IPSJ Transactions on Mathematical Modeling and Its Applications Vol.10 No.2 33–42 (July 2017)

Regular Paper

Stable and Energy Efficient Operation in a Large-scale Water Distribution Network

Toshihiro Hirano[1,†1,a)] Yuta Ashida[1] Lu Yu[2]
Sayaka Izukura[1] Itaru Nishioka[1] Ryohei Fujimaki[1]

Received: January 30, 2017, Accepted: February 28, 2017

Abstract: Recent advances in data analysis techniques have been attracting an increasing amount of attention from industry persons who operate complex social infrastructures. In this paper, we present our experience in using mathematical programming techniques to achieve energy efficient operations in a large-scale water distribution network. Our study mainly considers two issues. The first is how to achieve a stable water supply in the face of uncertain water demand while avoiding the occurrence of a water failure. The second is how to determine both water distribution and pump operations simultaneously in order to avoid unnecessary energy consumption. To address these issues, we propose a mathematical optimization solution under physical and operational constraints. It results in a stable water supply and fine-grain pump operations. Finally, with respect to a large-scale real water distribution network, we demonstrate the effectiveness of the proposed method in terms of energy consumption.

Keywords: mathematical programming, optimal water distribution and pump operation planning, pump modeling, stable water supply, water resource management

1. Introduction

In recent years, water resources are becoming more and more valuable due to growth in the world population. Rapid urbanization, in particular, has led to steadily increasing water demand in large cities. In order to provide a stable supply of clean water, water supply utilities consume a huge amount of electric energy in such processes as water purification and distribution. In Japan, for example, it has been reported that water supply utilities use approximately 1% of the total energy consumed in a city, and 60% of that energy consumption is used in the water distribution process. This fact motivates water supply utilities to try to manage water distribution more efficiently, not only to reduce energy costs but also to assume a social responsibility for mitigating the ecological impact of greenhouse gas emissions.

The task of increasing water-distribution efficiency, however, can be especially complex for water supply utilities in urban areas, whose large-scale water distribution networks include purification plants, reservoirs, tanks, and pump stations. In order to reduce energy consumption while meeting the growing and uncertain demand, the control of complicated water distribution networks needs to be intelligent.

In this paper, we focus on daily operations of a real water supply utility in an anonymous city and propose an optimization method that enables us to reduce energy consumption while meeting stable water supply requirements. It is designed to support a stable water supply with minimum energy consumption, to which purpose we have had to consider the following three problems. The first problem is non-coordinated operations among facilities, such as purification plants, reservoirs, and pump stations. Currently, individual facilities generally determine their own operational plans without taking overall energy efficiency into consideration. Therefore, it has often been observed that inefficient facilities treat greater amounts of water than do highly efficient ones. In order to treat appropriate water quantities at individual facilities, holistic optimization based on a valid metric for energy efficiency is required. The second problem is excessive water production. In the current operation, individual water purification plants tend to produce water in amounts that include large leeway in order to assure a stable water supply. Most of this leeway, however, is likely to go unutilized water and to be discarded by its expiration time. Roughly 30% of energy consumption in a water supply utility reportedly goes into the water purification process, and the energy consumed in the purification of eventually unutilized water is not negligible in the total energy consumption. Lastly, the third problem is how to create two separate plans for a single day: one for water distribution and the other for pump operations. Such plans are created in the pump operation planning for the water distribution, which determines how water distribution is to be conducted and how pumps are to be operated for each 15 minutes of that day. Currently, operators need one hour or more to create the two plans, and this is a significant burden. Additionally, since the plans are created simply on the basis of operators' experience, they do not necessarily result in the efficiency of water distribution and pump operations needed to achieve low energy consumption.

1 NEC Data Science Research Laboratories, Kawasaki, Kanagawa 211–8666, Japan
2 NEC Laboratories China, Beijing 100084, China
†1 Presently with College of Economics, Kanto Gakuin University
a) 1hirano2@kanto-gakuin.ac.jp

Our paper proposes an optimization method designed to address these problems for more efficient water distribution. In details, our contributions are as follows:

- Consistent energy consumption metric: from actual pump parameters, we develop a physical model to describe energy consumption and specify a flow range for which individual pumps work efficiently. From the physical model, we generate functions of the energy consumption metric by means of the piecewise linear approximation. This approximation simplifies our mathematical optimization problems and thus reduces computational time.
- Robust water production: we formulate a daily water production planning as a robust optimization (RO) problem. This formulation can significantly contribute to a stable water supply under uncertain water demand in the real-world water distribution network.
- Fine-grain pump operations: pump operations over multiple time periods are formulated as a mixed integer linear programming (MILP). Moreover, in the MILP, we propose physical and operational constraints that restrict the time intervals required for preparation of the next switch, switching counts, and the variation width of switching. Experimental results for the real-world water distribution network show that the fine-grain pump operations, by optimizing water distribution and pump operations simultaneously, reduce energy consumption by 14.0%.

The remainder of this paper is organized as follows. We first review related optimization work in Section 2 and give a configuration of a water distribution network model in Section 3. Section 4 presents a detailed description of two problems in the current water distribution management. Section 5 describes our pump modeling with a new energy consumption metric. In Section 6, we formulate the mathematical optimization problems for the operational problems discussed in Section 4. In Section 7, experimental results are given for the real-world water distribution network. Finally, we conclude this study in Section 8.

2. Related Work

There is an increasing interest with respect to optimization problems in water distribution networks, such as determination of stable water production quantities and efficiency in water distribution and pump behavior. We review some previous works related to stable water management and energy efficient operations.

To accomplish stable water management, it is first necessary to build a prediction model for water demand. Several studies predict water consumption with polynomial regression [8], a categorical approach [14], and an integrated method of the wavelet and artificial neural network [12]. As compared to these methods, our paper employs a piecewise sparse linear regression model [5] which has both the prediction accuracy and the interpretability required in practice.

Forecasted water demand values are used for a demand satisfaction constraint in the RO for stable water management. Reference [11] proposes the RO where uncertainty is a set of possible probability density functions, but Ref. [11] focuses only on reduction of overall pressure at selected points. Reference [13] minimizes energy consumption in the RO, but uncertainty of water demand is expressed as a box constraint. Our paper adopts a standard ellipsoidal constraint (see, e.g., Ref. [3]) to determine the daily water production, using a covariance matrix estimated from prediction errors in the piecewise sparse linear regression model.

The determined daily water production is distributed to each demand point on the basis of the pump operation planning for the water distribution. Reference [4] surveys recent researches related to the above planning in the field of the water network optimization. Reference [1] creates an energy metric function and determines efficient water distribution for a single day in the large-scale water distribution network. However, unlike our physical pump modeling approach, the flow range, in which all combinations of pumps work efficiently, is not specified because the energy metric is calculated using only historical data on a limited area of pump operations. Reference [2] finds the optimal pump scheduling using the combination of the grid search with the Hooke-Jeeves pattern search method. This optimization problem includes the pump constraint for the number of switches for each pump during a single day. However, pump on/off switches are individually defined, and Ref. [2] does not take interaction among pumps into account. In practice, we can often operate just combination of several pumps due to the specification of water facilities. Therefore, in this paper, we define on/off switch for the combination of several pumps, and we limit its switch counts. Reference [15] also develops the mathematical programming which minimizes the maximum value of energy consumption in a time zone and proposes the optimal operational plan of pumps. Reference [6] optimizes both water distribution and pump operations for dynamic settings and formulates a mixed integer nonlinear program with a linear objective function and quadratic constraints. The studies [6], [15] do not include pump constraints such as the pump switching-count limitation required by water supply utilities. There is also a rich literature regarding to heuristic optimization for water distribution problems (e.g., Refs. [9], [10], [16]). It is generally difficult to satisfy operational constraints with a heuristic approach compared to a mathematical programming approach.

This paper proposes a mathematical optimization solution to achieve a stable water supply based on the prediction for water demand and efficient pump operations for the water distribution. Particularly, unlike previous works, our mathematical programming introduces additional binary optimization variables so as to include the physical and operational constraints for: (1) time intervals required for preparation of the next switch, (2) switching counts, and (3) variation width of switching. Moreover, we develop a new metric of energy consumption based on a physical model. To the best of our knowledge, no existing study has formulated a series of entire water distribution process in consideration of a stable water supply and energy efficiency while still satisfying desirable pump behavior.

3. Water Distribution Network

To begin with, we explain detailed and comprehensive model

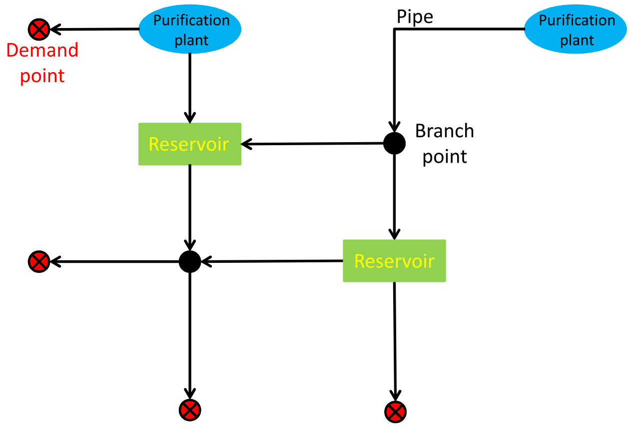

Fig. 1 Sketch of overall water distribution network of the water supply utility.

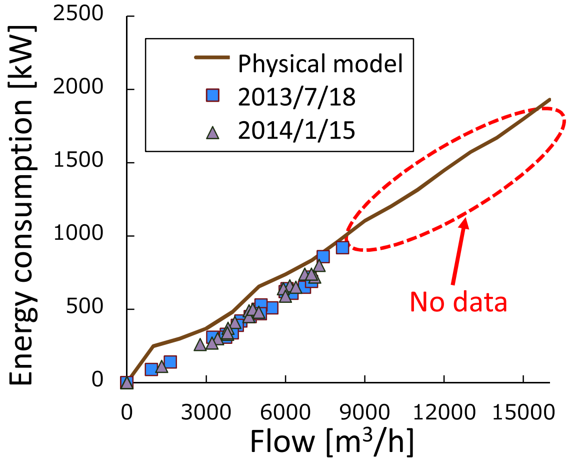

Fig. 2 Comparison between the energy consumption curve based on the physical model and past actual values at a typical pump station.

of the water distribution network of the water supply utility. As shown in **Fig. 1**, the water distribution network is composed of purification plants, reservoirs, branch points, demand points, and pipes. The purification plants and the reservoirs are facilities for clean water production and its storage respectively. There are one tank [*1] and pump stations inside the purification plant and the reservoir. The tank is capable of storing the water. Multiple pumps are equipped in each pump station, and the pumps are used to transfer kinetic energy to a mass of water along the pipes. The flow capacity can be increased by connecting two or more pumps in parallel by means of set procedures. Namely, the feasible range of the water flow at the pump station is determined by a combination of several pumps. In this paper, the combination of pumps is called as operating pump pattern for convenience. The branch points either integrate different pipes into a single pipe or separate a single pipe into multiple pipes. Using this water distribution network, the water supply utility offers clean water service and must satisfy time-varying water demand at end user points.

4. Current Water Distribution Management

This section provides a detailed description of a series of operating processes in the large-scale water distribution network. Individual facilities conduct two planning, that is, the daily water production planning and the pump operation planning for the water distribution.

The daily water production planning is conducted to determine water supply quantities to be produced at each purification plant in the next day, and it is based on predicted water demand. This prediction is currently performed by the empirical method with the past water demand. The use of heuristic methods in the prediction and in the determination of water supply quantities may often result in inaccurate plan. If the actual demand values are not likely to be satisfied, operators need to modify the original plan during daily operations. To make the water production process more stable, we need to create an accurate water production plan that meets actual daily demand.

The pump operation planning for the water distribution, as previously noted, generates two separate plans for a single day: one for water distribution and the other for pump operations every 15 minutes. In the current operation, operators spend one hour or more to develop these plans in the next day manually based on their experience. Therefore, it is great promise to apply mathematical optimization techniques to this problem for the reduction not only of energy consumption but also of operators' workload.

From the pump operation planning for the water distribution, we can also determine water supply quantities, but the water supply utility sets the daily water planning to conduct a plan modification surely for the contingent increase of water demand. Therefore, our optimization solution is applied to each planning.

5. Energy Consumption Modeling

In this section, we briefly describe our new energy consumption metric, which is based on the physical model of pumps.

5.1 Energy Consumption Metric

To achieve holistic energy optimization, we need to specify a consistent metric for energy consumption over an entire water distribution network. For such a metric, energy consumption per unit of water flow rate has been employed in Ref. [1]. This metric is a function of flow rate, which can be approximated with the piecewise linear function by fitting historical data. However, the derivation of the metric function has a problem in the energy optimization. As seen in **Fig. 2**, the historical data to be used for approximation consists of very few samples on the limited range of daily pump operations of the water supply utility, and consequently the approximated metric function is strongly biased. To avoid this problem, we propose a new energy metric function that is derived from the physical pump model.

5.2 Approximated Pump Energy Consumption

Water distribution pumps are used to transfer large quantities of water among purification plants, reservoirs, and demand points. The pumps impart energy to water, thereby raising its hydraulic head. Head is a measurement of the height of the liquid column and is created from the kinetic energy produced by the pump.

In our physical pump modeling, energy consumption P [kW] of the pump is calculated as

[*1] To be exact, several tanks are equipped in each facility. However, since this paper focuses on the total water quantity of tanks, we assume that the purification plants and reservoirs have one tank.

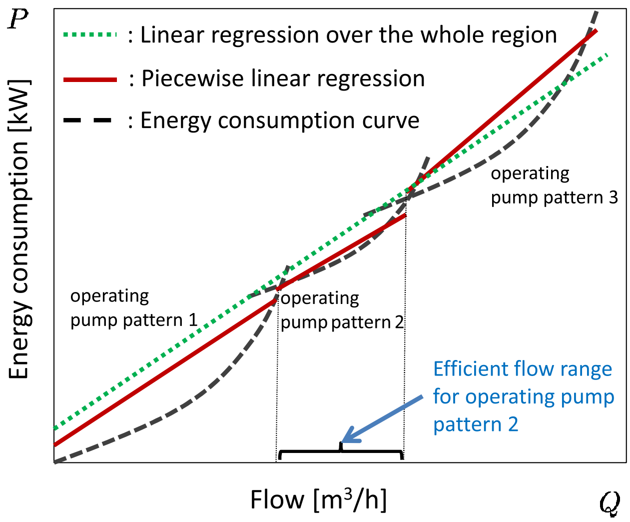

Fig. 3 Conceptual diagram of the energy consumption curve developed by pump modeling (dashed line) and two approximations by the linear regression over the whole flow range (dotted line) and the piecewise linear regression over each of the efficient flow ranges (solid line).

$$P = \frac{\rho H Q}{102\eta},$$

where ρ, H, Q, and η denote the fluid density [kg/m^3], the head [m], the flow [m^3/s], and the efficiency of the pump respectively. We assume that the efficiency η is a concave quadratic function of flow and passes through $(0, 0)$ and $(Q_{rated}, \eta_{rated})$. Q_{rated} and η_{rated} are given by the rated value of each pump.

To find out the operating pump pattern with the lowest energy consumption for different flow ranges, we keep the head as a constant target value. As a consequence, the energy consumption curve, such as the dashed line in **Fig. 3**, is developed for each pump station. These curves correspond reasonably well to actual pump data in the real water distribution network (see Fig. 2).

Finally, as shown in the dotted line and solid line in Fig. 3, an approximation of the energy consumption curve by a linear regression model is created in order to use it in the objective function of our mathematical programming. It is given by $P = \alpha Q + \beta$. Coefficients α and β are calculated using the least-squares method. We separately approximate the energy consumption curve for the daily water production planning and the pump operation planning for the water distribution. In the former case, the energy consumption curve is approximated by the linear regression over the whole flow range, while in the latter case, it is approximated by the piecewise linear regression over each of the efficient flow ranges, where the specific operating pump pattern shows the lowest energy consumption.

6. Solution by Mathematical Programming

In order to reduce energy consumption, we formulate the daily water production planning and the pump operation planning for the water distribution as the mathematical programming where the objective function includes our new energy consumption metric.

In the mathematical programming, a water distribution network is naturally represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where nodes $\mathcal{V}$ stand for purification plants, reservoirs, branch points, and demand points, while edges $\mathcal{E}$ stand for pipes (see Fig. 1). Each node is numbered in accordance with a particular rule, and we obtain $\mathcal{V} = \{1, \ldots, V\}$ where $V = |\mathcal{V}|$ and $|\cdot|$ denotes the cardinality. $\mathcal{V}$ is divided into purification plants set $\mathcal{P}$, reservoirs set $\mathcal{R}$, branch points set $\mathcal{B}$, and demand points set $\mathcal{D}$. $\mathcal{P}$, $\mathcal{R}$, $\mathcal{B}$, and $\mathcal{D}$ are pairwise disjoint, and $\mathcal{V} = \mathcal{P} \cup \mathcal{R} \cup \mathcal{B} \cup \mathcal{D}$.

Table 1 Common notations in optimization problems.

Notation	Definition
$\mathcal{V}$	Set of the V nodes
$\mathcal{E}$	Set of the E pipes
$\mathcal{P}$	Set of purification plants
$\mathcal{R}$	Set of reservoirs
$\mathcal{B}$	Set of branch points
$\mathcal{D}$	Set of demand points
$\mathcal{E}_{k,out}$, $\mathcal{E}_{k,in}$	Set of pipes from or to node k
$\mathcal{E}_{k,n,out}$	Set of pipes from pump station n in node k
N_k	Maximum number of pump stations in node k

Similar to the node case, we number each pipe, and the total count of pipes is $E = |\mathcal{E}|$. Note that we consider only nonnegative flow values, and the flow in the opposite direction is regarded as that through a different pipe. In the actual water distribution network, there is only a single pipe between two nodes, and this formulation is different from a real setting. However, since the optimal solution of our mathematical programming problem leads the flow in either normal or opposite direction to zero by virtue of the objective function including the network cost, the optimal flow is consistent with that through a real pipe.

The common notations that we use for the water distribution network through Section 6 are summarized in **Table 1**. Moreover, since we aim to develop a generic formulation, the mathematical programming optimization is described without the unit for optimization variables and parameters. In implementation, the consistency of the unit needs to be considered.

6.1 Optimization of Daily Water Production Planning by RO

The current methodology by the water supply utility does not necessarily achieve appropriate water supply quantities and constraint conditions. To address these problems, we adopt a mathematical programming approach via RO. Note that the tank and the operating pump pattern are considered only in the formulation of the pump operation planning for the water distribution.

Let q_i $(1 \le i \le E)$ denote the nonnegative flow through the pipe i. Since the flow is limited to the physical size of the pipe, the flow capacity constraint is given by

$$Q_i^L \le q_i \le Q_i^U, \quad 1 \le i \le E, \tag{1}$$

where Q_i^L and Q_i^U are the minimum and the maximum of the allowable flow through pipe i respectively.

Hereafter, $\mathcal{E}_{k,out}$ and $\mathcal{E}_{k,in}$ denote the set of numbers of pipes with the start and end node k $(k \in \mathcal{V})$ respectively. At reservoirs, branch points, and demand points, the flow conservation constraints must hold. For reservoirs and branch points, it is

$$\sum_{i \in \mathcal{E}_{k,in}} q_i - \sum_{i \in \mathcal{E}_{k,out}} q_i = 0, \quad k \in \mathcal{R} \cup \mathcal{B}. \tag{2}$$

Similarly, for demand points, it is expressed by

$$\sum_{i\in\mathcal{E}_{k,in}} q_i - \sum_{i\in\mathcal{E}_{k,out}} q_i = d_k, \quad k \in \mathcal{D}, \tag{3}$$

where d_k is the water demand at the demand point k in a single day.

In practice, since d_k is unknown, we use its predicted value $\hat{d}_k$ obtained from the piecewise sparse linear regression model on the basis of Ref. [5] where we select some features such as forecasted and past temperature, forecasted and past weather status, past water demand values, and holiday status. In accordance with the standard technique of RO (see, e.g., Ref. [3]), for a positive tuning parameter δ, we obtain the new constraint

$$\sum_{i\in\mathcal{E}_{k,in}} q_i - \sum_{i\in\mathcal{E}_{k,out}} q_i \geq \hat{d}_k + \delta^{\frac{1}{2}}\|\sigma_k\|, \quad k \in \mathcal{D}, \tag{4}$$

where σ_k is the column vector of $\Sigma^{1/2}$ corresponding to d_k, Σ is the covariance matrix of prediction errors $d_k - \hat{d}_k$, and $\|\cdot\|$ denotes the Euclidean norm. To set Eq. (4), we estimate Σ from past prediction errors of the piecewise sparse linear regression model by a common unbiased estimator $\hat{\Sigma}$.

Let L_k and U_k ($k \in \mathcal{P}$) denote the minimum and the maximum of the allowable water supply quantities in each purification plant respectively. The constraint on the water supply capability is given by

$$L_k \leq \sum_{i\in\mathcal{E}_{k,out}} q_i - \sum_{i\in\mathcal{E}_{k,in}} q_i \leq U_k, \quad k \in \mathcal{P}. \tag{5}$$

We define our objective function as a linear combination of the energy consumption function approximated by the linear regression model over the whole flow range in Section 5.2, and the network cost for the avoidance of the circumvention of the flow. Our objective function has the form

$$\sum_{k\in\mathcal{P}\cup\mathcal{R}} \sum_{n=1}^{N_k} \sum_{i\in\mathcal{E}_{k,n,out}} (\alpha_i q_i + \beta_i) + \sum_{k\in\mathcal{B}\cup\mathcal{D}} \sum_{i\in\mathcal{E}_{k,out}} W_i q_i, \tag{6}$$

where N_k means the maximum number of pump stations inside the node k, $\mathcal{E}_{k,n,out}$ means the set of numbers of pipes having the pump station n inside the node k ($k \in \mathcal{P}\cup\mathcal{R}$) as the start point, W_i is a positive weight parameter, and α_i and β_i are obtained by the least-squares method over the whole flow range for each pump station (see Fig. 3).

In summary, for the daily water production planning, we formulate the following linear programming (LP)

$$\min_{q_i, 1\leq i\leq E} (6)$$

subject to Eqs. (1), (2), (4), (5).

Finally, the robust daily water supply quantities in each purification plant are calculated by $\sum_{i\in\mathcal{E}_{k,out}} q_i - \sum_{i\in\mathcal{E}_{k,in}} q_i$ ($k \in \mathcal{P}$) from the optimal solution.

6.2 Optimization of the Pump Operation Planning for the Water Distribution by MILP

We introduce the formulation of the pump operation planning for the water distribution via the MILP. In the pump operation planning for the water distribution, a single day is divided into T time intervals, and we need to optimize both the flow and operating pump pattern in each time interval. Motivated by requirements from the water supply utility, this mathematical optimization problem needs to minimize energy consumption under constraints on the behavior of the pump switching.

Note that we have the same water distribution network as the daily water production planning case, and it does not change in time.

Similar to the daily water production planning, let $q_i(t)$ ($1 \leq i \leq E$, $1 \leq t \leq T$) denote the nonnegative flow through the pipe i in the time interval t, and the flow capacity constraint is given by

$$Q_i^L(t) \leq q_i(t) \leq Q_i^U(t), \quad 1 \leq i \leq E,\ 1 \leq t \leq T, \tag{7}$$

where $Q_i^L(t)$ and $Q_i^U(t)$ are the minimum and the maximum of the allowable flow through pipe i in the time interval t respectively.

With respect to the flow conservation constraints at branch and demand points, it follows that

$$\sum_{i\in\mathcal{E}_{k,in}} q_i(t) - \sum_{i\in\mathcal{E}_{k,out}} q_i(t) = 0, \quad k \in \mathcal{B},\ 1 \leq t \leq T, \tag{8}$$

$$\sum_{i\in\mathcal{E}_{k,in}} q_i(t) - \sum_{i\in\mathcal{E}_{k,out}} q_i(t) = \hat{d}_k(t), \quad k \in \mathcal{D},\ 1 \leq t \leq T, \tag{9}$$

where $\mathcal{E}_{k,out}$ and $\mathcal{E}_{k,in}$ are the same notations as those in the daily water production planning case, and $\hat{d}_k(t)$ is the predicted value of the water demand at the demand point k in the time interval t.

The flow constraint at purification plants is given by

$$\sum_{t=1}^{T} \sum_{i\in\mathcal{E}_{k,out}} q_i(t) - \sum_{t=1}^{T} \sum_{i\in\mathcal{E}_{k,in}} q_i(t) \leq S_k, \quad k \in \mathcal{P}, \tag{10}$$

where S_k is water supply quantities of the purification plant k derived from the optimal solution of the daily water production planning in Section 6.1.

To prepare for the contingent increase of water demand, the water supply utility adjusts the water volume in the tank. In each time interval t, the water tank balance constraint is written by

$$V_k^L \leq v_k(0) + \sum_{x=1}^{t} \left\{ - \sum_{i\in E_{k,out}} q_i(x) + \sum_{i\in E_{k,in}} q_i(x) \right\} \leq V_k^U,$$
$$k \in \mathcal{P}\cup\mathcal{R},\ 1 \leq t \leq T, \tag{11}$$

where $v_k(0)$ denotes the initial water volume in the tank, and V_k^L and V_k^U are the minimum and maximum of the tank capacity respectively.

To model constraints on the operating pump pattern, we define an optimization variable $P_{k,n,l}(t) \in \{0, 1\}$ ($k \in \mathcal{P}\cup\mathcal{R}$ $1 \leq n \leq N_k$, $1 \leq l \leq L_{k,n}$, $1 \leq t \leq T$) where n and l mean the pump station number inside the node k and the operating pump pattern of the pump station respectively. $L_{k,n}$ is the maximum number of the possible operating pump pattern at the pump station n inside the node k. If $P_{k,n,l}(t) = 1$, the pump station n inside the node k in the time interval t is operating using the combination of pumps corresponding to the pattern l, otherwise its operating pump pattern is not selected. To select the single pump pattern among all of the operating pump patterns in the optimal solution, we give the following constraints

$$P_{k,n,l}(t) \in \{0,1\},$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le l \le L_{k,n},\ 1 \le t \le T, \quad (12)$$
$$\sum_{l=1}^{L_{k,n}} P_{k,n,l}(t) = 1,$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le t \le T. \quad (13)$$

In this paper, if there exists l such that $P_{k,n,l}(t+1) - P_{k,n,l}(t) \neq 0$ for some k, n, and t, it is regarded as one switch of the operating pump pattern between the time interval t and $t+1$. Switching the operating pump pattern needs to be limited from both physical and operational perspectives. To formulate these limitations, we introduce a new binary optimization variable $P'_{k,n,l}(t)$ ($k \in \mathcal{P} \cup \mathcal{R}$, $1 \le n \le N_k$, $1 \le l \le L_{k,n}$, $1 \le t \le T-1$) which controls the switch of the operating pump pattern between the time intervals t and $t+1$ through the following constraints

$$-P'_{k,n,l}(t) \le P_{k,n,l}(t+1) - P_{k,n,l}(t) \le P'_{k,n,l}(t),$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le l \le L_{k,n},\ 1 \le t \le T-1, \quad (14)$$
$$P'_{k,n,l}(t) \in \{0,1\},$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le l \le L_{k,n},\ 1 \le t \le T-1. \quad (15)$$

As for the physical constraint, once the operating pump pattern changes, the next switch requires more preparation time than one time interval. If this constraint is violated, the risk of mechanical failure increases. This constraint is given by

$$\frac{1}{2} \sum_{l=1}^{L_{k,n}} \sum_{x=0}^{T_{k,n}} P'_{k,n,l}(t+x) \le 1,$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,$$
$$1 \le t \le T-1-T_{k,n},\ 1 \le T_{k,n} \le T-2, \quad (16)$$

where $T_{k,n}+1$ means the number of time intervals required for preparation of the next switch. In the feasible solution satisfying Eq. (16), at most one switch of the operating pump pattern occurs from the time interval t to $t + T_{k,n} + 1$ ($1 \le t \le T-1-T_{k,n}$). In other words, after one switch, the next switch does not occur in at least $T_{k,n}+1$ successive time intervals. Since $P'_{k,n,l}(t) = 1$ and $P'_{k,n,l'}(t) = 1$ if the operating pump pattern l changes to l', Eq. (16) is multiplied by $1/2$.

In the current operation, some operators change the combination of pumps by hand. Consequently, the water supply utility makes the plan for pump operations in such a way as to reduce the count of switching the operating pump pattern as far as possible. To satisfy this operational requirement, switching counts in a single day need to be limited. It can be written by

$$\frac{1}{2} \sum_{l=1}^{L_{k,n}} \sum_{t=1}^{T-1} P'_{k,n,l}(t) \le C_{k,n},$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ C_{k,n} \in Z_+, \quad (17)$$

where $C_{k,n}$ is the upper bound of the count of the pump switching in a single day at the pump station n inside the node k.

Additionally, it is necessary to restrict the fluctuation of the operating pump pattern in one switch because huge change of the combination of pumps causes damage to pump station. Therefore, we need to have

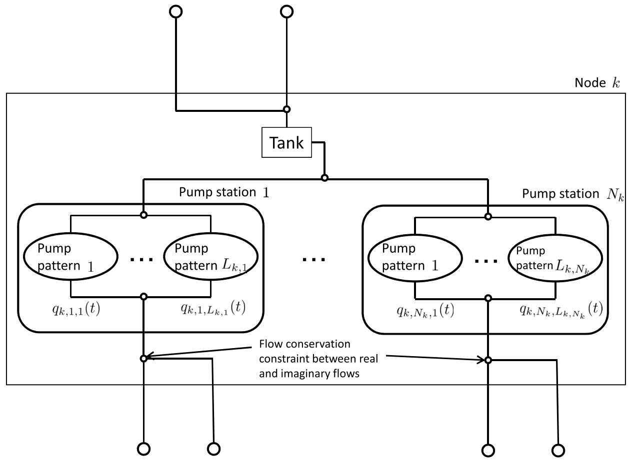

Fig. 4 Overview of the imaginary flow at the node $k \in \mathcal{P} \cup \mathcal{R}$ in the time interval t.

$$-C'_{k,n} \le \sum_{l=1}^{L_{k,n}} l P_{k,n,l}(t+1) - \sum_{l=1}^{L_{k,n}} l P_{k,n,l}(t) \le C'_{k,n},$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le t \le T-1,\ 0 \le C'_{k,n} \le L_{k,n} - 1, \quad (18)$$

where $C'_{k,n}$ is the upper bound of the variation width of the operating pump pattern in one switch.

For the pump behavior, physical and operational constraints such as Eqs. (16)–(18) are not imposed in earlier studies [1], [2], [4], [6], [15]. As a consequence, the operating pump pattern in the optimal solution of these previous works may very often switch.

Next, we introduce an imaginary flow $q_{k,n,l}(t)$ ($k \in \mathcal{P} \cup \mathcal{R}$, $1 \le n \le N_k$, $1 \le l \le L_{k,n}$, $1 \le t \le T$) arising from each operating pump pattern according to **Fig. 4**. The flow range of $q_{k,n,l}(t)$ depends on the combination of pumps, i.e., the operating pump pattern, and is not overlapping except for the end point (see Fig. 3). In addition, if $q_{k,n,l'} > 0$, $q_{k,n,l} = 0$ ($l \neq l'$) because only a single operating pump pattern is realized. Therefore, we impose the following constraint

$$Q^L_{k,n,l} P_{k,n,l}(t) \le q_{k,n,l}(t) \le Q^U_{k,n,l} P_{k,n,l}(t),$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le l \le L_{k,n},\ 1 \le t \le T, \quad (19)$$

where $Q^L_{k,n,l}$ and $Q^U_{k,n,l}$ are obtained from the energy consumption curve. In general, $Q^L_{k,n,1} = 0$ because the pump station often does not provide the flow. As a result, Eq. (19) leads to the linear objective function in this optimization problem. Reference [1] also introduces the flow as in Eq. (19), but it is just two types of efficiency or inefficiency.

As shown in Fig. 4, the relation between the real flow and imaginary one can be expressed as the following flow conservation constraint

$$\sum_{l=1}^{L_{k,n}} q_{k,n,l}(t) = \sum_{i \in \mathcal{E}_{k,n,out}} q_i(t),$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \le n \le N_k,\ 1 \le t \le T. \quad (20)$$

Although the pump station has the ability to put the flow out into the another node, the fluctuation of the flow from the pump station is physically limited. Therefore, we set the following constraint

$$-U_{k,n} \leq \sum_{l=1}^{L_{k,n}} q_{k,n,l}(t+1) - \sum_{l=1}^{L_{k,n}} q_{k,n,l}(t) \leq U_{k,n},$$
$$k \in \mathcal{P} \cup \mathcal{R},\ 1 \leq n \leq N_k,\ 1 \leq t \leq T-1, \qquad (21)$$

where $U_{k,n}$ is the upper bound of the allowable range of the flow variation.

The objective function is expressed as a linear combination of the approximation of the energy consumption curve by the piecewise linear function and the network cost. For a positive weight constant W'_i $(1 \leq i \leq E)$, we obtain the following objective function

$$\sum_{t=1}^{T} \sum_{k \in \mathcal{P} \cup \mathcal{R}} \sum_{n=1}^{N_k} \sum_{l=1}^{L_{k,n}} (\alpha_{k,n,l} q_{k,n,l}(t) + \beta_{k,n,l} P_{k,n,l}(t))$$
$$+ \sum_{t=1}^{T} \sum_{k \in \mathcal{V}} \sum_{i \in \mathcal{E}_{k,out}} W'_i q_i(t), \qquad (22)$$

where coefficients $\alpha_{k,n,l}$ and $\beta_{k,n,l}$ are calculated from the approximation of the energy consumption curve by the linear regression model over the piecewise range disaggregated by the operating pump pattern in each pump station (see Fig. 3).

Finally, we formulate the following optimization problem

$$\min_{\substack{q_i(t), q_{k,n,l}(t), P_{k,n,l}(t), P'_{k,n,l}(t'),\\ 1 \leq i \leq E, 1 \leq t \leq T, 1 \leq t' \leq T-1,\\ k \in \mathcal{P} \cup \mathcal{R}, 1 \leq n \leq N_k, 1 \leq l \leq L_{k,n}}} (22)$$

subject to Eqs. (7)–(21).

It is known as the MILP. The plans for water distribution and pump operations are determined by $q_i(t)$, $q_{k,n,l}(t)$, and $P_{k,n,l}(t)$ in the optimal solution.

7. Experimental Results

In this section, we discuss optimization results for the large-scale real water distribution network in the anonymous city. We demonstrate the daily water production by the RO and both water distribution and pump operations by the MILP as described in Section 6. The number T of time intervals in Section 6.2 is 96, i.e., every 15 minutes, and it starts from 22:00.

Figure 5 shows the maximum values and mean values of water demand values every 15 minutes, and they exhibit two peaks at around 8:00 and 21:00. This is a typical variation pattern in water demand, corresponding to the lifestyle of people in the city. We used water demand values employed in description of Fig. 5 as $\hat{d}_k(t)$ $(k \in \mathcal{D},\ 1 \leq t \leq 96)$ in Eq. (9). $\hat{d}_k = \sum_{t=1}^{96} \hat{d}_k(t)$ $(k \in \mathcal{D})$ was adopted in Eq. (4) of the RO formulation, but $\hat{\Sigma}$ was calculated from past prediction errors of the piecewise sparse linear regression model using the estimation method in Section 6.1.

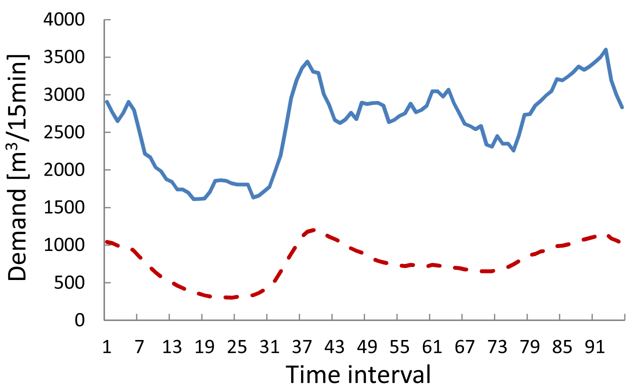

Fig. 5 Maximum values (solid line) and mean values (dashed line) for water demand at all the demand points every 15 minutes from 22:00 in a typical day.

7.1 Daily Water Supply Quantities by RO

The first experiment was the optimization of the daily water production using RO. We evaluate the stability of our RO in terms of the change count of the plan over the course of one year. To evaluate the number of plan modifications in one year, 365 patterns of water demand were generated randomly by adding a prediction error vector following a normal distribution with mean 0 and covariance matrix $\hat{\Sigma}$ to $\hat{d}_k$ $(k \in \mathcal{D})$, and we compared the number of plan modifications between our RO result and the current operation.

Also, our RO determines the optimal margin rate in purification plants to achieve a stable water supply. The margin rate is defined as the excess rate of the determined gross water supply quantity for the total amount of uncertain daily water demand. In order to obtain the reasonable margin rate for RO, we employed Eq. (4) for uncertainty in the predicted water demand. This constraint requires the covariance matrix Σ and the tuning parameter δ. We set $\hat{\Sigma}$ instead of Σ, and individual margin rates were calculated for each of various values of δ using $\hat{d}_k$ $(k \in \mathcal{D})$. For the weight parameter required in RO, we used $W_i = 1$ $(i \in \mathcal{E}_{k,out},\ k \in \mathcal{B} \cup \mathcal{D})$.

Figure 6 shows our numerical result for the number of plan modifications in RO, as a function of the margin rate. The margin rate in the current operation was based on the experience of operators, while that in RO was calculated by holistic optimization on the basis of Eq. (4). This enables RO to achieve lower water production in meeting the specific number of plan modifications. The number of plan modifications in the current operation was more than one hundred in one year, and the margin rate in the day of Fig. 5 was 18.8%. If we consider the goal for plan modification count to be less than 10, the optimal margin rate via our RO was 12.1% at $\delta = 7.03$, while the current operation may require a higher margin rate. Thus, the daily water production planning with RO eliminated the unnecessary water production, while establishing a stable water supply with fewer plan modifications

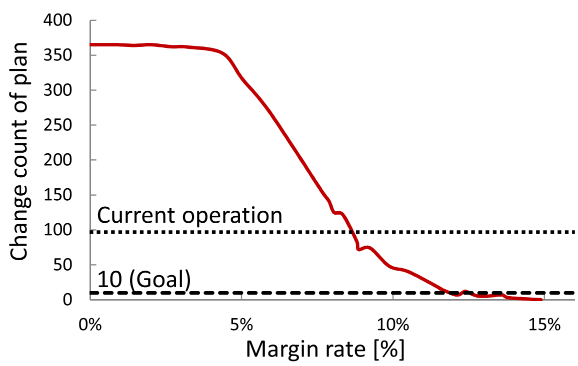

Fig. 6 Relation between the margin rate of gross water supply quantity for the total amount of water demand and the change count of the original plan with RO.

than in the current operation.

7.2 Water Distribution and Pump Operations by MILP

In the second experiment, we demonstrated the effectiveness of the water distribution and pump operations every 15 minutes by the MILP. Our MILP formulation achieves the lowest-energy water distribution solution under two requirements: 1) the stable water supply quantity obtained in Section 7.1 for the water demand in Fig. 5, and 2) pump behavior restricted by the proposed constraints Eqs. (16)–(18).

Below is a detailed configuration for pumps and tanks. To reduce the risk of mechanical failure, we restricted pump switching based on requirements from the water supply utility. For the pump switching interval, we set $T_{k,n}$ in Eq. (16) as 3. In other words, after the change of the operating pump pattern, it cannot be changed again for one hour. For maximum pump switching counts per day, we took $C_{k,n}$ in Eq. (17) to be equal to 4. Also, pumps in the pump stations were turned on or off one by one, which is equivalent to $C'_{k,n} = 1$ in Eq. (18). Note that some pumps, i.e., pumps directly connected to demand points, could not satisfy the above conditions to meet water demands. For these specific pumps, we mitigated the pump switching constraints in Eqs. (16)–(18). For tanks in the purification plants in Eq. (10), we used the water supply quantity S_k ($k \in \mathcal{P}$) calculated by the RO with $\delta = 7.03$. At the reservoirs' tanks, the last water volumes in the 96th time interval were kept within ±5% of the initial water volume. Thus, we substituted $1.05v_k(0)$ and $0.95v_k(0)$ for the upper and lower bound of the water tank balance constraint Eq. (11) in the 96th time interval respectively. W'_i ($1 \leq i \leq E$) was calculated by the ratio of the length of the pipe i divided by its diameter. Finally, we obtained the MILP with 136,604 optimization variables including 55,772 binary variables and 179,511 constraints. This problem was solved in a few minutes using the branch-and-cut algorithm of Gurobi [7].

Figures 7 and **8** display an optimal solution for water distribution and operating pump pattern obtained by the proposed MILP, as compared to the optimal one obtained by the MILP without Eqs. (16)–(18). In contrast to the proposed MILP, the MILP without pump constraints violated requirements of the water supply utility and resulted in large fluctuations in both water distribution and operating pump pattern. **Figures 9–11** show summary statistics for pump operations with/without Eqs. (16)–(18). The solution with pump constraints had 91 fewer pump switching counts than that without constraints. Thus, the constraints of Eqs. (16)–(18) work well and are quite effective for avoiding mechanical failure by pump operations. Furthermore, **Fig. 12** shows that frequent and rapid pump control is not likely to contribute to reductions in energy consumption in this water distribution network. In fact, with the proposed energy consumption metric, energy consumption for the plan via the MILP with or without Eqs. (16)–(18) was 14.0% and 14.2% less, respectively, than that for the actual plan used in the current operation. These optimization results indicate the usefulness of our MILP formulation with proposed pump constraints.

Figure 13 is a histogram of reductions in energy consumption by the MILP at each pump station. While most pump stations

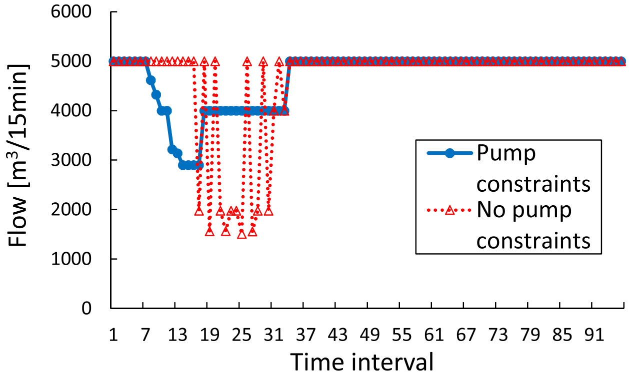

Fig. 7 Example of optimal water distribution from a pump station over 96 time intervals.

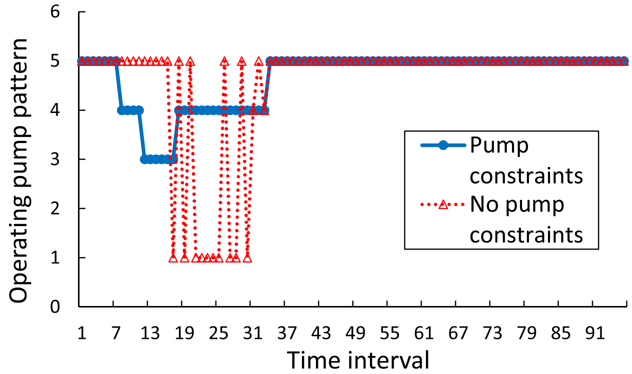

Fig. 8 Example of optimal pump operations at a pump station over 96 time intervals.

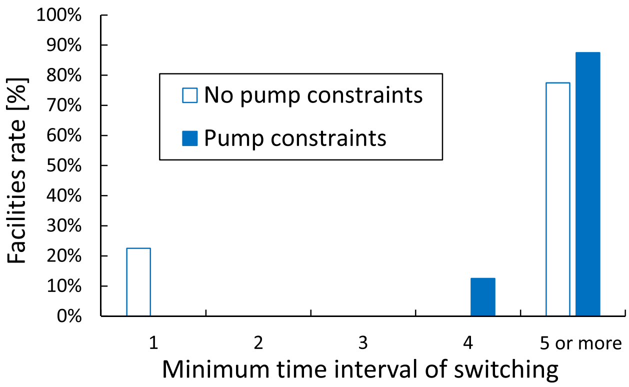

Fig. 9 Comparison of the MILP with and without proposed pump constraints in light of the time intervals for switching.

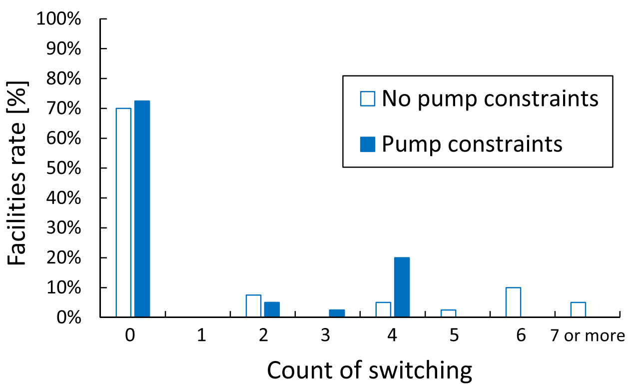

Fig. 10 Comparison of the MILP with and without proposed pump constraints in light of the switching counts.

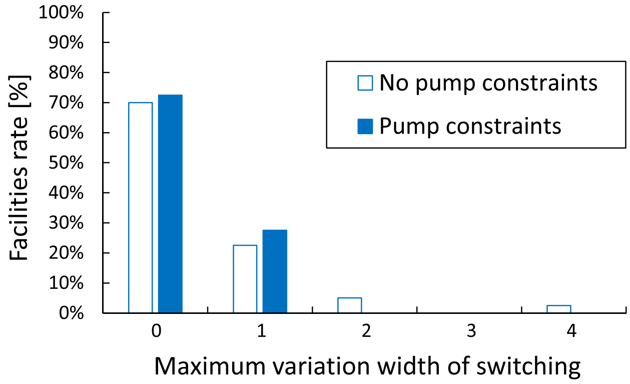

Fig. 11 Comparison of the MILP with and without proposed pump constraints in light of the variation width of switching.

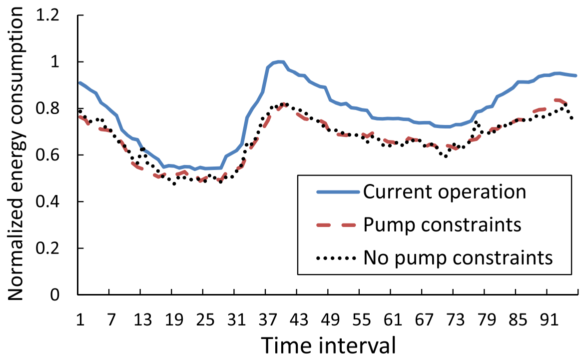

Fig. 12 Comparison of the current operation (solid line), the proposed MILP (dashed line), and the MILP without pump constraints (dotted line) under the proposed energy consumption metric. Energy consumption is normalized so that the upper bound is equal to 1.

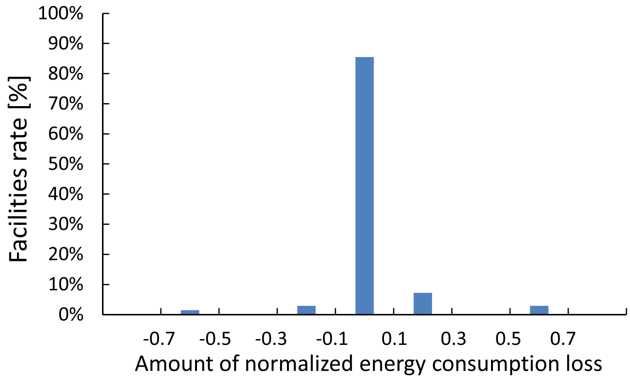

Fig. 13 Histogram of the magnitude of reductions in energy consumption by the MILP in each pump station. Energy consumption is normalized so that the upper bound is equal to 1.

were improved slightly, major contributions were from a small number of pump stations, especially from large pump stations. This may be because it is difficult to make the efficient water distribution plan of large facilities by hand.

Finally, we consider the optimal water volume of tanks in **Fig. 14**. The water volume tends to have peaks shortly before the two peaks in water demand shown in Fig. 5 in order to prepare for the coming high water demand. This behavior is quite natural because our MILP formulation is based on holistic optimization.

8. Conclusion and Future Work

In this paper, we have focused on the industrial application of the mathematical programming technique to the minimization

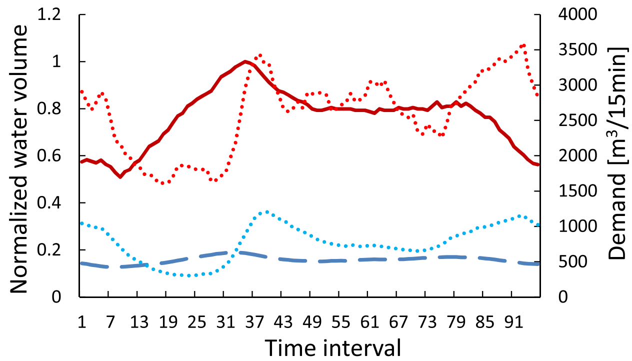

Fig. 14 Maximum values (solid line) and mean values (dashed line) of optimal tank behavior at reservoirs over 96 time intervals. Water volume is normalized so that the upper bound is equal to 1. Maximum values and mean values (dotted line) of water demand shown in Fig. 5 are also plotted.

problem of energy consumption for the large complex water distribution network in the anonymous city. For the case study on the subject of the efficient water management, we have proposed the practical planning method employing the RO and MILP for the water production, the water distribution, and pump operations. The MILP has novel constraints to satisfy physical and operational requirements for the pump behavior. In addition, we have developed the metric for energy consumption based on the physical model, which was applied to the proposed method.

Due to the increasing water demand in rapid urbanization, energy consumption in water supply utilities has become a societal issue, and they have an interest in advanced data analysis techniques for the efficient water management. We believe that our proposed optimization method enables them to reduce energy consumption as well as the number of plan modifications in the daily water production. The experiment with respect to the water distribution network of the actual water supply utility has shown that we were able to reduce energy consumption by 14.0%, which was nearly equivalent to 0.1% of total energy consumption of the city.

The future work is to refine the mathematical formulation so as to describe the water distribution network model more precisely. Firstly, we can take asymmetric switching costs into account. This will be helpful to reduce the risk of machine failure more. Secondly, to determine more efficient water distribution through pipes, it is interesting to consider head loss inside pipes in our mathematical programing.

Acknowledgments The authors are grateful to Yoshio Kameda, Yusuke Muraoka, and Riki Eto for their encouragement and helpful comments. We also acknowledge the suggestions from the editor and two anonymous referees that refined and improved the manuscript.

References

[1] Arai, Y., Nishie, K., Koizumi, A., Inakazu, T., Ishida, N., Yamazaki, C. and Moriyasu, J.: Application of a milp model to a large-scale water distribution system with the objective of reducing electric power usage, *Journal of Japan Society of Civil Engineers, Ser. G* (*Environmental Research*), Vol.69, pp.II_149–II_156 (2013).

[2] Bagirov, A., Barton, A., Mala-Jetmarova, H., Nuaimat, A.A., Ahmed, S., Sultanove, N. and Yearwood, J.: An algorithm for minimization of pumping costs in water distribution systems using a novel approach

to pump scheduling, *Mathematical and Computer Modelling*, Vol.57, pp.873–886 (2013).
[3] Bertsimas, D., Brown, D.B. and Caramanis, C.: Theory and applications of robust optimization, *SIAM Review*, Vol.53, pp.464–501 (2011).
[4] D'Ambrosioa, C., Lodi, A., Wieseb, S. and Bragallic, C.: Mathematical programming techniques in water network optimization, *European Journal of Operational Research*, Vol.243, pp.774–788 (2015).
[5] Eto, R., Fujimaki, R., Morinaga, S. and Tamano, H.: Fully-automatic bayesian piecewise sparse linear models, *Proc. 17th International Conference on Artificial Intelligence and Statistics*, pp.238–246 (2014).
[6] Fooladivanda, D. and Taylor, J.A.: Optimal pump scheduling and water flow in water distribution networks, *IEEE 54th Annual Conference on Decision and Control*, pp.5265–5271 (2015).
[7] Gurobi Optimization, Inc.: *Gurobi optimizer reference manual* (2014).
[8] Kermany, E., Mazzawi, H., Baras, D., Naveh, Y. and Michaelis, H.: Analysis of advanced meter infrastructure data of water consumption in apartment buildings, *Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1159–1167 (2013).
[9] Lopez-Ibanez, M., Prasad, T.D. and Paechter, B.: Ant colony optimization for optimal control of pumps in water distribution networks, *Journal of Water Resources Planning and Management*, Vol.134, pp.337–346 (2008).
[10] McCormick, G. and Powell, R.S.: Derivation of near-optimal pump schedules for water distribution by simulated annealing, *Journal of the Operational Research Society*, Vol.55, No.7, pp.728–736 (2004).
[11] Mevissen, M., Ragnoli, E. and Yu, J.Y.: Data-driven distributionally robust polynomial optimization, *Advances in Neural Information Processing Systems 26*, pp.37–45 (2013).
[12] Mohammed, J.R. and Ibrahim, H.M.: Hybrid wavelet artificial neural network model for municipal water demand forecasting, *ARPN Journal of Engineering and Applied Sciences*, Vol.7, No.8, pp.1047–1065 (2012).
[13] Morsi, A., Geißler, B. and Martin, A.: *Mixed integer optimization of water supply networks*, chapter 3, pp.35–54, Springer, Basel (2012).
[14] Tachibana, Y. and Ohnari, M.: Prediction model of hourly water consumption in water purification plant through categorical approach, *Systems, Man, and Cybernetics, 1999*, pp.569–574 (1999).
[15] Takahashi, S., Koibuchi, H., Adachi, S., Takemoto, T. and Koizumi, K.: Pump operation scheduling for power demand response in water transmission systems, *IEEJ Trans. Electronics, Information and Systems*, Vol.8, pp.1200–1208 (2016).
[16] Ulanicki, B., Kahler, J. and See, H.: Dynamic optimization approach for solving an optimal scheduling problem in water distribution systems, *Journal of Water Resources Planning and Management*, Vol.133, pp.23–32 (2007).

Toshihiro Hirano received B.E. degree from Tokyo Institute of Technology in 2007, and the M.E. degree from the University of Tokyo in 2009. From April 2014 to March 2017, he worked at Data Science Research Laboratories, NEC as researcher. Since April 2017, he is a lecturer at College of Economics, Department of Economics, Kanto Gakuin University.

Yuta Ashida received B.E. degree from Kobe University, Japan in 2007, and received M.E. degree from Graduate School of Engineering, Kobe University, Japan in 2009. Since April 2009, he is working at Data Science Research Laboratories, NEC as researcher.

Lu Yu received B.E. degree from Beijing Institute of Technology, China in 2005; and received M.E. and Ph.D. degrees from Waseda University, Japan in 2007 and 2010, respectively. From September 2010 to March 2011, she was a research fellow of Japan Society for the Promotion of Science. From January 2012 to January 2017, she worked at NEC Labs China as researcher. Since February 2017, she is a game designer at Otakugame Ltd.

Sayaka Izukura received B.E. and M.E. degrees from Sophia University in 2006 and 2008, respectively. Since April 2008, she is working at Data Science Research Laboratories, NEC as researcher.

Itaru Nishioka received B.E. and M.E. degrees from Osaka University. Since joined NEC in 2000, he was engaged as a researcher of network control/management system, embedded control software, and multi-layer resource planning algorithm. From 2012 to 2014, he was a visiting researcher at Raytheon BBN technologies, where he conducted R&D on SDN/NFV. After returned NEC, he works at several applied machine learning fields as a principal researcher.

Ryohei Fujimaki (Ph.D.) is a research fellow at NEC's Data Science Research Laboratories. He is in charge of developing NEC's advanced analytics technologies and their applications. In addition to his industrial contributions, he has been publishing high-quality academic papers in such premier conferences as KDD, NIPS, ICML, etc.