approached that can capture such complex dynamics.

In this section, three different modelling approaches are presented to model the water demand from a DWN case study. These models are trained using the same dataset of 2700 demand measurements from the DWN of Barcelona, out of 8760 data points that are available. The rest is used as an external test-set for validating these models. These data were provided by AGBAR (Aguas de Barcelona, s.a.), which is the company that manages the Barcelona DWN).

2.1 Seasonal ARIMA Time-Series Models

ARIMA models are widely used as they can capture complex linear dynamics of stationary processes or processes that become stationary after one applies the difference operator finitely many times. ARIMA models put more emphasis on the recent past of the time series they intercept, so, they are considered to be suitable for short-term forecasting (see Box et al. [1994]). Here we consider Seasonal ARIMA models seasonally integrated with seasonal AR (Auto-Regressive) terms which describe well time series that follow a periodic pattern.

In order to derive ARIMA models, the following notation is introduced. For a time series d_k, let $Ld_k = d_{k-1}$ be the *backward shift operator*. For $i \geq 2$, define $L^i = LL^{i-1}$, then, $L^i d_k = d_{k-i}$. Denoting by 1 the identity operator $1z_t = z_t$ and $\nabla = 1 - L$, it follows that $\nabla d_t = d_t - d_{t-1}$. Let α_t be a white noise process, i.e., a time series such that $\mathbb{E}\alpha_t = 0$, $\alpha_t \sim N(0, \sigma_\alpha^2)$ and $\mathrm{cov}(a_t, a_{t+k}) = 0$ for all $k \neq 0$.

Let ϕ be a polynomial of L of order p with unitary constant term, symbolically $\phi \in \mathrm{K}_p^1[L]$, i.e., $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 -$

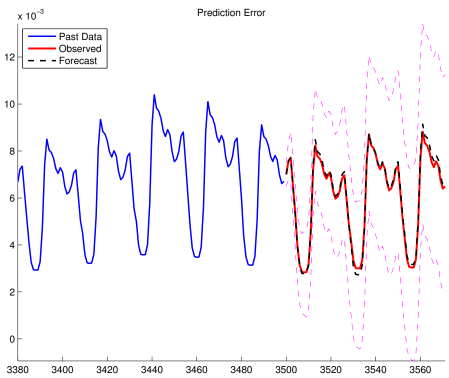

Fig. 1. Stochastic forecasts using the SARIMA model (8). The purple dashed thin lines represent the 99% upper and lower bounds computed by a Monte Carlo simulation using 10^5 seeds.

$\ldots - \phi_p L^p$ and let $\psi \in \mathrm{K}_q[L]$ be a polynomial with of the form $\psi(L) = 1+\psi_1 L + \ldots + \psi_q L^q$. Then, an ARIMA$(p, d, q)$ model has the form $\phi(L)(1-L)^d \bar{d}_t = \psi(L)\alpha_t$, where $\bar{d}_t = d_t - \mu$ where $\mu = \mathbb{E}d_t$ is the (known) expected value of d_t. In certain cases, ϕ and ψ are assumed to have some of their coefficients fixed to 0. Then, if χ_p and χ_q are the sets of non-zero coefficients of ϕ and ψ repectively (with $\max \chi_p = p$ and $\max \chi_q = q$), the respective ARIMA model is denoted by ARIMA(χ_p, d, χ_q).

Water demand generally exhibits a periodic variation which in time-series analysis is referred to as *seasonality*, which might follow some calendric trend such as daily, weekly or monthly. This periodicity is captured by a term of the form $1 - L^s$ (that acts on $\bar{d}_t$) where s is the seasonality of the process and the respective model is denoted as SARIMA$(\chi_p, d, \chi_q; s)$. The model can be enhanced by a multi-seasonal auto-regressive polynomial operator $\Psi(L) = 1 - \Psi_{t_1} L^{t_1} - \ldots - \Psi_{t_Q} L^{t_Q}$ whose exponents form the set χ_Q^{Ψ}. Such a model is denoted by SARIMA$(\chi_p, d, \chi_q; s) \times$ SAR(χ_Q^{Ψ}).

A set of models was created for each of the 88 demands of the DWN of Barcelona. Indicatively, we present the following model for the demand form the output node `c450BEG` of the Barcelona DWN:

$$\begin{aligned}&\mathrm{ARIMA}(\{1:4, 6:9\}, 1, \{1:13, 15, 17\}; 168)\times\\ &\mathrm{SAR}(\{168, 336\}),\end{aligned} \tag{8}$$

which was trained using 2700 samples and was formulated as a maximum likelihood problem. All the parameters of this model were determined with high statistical significance (based on their t-statistic).

For the evaluation of the predictive power of each model we forecast the water demand throughout a horizon of $H_p = 24$h ahead, but will enable us to demonstrate how each model performs in the long run. Let $\hat{d}_{k+j|k}$ be the predicted expected demand of water for the future time instant $k+j$ performed at time k. Let d_{k+j} be the actual value of the demand. The total prediction error along the period $[k, k+H_p]$ is quantified by the prediction mean squared error (PMSE), defined as:

$$\mathrm{PMSE}_{H_p,k} = \frac{1}{H_p}\sum_{i=1}^{H_p}(\hat{d}_{k+i|k} - d_{k+i})^2, \tag{9}$$

and its square root is the PRMSE, i.e., $\mathrm{PRMSE}_{H_p,k} = \sqrt{\mathrm{PMSE}_{H_p,k}}$. Model (8) was found to have $\mathrm{PMSE}_{24} = 0.0158$ (with st. dev. 0.0049) and a $\mathrm{PRMSE}_{24} = 0.1311$ on average [1]. This model passed the Ljung-Box test for uncorrelated residuals with a p-value of 0.2908, the value of the Ljung-Box statistic being 22.96 with critical value 31.41. This model was selected among a set of many other models using the Akaike Information Criterion given by $\mathrm{AIC} = \ln(\hat{\sigma}_\alpha^2) + 2k/T$, where $\hat{\sigma}_\alpha^2$ is the statistical estimate of σ_α^2, k is the number of parameters of the model and T is the number of observations used for the estimation of the model. For the model in (8) there was obtained an $\mathrm{AIC} = -8.5044$.

Conclusively, model (8) interpolates very well the demand time series, it has high predictive ability, its residuals are uncorrelated, it is determined with high statistical significance and it is invertible.

2.2 Box-Cox transformation, ARMA Errors, Trends and Seasonality (BATS) Modelling

BATS models were introduced by De Livera et al. [2011] and have proven to be well-suited for modelling time series with multiple seasonal patterns and complex dynamics.

Let d_k, $k \in \mathbb{N}$, denote an observed time series of any water demand, and $d_k^{(\omega)}$ its Box-Cox transformation with the parameter ω. The transformed series is then decomposed into an *irregular* component h_k, a *level* component l_k, a *growth* component b_k and possible seasonal ones $s_k^{(i)}$ with seasonal frequencies m_i, for $i = 1, \ldots, P$, where P is the total number of seasonal patterns in the series. The irregular component h_k is described by an ARMA(p, q) process with parameters ϕ_i for $i = 1, \ldots, p$ and θ_i for $i = 1, \ldots, q$, and an error term ε_k which is assumed to be a Gaussian white noise process with zero mean and constant variance σ^2. The smoothing parameters, given by α_d, β_d, $\gamma_{d,i}$ for $i = 1, \ldots, P$, determine the extent of the effect of the irregular component on the states l_k, b_k, $s_k^{(i)}$ respectively. The equations of the model are:

$$d_k^{(\omega)} = \begin{cases} \dfrac{d_k^{(\omega)} - 1}{\omega}, & \omega \neq 0, \\ \log(d_k), & \omega = 0, \end{cases} \tag{10a}$$

$$d_k^{(\omega)} = l_{k-1} + \phi b_{k-1} + \sum_{i=1}^{P} s_{k-m_i}^{(i)} + h_k, \tag{10b}$$

$$l_k = l_{k-1} + \phi b_{k-1} + \alpha_d h_k, \tag{10c}$$

$$b_k = \phi b_{k-1} + \beta_d h_k, \tag{10d}$$

$$s_k^{(i)} = s_{k-m_i}^{(i)} + \gamma_{d,i} h_k, \tag{10e}$$

$$h_k = \sum_{i=1}^{p} \varphi_i h_{k-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{k-i} + \varepsilon_k. \tag{10f}$$

[1] Based on external data and using 150 samples. The computation was carried out against an external test-set, not available to the model.

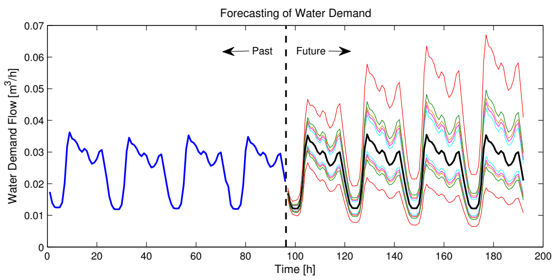

Fig. 2. Forecasting of water demand using the proposed BATS model. (Black thick line) expected forecast, (surrounding lines) confidence intervals for 90%, 95%, 97%, 99% and 99.9999% confidence levels.

The decomposition of the time series into such components and the use of them for modelling and forecasting offers also a qualitative insight into the dynamics of the process.

A set of BATS models was trained for the demand time series by maximizing their log-likelihood according to De Livera et al. [2011] using two seasonal patterns: one daily and one weekly. The best BATS model for each demand time series was chosed on the basis of the Akaike Information Criterion given by AIC $= L^\star + 2k$, where $L^\star$ is the maximum log-likelihood of the model, and k is the number of tunable parameters of the model.

2.3 RBF-based Support Vector Machine

Excellent results were obtained using a Support Vector Machine (SVM) model with a Radial Basis Function (RBF) kernel with $\gamma = 0.015$. The problem was formulated as an ϵ-SVM (see Cortes and Vapnik [1995]) with $\epsilon = 10^{-5}$ and was solved using the celebrated library libSVM by Chang and Lin [2011]. The parameter C of the cost function of the problem was set to 1000. The explanatory variables used were 200 past values of the demand and a set of binary calendar variables as follows: Let m_i be 1 if the day when the measurement was taken was the i-th day of the week (for $i = 0, \ldots, 6$) and 0 otherwise. Let h_j be the corresponding variable referring to the hour of the day for $j = 0, \ldots, 23$. In this way, the information about the seasonal pattern of the time series is encoded. All the features of the dataset were scaled to the interval $[0, 1]$.

A 10-fold cross-validation of the model produced a $q^2 = 0.9952$. It was found that $\text{PMSE}_{24} = 0.0065$ (with st. dev. 0.0051) and $\text{PRMSE}_{24} = 0.0743$ on average based on 150 samples. The performance of the SVM-based predictor and the stringency of the confidence intervals for its forecasts is illustrated in Fig. 3.

2.4 Comparison and Evaluation

The three proposed models are concisely compared in Table 1 in regard to their complexity and predictive ability against external data.

3. OPTIMAL OPERATING MANAGEMENT

3.1 Control Objectives

The formulation of the MPC problem amounts to determining optimal control actions so as to minimise the water production and transportation cost, minimise the deviation of the volume of water in storage tanks from the prescribed operating limits and deliver smooth control actions. These costs were quantified following the approach reported by Barcelli et al. [2010].

Briefly, the *water production and transportation cost* is quantified by the cost function:

$$\ell^w(u_k, k) \triangleq W_\alpha \left(\alpha_1 u_k + \alpha_{2,k} u_k\right), \tag{11}$$

where α_1 is a vector related to the production costs of water according to the selected source (treatment plant, dwell, etc.), and $\alpha_{2,k}$ is associated with the pumping cost for the transportation of the water through certain paths. W_α is a proper scaling factor. The *safety-storage cost* is given by $\ell^s(x_k) = s_k' W_x s_k$, where $s_k = \max\{0, x^s - x_k\}$. This cost term is, however, non-quadratic, but we may replace it with the term:

$$\ell^S(\xi_k) = \xi_k' W_x \xi_k, \tag{12}$$

where $\xi_k = \xi_k(x_k) \geq x^s - x_k$ accompanied by the convex constraint $\xi_k \geq 0$. The *smooth operation cost* is defined as:

$$\ell^\Delta(\Delta u_k) = (\Delta u_k)' W_u \Delta u_k, \tag{13}$$

where W_u is a n_u-by-n_u square positive semi-definite matrix, and $\Delta u_k = u_k - u_{k-1}$.

3.2 Formulation of the MPC problem

MPC algorithms recently started being used for the control of DWNs (see for example Ocampo-Martinez et al. [2012]) as they guarantee certain quality of service, satisfaction of the operating constraints of the plant and optimal operation (according to the prescribed performance indices). In this section, an MPC problem is formulated taking into account the nominal demand forecasts and the estimated bounds for the prediction error (which are generated by the demand forecast module). It is assumed that, at every

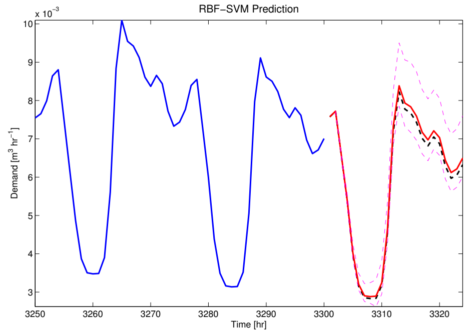

Fig. 3. Forecasting of the demand time series using the proposed RBF-SVM model. (Blue line): past demand data, (Red line): Actual demand, (Black dashed line): SVM predictions, (Light magenta dashed lines): 99% confidence intervals calculated by Monte Carlo simulations using 5000 seeds.

Table 1. Comparison of the Predictive Models

Performance Index	sARIMA	BATS	RBF-SVM
Average PMSE_{24}	0.0158	0.0043	0.0065
Average PRMSE_{24}	0.1311	0.0584	0.0743
No. Parameters	25	26	229

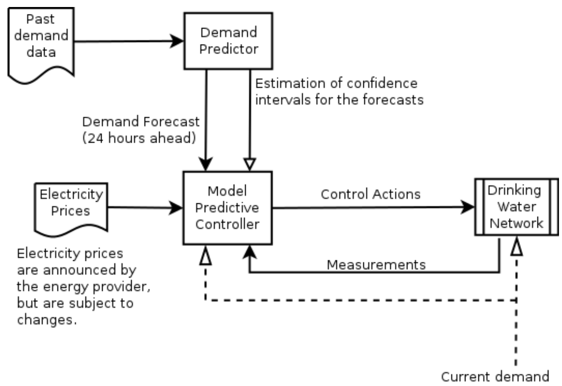

Fig. 4. The closed loop system with the MPC controller and a demand estimator.

time instant k, the controller has access to the current demand measurement d_k and to a set of H_p future demand estimates, namely $\hat{d}_{k+j|k}$ such that

$$d_{k+j|k} = \hat{d}_{k+j|k} + \epsilon_{k+j|k}, \forall j \in \mathbb{N}_{[0,H_p]} \tag{14}$$

where $\epsilon_{k+j|k}$ is an error term drawn from a compact set $\mathcal{E}_{k+j|k} = \{\epsilon : \ \epsilon^{min}_{k+j|k} \leq \epsilon \leq \epsilon^{max}_{k+j|k}\}$ which contains the origin in its interior; in practice, such sets can be chosed to be the 99.9% (or higher) confidence intervals of the prediction error. Henceforth, the index $k+j|k$ will refer to a prediction at time k for the future time instant $k+j$.

The total cost function and performance index is made up of the aforementioned costs; it is:

$$J(x_k, \mathbf{u}_k, \Xi_k, \Delta\mathbf{u}_k, k) = L^w(\mathbf{u}_k, k) + L^\Delta(\Delta\mathbf{u}_k) + L^S(\Xi_k),$$

where H_p is the *prediction horizon* and H_u is the *control horizon* (normally $H_u \in \mathbb{N}_{[1,H_p-1]}$). L^w is the *total water production cost*, L^Δ is the *total smooth operation cost* and L^S is the *total safety volume cost* given by:

$$L^w(\mathbf{u}_k, k) = \sum_{i \in \mathbb{N}_{[0,H_u]}} \ell^w(u_{k+i|k}, k), \tag{15a}$$

$$L^\Delta(\Delta\mathbf{u}_k) = \sum_{i \in \mathbb{N}_{[0,H_u-1]}} \ell^\Delta(\Delta u_{k+i|k}), \tag{15b}$$

$$L^S(\Xi_k) = \sum_{i \in \mathbb{N}_{[0,H_p-1]}} \ell^S(\xi_{k+i|k}). \tag{15c}$$

Hereinafter, it is denoted $\mathbf{u}_k \triangleq (u_k, u_{k+1|k}, \ldots, u_{k+H_u|k})$ for the sequence of control actions, $\Delta\mathbf{u}_k \triangleq (u_{k+1|k} - u_k, \ldots, u_{k+H_u|k} - u_{k+H_u-1|k})$ for the successive differences of $\mathbf{u}$, and $\Xi_k = (\xi_{k+1|k}, \ldots, \xi_{k+H_p|k})$. The vector $\mathbf{d}_k$ that appears in the formulation of the MPC problem is defined as $\mathbf{d}_k = (d_k, \hat{d}_{k+1|k}, \ldots, \hat{d}_{k+H_p-1|k})$ and is provided to the controller from the demand forecast module along with the vector of maximum/minimum estimated prediction errors:

$$\mathbf{e}_k = (\epsilon^{min}_{k+1|k}, \ldots, e^{min}_{k+H_p|k}, \epsilon^{max}_{k+1|k}, \ldots, e^{max}_{k+H_p|k}).$$

It is known that the prediction error $\epsilon_{k+j|k} \in \mathcal{E}_{k+j|k}$ and $\mathcal{E}_{k|k} = \{0\}$ because $d_{k|k} = d_k$ is measured. This implies that the predicted sequence of states is:

$$x_{k+j|k} = \hat{x}_{k+j|k} + \sum_{l=1}^{j} A^{l-1} G_d \epsilon_{k+l|k}, \tag{16}$$

where $\hat{x}_{k+j|k}$ stands for the nominal predicted state, whose dynamics, starting from $\hat{x}_{k|k} = x_k$ is described following (5), i.e.,

$$\hat{x}_{k+j+1|k} = A\hat{x}_{k+j|k} + Bu_k + G_d\hat{d}_{k+1|k}, \tag{17}$$

The MPC problem is then formulated as follows:

$$\mathbb{P}_{H_p,H_u}(x_k, \mathbf{d}_k, \mathbf{e}_k, k):$$
$$J^\star(x_k, \mathbf{d}_k, k) = \min_{\mathbf{u}_k, \Xi_k} J(x_k, \mathbf{u}_k, \Xi_k, \Delta\mathbf{u}_k, k) \tag{18a}$$

subject to the constraints:

$$\hat{x}_{k+i|k} \in \mathcal{X} \ominus \bigoplus_{j=1}^{i} A^{j-1} G_d \mathcal{E}_{k+j|k}, \forall i \in \mathbb{N}_{[1,H_p-1]} \tag{18b}$$

$$u^{min} \leq u_{k+i|k} \leq u^{max}, \forall i \in \mathbb{N}_{[0,H_u]} \tag{18c}$$

$$\hat{x}_{k+i+1|k} = A\hat{x}_{k+i|k} + Bu_{k+i|k} + G_d\hat{d}_{k+i|k}, \quad \forall i \in \mathbb{N}_{[0,H_p-1]} \tag{18d}$$

$$Eu_{k+i|k} + E_d\hat{d}_{k+i|k} = 0, \forall i \in \mathbb{N}_{[0,H_u]} \tag{18e}$$

$$u_{k+j|k} = u_{k+H_u|k}, \forall j \in \mathbb{N}_{[H_u+1,H_p-1]} \tag{18f}$$

$$\xi_{k+i|k} \geq x^s - \hat{x}_{k+i|k}, \forall i \in \mathbb{N}_{[0,H_p]} \tag{18g}$$

$$\xi_{k+i|k} \geq 0, \forall i \in \mathbb{N}_{[0,H_p]} \tag{18h}$$

$$\hat{d}_{k|k} = d_k, \text{ and } \hat{x}_{k|k} = x_k \tag{18i}$$

Equation (18b) implies that for all $i \in \mathbb{N}_{[0,H_p-1]}$, then $x_{k+i|k} \in \mathcal{X}$ as long as $\epsilon_{k+j|k} \in \mathcal{E}_{k+j|k}$ for all $j \in \mathbb{N}_{[1,i]}$. Normally, in order to perform the set operations in (18b), it is required to iterate over all vertices of the sets $\mathcal{E}_{k+j|k}$ (i.e., 2^{n_d} elements). However, in the case study of this paper, G_d is a very sparse matrix (maximum 3 non-zero elements per row) so the complexity is such that allows the on-line implementation of the proposed algorithm. Notice that the (predicted) state is constrained in a set that is smaller than $\mathcal{X}$ and that the constraints are time-varying along the prediction horizon and are conditioned by the estimated prediction error.

The above optimisation problem can be easily formulated as a convex quadratic problem and solved on-line, very efficiently, using state-of-the-art methods such as the global piece-wise smooth Newton method by Patrinos et al. [2011] and the accelerated dual gradient-projection method by Patrinos and Bemporad [2012].

Let us denote by:

$$(\mathbf{u}^\star_k(x_k, k), \Xi^\star_k(x_k, k)) = \operatorname{argmin} J(x_k, \mathbf{u}_k, \Xi_k, \Delta\mathbf{u}_k, k),$$

where:

$$\mathbf{u}^\star_k(x_k, k) = (u^\star_{k|k}(x_k, k), \ldots, u^\star_{k+H_u|k}(x_k, k)).$$

Then, the associated MPC control law is $\kappa(x_k, k) = u^\star_{k|k}(x_k, k)$, i.e., the first element of the sequence $\mathbf{u}^\star_k(x_k, k)$ is applied to the system according to the *receding horizon control* strategy.

This approach allows for the decoupling of the predictor from the controller. The effect of the predictor's estimated error on the controller is clear from the formulation of the problem as in (18): the bigger the error is, the stricter the constraints and the more conservative the control policy will be.

3.3 Control of the Barcelona DWN

The case study addressed in this paper is the Barcelona DWN, which has been previously reported by Ocampo-

Martinez et al. [2011]. The system is modelled following Section 1.2, using a linear time-invariant dynamical system with 63 state variables, 114 manipulated inputs, 88 disturbances and 17 flow intersection nodes. The time series models calibrated in Section 2 are used in a closed-loop setting with the MPC controller described above. The weighting matrices in (11), (12) and (13) are chosen to be $W_\alpha = 1$, $W_u = I$, and $W_x = 10^{-6} \cdot I$, respectively. All other parameters and technical characteristics, such as x^{min}, x^{max}, x^s were specified by the network manager (AGBAR). The prediction horizon $H_p = 24$h and the control horizon $H_u = 23$h were considered with a sampling time of 1h.

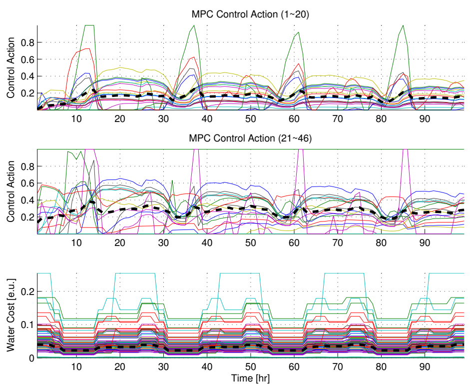

Fig. 5. The MPC control action (scaled in the range $[0, 1]$) and the water production cost. The bold dashed lines represent the average values (average control action and average production cost). The input trajectories are split in two graphs for clarity. Only pumping actions are presented here.

As shown in Fig. 5, the controller tends to operate the pumps when the electricity cost is low. Moreover, Fig. 6 shows that the volume of water in the tanks remains always between the prescribed bounds and tends to stay over the safety storage limit.

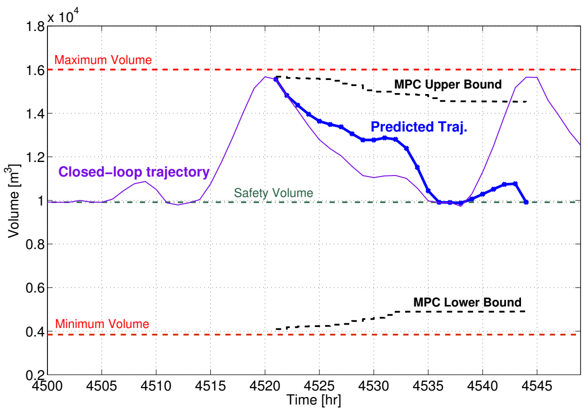

Fig. 6. The controlled trajectory of the volume of water in the tank `d130BAR` of the Barcelona DWN.

The optimisation problem (18) was solved online very efficienty using CPLEX. In terms of complexity, it comprises 4248 decision variables, 1512 bound constraints, 408 linear equations and 4536 linear inequality constraints. On average (based on 500 samples) the computational time for the solution of the optimisation problem was 1.85s (st. dev.: 0.055s) and the time needed for the its formulation was 0.018s (st. dev.: 0.005s).

REFERENCES

D. Barcelli, C. Ocampo-Martinez, V. Puig, and A. Bemporad. Decentralized model predictive control of drinking water networks using an automatic subsystem decomposition approach. In *Symposium on large scale complex systems theory and applications*, 2010.

G. E. P. Box, G. M. Jenkins, and G. C. Reisnel. *Time Series Analysis, Forecasting and Control*. Prentice-Hall International, Inc., 1994.

M. Brdys and B. Ulanicki. *Operational Control of Water Systems: Structures, algorithms and applications*. Prentice Hall International, 1994.

C. C. Chang and C. J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

C. Cortes and V. Vapnik. *Support-Vector Networks*. Springer, 1995.

A. M. De Livera, R. J. Hyndman, and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.

J. M. Grosso, C. Ocampo-Martinez, and V. Puig. Learning-based tuning of supervisory model predictive control for drinking water networks. *Engineering Applications of Artificial Intelligence*, 26(7):1741–1750, 2013.

C. Ocampo-Martinez, V. Puig, G. Cembrano, R. Creus, and M. Minoves. Improving water management efficiency by using optimization-based control strategies: the barcelona case study. *Water Science & Technology: Water supply*, 9(5):565–575, 2009.

C Ocampo-Martinez, S Bovo, and V Puig. Partitioning approach oriented to the decentralised predictive control of large-scale systems. *Journal of Process Control*, 21(5): 775–786, 2011.

C. Ocampo-Martinez, D. Barcelli, V. Puig, and A. Bemporad. Hierarchical and decentralised model predictive control of drinking water networks: Application to barcelona case study. *IET Control Theory and Applications*, 6(1):62–71, 2012.

P. Patrinos and A. Bemporad. An accelerated dual gradient-projection algorithm for linear model predictive control. In *51st Conference on Decision and Control*, pages 662–667, 2012.

P. Patrinos, P. Sopasakis, and H. Sarimveis. A global piecewise smooth newton method for fast large-scale model predictive control. *Automatica*, 47(9):2016 – 2022, 2011. ISSN 0005-1098.

S.J. Qin and T.A. Badgwell. A survey of industrial model predictive control technology. *Control engineering practice*, 11(7):733–764, 2003.

J. Quevedo, V. Puig, G. Cembrano, J. Aguilar, C. Isaza, D. Saporta, G. Benito, M. Hedo, and Molina A. Estimating missing and false data in flow meters of a water distribution network. In *Proceedings of the IFAC SAFEPROCESS*, Beijing (China), 2006.

J. B. Rawlings and D. Q. Mayne. *Model predictive control: theory and design*. Madison: Nob Hill Publishing, 2009.