Supporting Information for

Marginal energy intensity of water supply

Yang Liu^a, Meagan Mauter^{a,*}

^aDepartment of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305, USA.

*Corresponding author. E-mail: mauter@stanford.edu

Supporting Information Summary:

This document contains descriptions of 1) Comparison of the proposed computational framework that calculates marginal energy intensity (MEI) with previous methods that compute high-resolution energy intensity of water supply systems (WSS); 2) details of the implemented genetic algorithm for optimizing the pumping schedule; 3) details of the three-source water distribution network; 4) electricity price patterns used in the study; 5) hourly MEI values in the sensitivity analysis; 6) perturbation analysis of MEI; 7) Hazen-Williams formula; 8) details of the c-town water distribution network; 9) Adaptation of the backtracking algorithm for computing marginal electricity cost of water supply

This supporting information is 15 pages long and contains eight figures and two tables.

Supplementary Note 1. Comparison to previous methods that compute highresolution energy intensity of water supply systems (WSS)

Compared to the method proposed in previous work,¹ which estimated the energy intensity of water supply at the resolution of individual pressure zones, our computational framework improves the resolution to individual consumers and features a rigorous modeling of tanks. Such explicit modeling of tanks is absent in the previous work,¹ Similar to batteries in a power supply system, tanks in a WSS can temporally decouple supply and demand, which complicates the backtracking of water flows and the computation of flow path-specific energy intensity of delivered water. In order to more accurately account for the embedded energy of water discharged from tanks to consumers, an alternative approach is to treat energy as a conservative property (e.g., concentration of NaCl) and uses EPANET's function that simulates the movement of chemicals to obtain the 'concentration of energy' (energy intensity of water supply) for each consumer². Despite the high resolution of the results achieved by this EPANET-based method, treating energy as a conservative property requires the user of the method to specify a constant of a timeseries of pre-determined energy intensity of each pump before running the simulation, which ignores the fact that the energy intensity of a pump varies with its operating status (e.g., flow rate, efficiency). Assuming the inherently time-varying operating status of a pump to be static or known can result in a large distortion of the results. In contrast, MEI calculated by our computational framework reflects the instantaneous operating status of each pump and incorporates the frictional and minor losses, which are infeasible to include using the method that treats energy as a conservative property.

Supplementary Note 2. Genetic algorithm

To solve for the energy cost-minimizing pumping schedules, we implement a genetic algorithm (GA) that is adapted from methods described in previous studies.^{3–8} GA is a widely used evolutionary algorithm that incrementally improves the solution through two mechanisms – crossover and mutation. Crossover takes fractions (i.e., values of a subset of decision variables) from two or more different solutions and merge the fractions into a new solution. We call a new solution generated by crossover an offspring. To avoid premature convergence to a sub-optimal solution, GA usually uses mutation to randomly modify newly generated offspring. To make sure that the best solution does not get worse from generation to generation over to the next generation. Unlike mathematical programming-based methods, GA cannot rigorously enforce all constraints. As a result, most GA implementations, including ours, account for the constraints as penalty functions and add the values of such penalty functions to the objective function, which suggest that both their objective function and penalty function values tend to be low. Details about our implementation of GA in this study are discussed in the following sub-sections.

Fitness function

The fitness function value F is the sum of four components – normalized electricity cost (C_{elec}), penalty for low tank levels (P_{tank}), penalty for detection of under-pressurized water delivery ($P_{pressure}$), and penalty for large deviation from the target fraction of water supply by each source ($P_{fraction}$). Given a timeseries of electricity prices, we first compute the total electricity cost by multiplying the hourly electricity consumption E_t by the hourly electricity price c_t . To avoid poor scaling under different electricity prices, we normalize the total electricity cost by dividing it by the daily average electricity price. Subscript t in eqn (S2) represents time. Similarly, the penalty

function for low tank levels at the end of the day are also normalized. In the numerator of eqn (S3), $H_{T,f}$ and $H_{T,i}$ are final and initial water level of tank T, respectively. In the denominator, $H_{T,max}$ and $H_{T,min}$ bound the entire range of tank level fluctuation. eqn (S4) is the penalty function that penalizes delivered water with pressure lower than the required minimum pressure p_{min} (i.e., 20 psi/14.06 m). p_{low} is the lowest pressure of delivered water across all consumers throughout the day. In eqn (S5), $f_{target,i}$ and $f_{actual,i}$ are the target and actual fraction of water injection into the WDN through injection point *i*. $I_{Threshold,i}$ is an indicator function whose value is 1 when the difference between $f_{target,i}$ and $f_{actual,i}$ are greater than 0.02 and assumes a value of 0 otherwise. For example, if we plan to let injection point 1 account for 30% of the daily water injection (i.e., $f_{target,1} = 0.3$) but our pumping schedule results in a $f_{actual,i}$ are not constants with known values but can be retrieved from the results of the hydraulic simulation of a given pumping schedule.

$$F = C_{elec} + P_{tank} + P_{pressure} + P_{fraction}$$
(S1)

$$\forall t \qquad \qquad C_{elec} = \frac{24}{\sum_t c_t} \cdot \sum_t E_t \cdot c_t \qquad (S2)$$

$$\forall T \qquad P_{tank} = \sum_{T} \left(\left(\frac{H_{T,f} - H_{T,i}}{H_{T,max} - H_{T,min}} + 0.2 \right) \cdot 100 \right)^2 \tag{S3}$$

$$P_{pressure} = (p_{min} - p_{low})^2 \cdot 10 \tag{S4}$$

$$P_{fraction} = \sum_{i} (f_{actual,i} - f_{target,i})^2 \cdot I_{Threshold,i} \cdot 250000$$
(S5)

Generate the initial population

Our GA has 500 generations and each generation consists of 500 feasible solutions (i.e., pumping schedules). Since all pumps are fixed-speed in our case study, their status can be represented as 0-1 integers -0 indicates 'shut down' and 1 indicates 'operating'. As a result, each solution consists of 11 strings of 0-1 integers and each string has 24 integers. In other words, each solution can be represented as a 11 by 24 matrix consisted of 0s and 1s. To generate the initial population, we randomly and repeatedly generate such 0-1 matrices until we obtain 500 feasible pumping schedules, which result in successful hydraulic simulations by EPANET.

Crossover

After the first generation, each new generation goes through a crossover stage. Out of the 500 new solutions, five are directly carried over from the best five solutions in the previous generation, the remaining 495 new solutions are all generated through crossover. After ranking the 500 solutions in the previous generation by fitness function value, we randomly select pairs of 'parent solutions' from the best 250 solutions (parent pool) and have each pair generate an offspring through a two-point crossover ⁷. For each crossover, the crossover points are randomly chosen but are the same for both parent solutions. We implement a roulette wheel selection here – for solution *j* in the parent pool, its probability of being selected as a parent solution is set to be proportional to $(500 - rank_i)$, where $rank_i$ is the ranking of solution *j*.⁹ If the crossover of a pair of randomly selected

parent solutions generates an infeasible pumping schedule that leads to an unsuccessful hydraulic simulation, we re-select a pair of parents until a feasible new pumping schedule is obtained. It is worth noting that each solution can be selected as a parent multiple times and there is no guarantee that every solution in the parent pool will be selected at once.

Mutation

Once a feasible new pumping schedule is generated through crossover, there is an 80% probability that the pumping schedule is selected to undergo a mutation. For each mutation, 1-4 pumps are randomly chosen first. Afterwards, pump status over 1-6 randomly chosen and adjacent time steps (i.e., hours) are changed to either 0s or 1s for each pump selected to undergo mutation. The number of pumps to choose, the number of time steps to choose, and the pump status to change to are all random variables with uniform probability distribution.

Parallel computing

To accelerate the GA, we parallelize the crossover and mutation mechanisms. Instead of generating 495 new solutions one by one, we launch separate and parallel processes that each works towards finding a new feasible solution through crossover and mutation. After this parallelization, hydraulic simulations of different pumping schedules, which are the most time-consuming computational task in the GA, can be completed by parallel CPUs.

Consistency of the algorithm

As a heuristic method, GA cannot solve a problem to a given optimality gap. Instead, GA converges to a near-optimal solution efficiently. Therefore, it is important to check the reliability of the implemented GA and make sure that it produces consistent solutions given the same optimization inputs. Such consistence is particularly critical as we are using the GA to compute energy cost-minimizing pumping schedules for our sensitivity analysis. To test the consistency of the GA, we run the algorithm five times for the base case and compare the results. As Figure S1 shows, the five optimized pumping schedules are very similar in their pattern and they result in very close daily energy costs, which are \$56.8, \$57.6, \$57.8, \$57.7, \$57.8.



Figure S1. Comparison of five pumping schedules optimized by the implemented genetic algorithm under the base case scenario.

Supplementary Note 3. Details of the three-source water distribution network

The three-source WDN shown in Fig. 2a is adapted from a real-world system in Kentucky. The real-world WDN has 287,609 ft (87.7 km) of pipe and provides 1.04 million gallons (4581 m³) of water per day to 2682 consumers. The skeletonized network maintains the overall configuration of the WDN but merges some nodes and pipes, which results in 331 consumers (one consumer is converted to a non-consumer junction as shown in Table S1).

As described in the main manuscript, we manipulate the locations of injection points I1, I2, and I3 from their true locations to improve their visualization in Fig. 2a. The true locations of the three injection points relative to the rest of the WDN are shown in Figure S2. As Figure S2 shows, I1 is closer to its downstream consumers than shown in Fig. 2a. On the other end of the WDN, I3 is much farther away from its downstream consumers than shown in Fig. 2a. For all the calculations in this study, we use the true locations of the injection points as well as their downstream pumps, pipes, and valves.

We make several small modifications to the nomenclature of the original network. The original WDN is downloaded as an EPANET file.¹⁰ In the original WDN, there are four reservoirs named R-1, R-2, R-3, and WTP. All reservoirs have known water levels and elevations. From the given information in the original EPANET file, it is obvious that R-2, R-3 are reservoirs storing water that requires no extra treatment for potable use, and WTP treats and stores water from its upstream reservoir R-1. We remove all the hydraulic components upstream WTP, including R-1, and rename WTP, R2, and R3 as I1, I2, and I3, respectively. Pumps originally named Pump 1, Pump 2, Pump 3, and Pump 4 are renamed as P2, P3, P1, and P4, respectively.

In addition to the removal of R-1 and the renaming of components, five additional modifications to the EPANET file are made for the base case scenario (shown in Table S1). In the original EPANET file, only a rated power is provided for each pump and all pumps are assumed to operate at a constant mechanical efficiency of 75%, which is unrealistic. To mimic more realistic operations of pumps, we use a previously proposed method to re-construct reasonable pump curves for the pumps.¹¹ Such curves include two sets – one set is head-flow (H - Q) curves and the other set is efficiency-flow $(\eta - Q)$ curves. According to past work by Kuritza et al.,¹¹ the two curves for each pump can be estimated from the best-efficiency point (BEP), which is a pair of operating head and flow rate values. We run 1000 feasible one-time-step hydraulic simulations for the WDN under randomized conditions (e.g., initial tank levels, water demand) and use the mean head and flow rate values as its BEP. In the original EPANET file, P2 only has a rated power of 3 horsepower, which makes it impossible to withdraw large amounts of water from I2. We change the rated power of P2 to 150 horsepower before running the 1000 simulations. The re-constructed pump curves are shown in Figure S3 below. Since the re-construction of H - Q and $\eta - Q$ curves use the same two functions whose independent variables are both the ratio of flow rate to the BEP, the curves shown in Figure S3 share the same shape but not the same range of values. Also, it is worth noting that pumps P3 and P4, which are connected in series, have very similar flow rates at their respective BEPs.

Table S1. Modifications to the EPANET file of the three-source WDN

Modification 1	Increase the water demand of all consumers at all times by a factor of four
Modification 1	Make junction J-106 a non-consumer junction

Modification 2	Move the start time of the time horizon (for hydraulic simulation) from 12 am to 6
	am
Modification 3	Increase the elevation of I1, I2, I3 and tanks T1, T2, T3 by 30 ft (9.14 m)
Modification 4	Re-construct head-flow and efficiency-flow curves for pumps P1, P2, P3, and P4



Figure S2. Locations of injection points I1, I2, and I3 relative to the rest of the WDN. In the case study, all head losses in pipes are computed using the true pipe lengths that are proportional to the black lines in this figure.



Figure S3. Re-constructed pump curves. Head-flow curve and efficiency-flow curve for each pump in the three-source WDN. The curves are re-constructed from the best-efficiency points, which are obtained through hydraulic simulations with randomized conditions.



Supplementary Note 4. Electricity prices used in the case study

Figure S4. Electricity prices used in the case study. EP0, EP1, and EP2 are averaged to hourly resolution from 15-minute-resolution data downloaded from online databases.^{12–14}EP0 is a 'duck curve' in which the electricity price reaches the lowest point in the afternoon when the power generation from solar farms are at its peak. EP1 is a relatively flat price timeseries. EP2 is a single-peak electricity price timeseries in which the electricity price peaks around evening.



Supplementary Note 5. Hourly MEI values in the sensitivity analysis

Figure S5. Hourly MEI of the 330 consumers in the three-source WDN in the eight scenarios of the sensitivity analysis. 'D-' and 'D+' are scenarios where the hourly water consumption rates (i.e., demand) of all consumer nodes are decreased or increased by 25%, respectively. 'EP1' and 'EP2' are scenarios where alternative electricity price profiles are used to optimize the pumping schedule. 'R-' and 'R+' are scenarios where the roughness of all pipes is decreased or increased by 50%, respectively. 'I3-' and 'I3+' are scenarios where the target percentage of daily water injection through I3 is adjusted from 50% in the base case to 30% or 70%, respectively.

Supplementary Note 6. Perturbation analysis of MEI

Since the water consumption rate of a single consumer is marginal (i.e., negligible) compared to the total injection rate through an injection point or flow rate in a water main, MEI is inherently insensitive to perturbations in the water consumption behavior of a single or a small number of consumer nodes. Such insensitiveness, or marginality of MEI, is demonstrated in Figure S6. In this perturbation analysis, we randomly pick five consumers from the three-source WDN and vary their hourly water consumption rate by 10%, 20%, and 40% respectively. At each level of perturbation, the consumption rate of each perturbed consumer node could either increase or decrease, with equal probabilities, by the specified percentage. We repeat the perturbation at each level ten times, which generates 50 lines per level in Figure S6. Since such perturbations have negligible effect on the overall water demand in the WDN, the base case pumping schedule remains feasible and is used to simulate water flows and calculate MEI values.

As the deviations in hourly MEI values suggest, even if the five consumer nodes' water consumption rates are perturbed by 40%, their hourly MEI values remain virtually unchanged compared to the unperturbed condition (Fig. 3c). The maximum deviation from the base case MEI values among the 150 lines in Figure S6 is merely 0.077 kWh/m³, which corresponds to a 5.7% increase. However, it is worth noting that larger perturbation in the water consumption behavior does lead to larger deviation in MEI.

Potential users of MEI and the computational framework must pay close attention to the marginality of MEI. On one hand, for example, the small deviations shown in Figure S6 prove that the energy saving potential of a water-saving toilet can be directly computed by multiplying the water demand per flush by the MEI at the location. On the other hand, one should be careful when doing such multiplication over a large number of water consumers whose water consumption behaviors significantly deviate from the base case scenario where the MEI values are calculated. Once the change in water supply or demand is no longer marginal, the MEI values calculated before the change become inaccurate and need to be re-calculated.



Figure S6. Results of the perturbation analysis. Hourly MEI values of perturbed consumer nodes whose hourly water consumption rates are perturbed by 10%, 20%, and 40%. **a.** Deviation in hourly MEI values after perturbation measured in kWh/m³. **b.** Deviation in hourly MEI values after perturbation measured in %.

Supplementary Note 7. Hazen-Williams formula

The Hazen-Williams formula is a classical formula that correlates the head loss with the length, diameter, roughness of a pipe and the flow rate in the pipe. In eqn (S6), h is the head loss (m), L is the pipe length (m), Q is the flow rate (m³/s), and d is the pipe diameter (m). C is the roughness coefficient, which is inversely proportional to the roughness of a pipe.

$$h = \frac{10.67 \cdot L \cdot Q^{1.852}}{C^{1.852} \cdot d^{4.87}} \tag{S6}$$

As eqn (S6) suggests, longer, thinner, and rougher pipes result in larger head losses. Therefore, by increasing the pipe roughness in the sensitivity analysis, we essentially achieve the same effect as stretching the pipes and expanding the size of the WDN. In the sensitivity analysis in the main manuscript, as we increase the roughness of each pipe by 50% (i.e., roughness coefficient decreases by 33.3%), the head loss increases by 112% with all else held equal.

Supplementary Note 8. Details of the c-town water distribution network

Similar to the three-source WDN, the c-town WDN (Fig. 4a) is also adapted from an EPANET file with several modifications, which are listed in Table S2. This WDN is widely used as a benchmark system for studies that aim to optimize the design and operation of water distribution systems.

The original EPANET file comes with head-flow curves for the pumps but assumes a constant mechanical efficiency. As a result, we follow the same procedure applied to the three-source WDN to re-construct the efficiency-flow curves. Due to the similarity in both the procedure and the shapes of the resulting curves, we do not visualize each of the re-constructed curves for all eleven pumps. Using the GA described above, we optimize the pumping schedule for this WDN and plot the hourly pumping load, hourly water injection through the injection point and tanks, and hourly MEI values in Figure S7.

Modification 1	Move the start time of the time horizon (for hydraulic simulation) from 12 am to 6
	am
Modification 2	Increase the elevation of the only injection point and the 7 tanks by 30 ft (9.14 m)
Modification 3	Re-construct efficiency-flow curves for the eleven pumps

Table S2. Modifications to the EPANET file of the c-town WDN



Figure S7. Pumping schedule and hourly MEI values of the c-town WDN. a. Hourly load profile of the optimal pumping schedule that minimizes the electricity cost under electricity price profile EP0. Primary pumps are the three parallel pumps that are immediately downstream the injection point. Booster pumps are the remaining eight pumps. b. Hourly water injection through the injection point and the net water flows into tanks and consumers. 'Inj' is the abbreviation for 'injection point'. **c.** Hourly MEI values of 334 consumer nodes.

Supplementary Note 9. Adaptation of the backtracking algorithm for computing marginal electricity cost of water supply

As discussed in the manuscript, the backtracking algorithm can be applied to trace other inputs into water flows. For instance, the computing marginal operating cost of water supply is determined by the locations where operating costs are proportional to the volume of water flow. Such points include any 'energy input' point that incurs an energy cost, any 'chemical input' point (e.g., treatment plant) that incurs a chemical cost, and (or) any raw water source whose effluent must be purchased by the water utility. Here, we focus on the electricity component of the marginal operating cost and demonstrate the derivation of marginal electricity cost (MEC) of water supply.

In general, MEC of water supply is equal to the product of MEI (kWh/m³) and marginal electricity price c_t (\$/kWh). Such straightforward conversion is indicated in eqn (S7)-(S9) below. However, since energy dissipated in pipes and valves do not need to be paid, we must exclude the energy dissipation component in the previously calculated MEI_{*i*-*j*,*t*} values. In addition, when a consumer is receiving a fraction or all of her water from a tank, it is incorrect to directly multiply MEI with electricity price because a fraction of the MEI is the pre-injection energy intensity of the tank (MEI_n^{Pre-inj}), which is a constant that equals to the daily average energy intensity of delivering water into the tank (eqn (7)). In other words, multiplying the MEI_n^{Pre-inj} value of a tank by the real-time electricity price would be mistakenly charging electricity consumed during other time steps at the price of the current time step. Instead, as eqn (S10) shows, the pre-injection component of MEC for a tank discharging water should be calculated by averaging the electricity costs of delivering water into the tank throughout the day.

$$MEC_{j,t} = \sum_{i} r_{i-j,t} \cdot (MEC_{i,t}^{Pre-inj} + MEC_{i-j,t}^{Dist})$$
(S7)

$$MEC_{i-j,t}^{Dist} = MEI_{i-j,t}^{Dist} \cdot c_t$$
(S8)

$$i \in R$$
 $MEC_{i,t}^{Pre-inj} = (MEI_i^{Trans} + MEI_i^{Treat}) \cdot c_t$ (S9)

$$n \in T \\ i \in (T \cup R) \setminus n \qquad \qquad \mathsf{MEC}_n^{\mathsf{Pre-inj}} = \sum_t \frac{Q_{n,t}}{Q_n} \sum_i r_{i-n,t} \cdot (\mathsf{MEI}_i^{\mathsf{Pre-inj}} + \mathsf{MEI}_{i-n,t}^{\mathsf{Dist}}) \cdot c_t \tag{S10}$$

In Figure S8 below, the timeseries of MEC values of each consumer over the 24-hour period in the base case scenario is shown as a separate black line (Figure S8c). Overall, the MEC values vary between 0 and 0.03 ^{m^3} . As Fig. 3a and Fig. 3b in the manuscript show, the pumping load (i.e., electricity consumption for water distribution) is non-zero during 10:00-18:00 and 23:00-5:00, implying that most consumers directly receive water from injection points I1-I3 and tanks are being filled. As a result, the segments of MEC curves corresponding to these two periods in Fig. S8c take shapes that are the products of electricity prices and MEI values. During the remaining periods, however, consumers receive water from tanks, resulting in relatively stable MEC values between 0.01 and 0.02 \$/m^3. It is worth noting that, while the electricity price peaks around 19:00, the MEC values do not peak at the same time because consumers are receiving water that was pumped into the tanks during hours of much cheaper electricity.



Figure S8. Derivation of marginal electricity cost of water supply. (a) hourly electricity price. (b) hourly MEI of consumers (excluding dissipated energy). (c) hourly marginal electricity cost of water supply.

References

- 1. Spang ES, Holguin AJ. WEMap : A Software Tool for Estimating Water Utility Energy Savings from Water Conservation. *9th Int Congr Environ Model Softw*, https://scholarsarchive.byu.edu/iemssconference/2018/Stream-B/42/ (2018).
- 2. Sowby RB, Burian SJ. High-Resolution Energy Intensity Modeling to Improve Water Distribution System Performance. *J Sustain Water Built Environ* 2020; 6: 04019009.
- 3. Savic DA, Walters GA, Schwab M. Multiobjective genetic algorithms for pump scheduling in water supply. In: Corne D, Shapiro JL (eds) *Evolutionary Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 227–235.
- 4. Van Zyl JE, Savic D a, Walters G a. Operational optimization of water distribution systems using a hybrid genetic algorithm. *J Water Resour Plan Manag* 2004; 130: 160–170.
- 5. Mackle G, Savic DA, Walters GA. Application of genetic algorithms to pump scheduling for water supply. *IEE Conf Publ* 1995; 400–405.
- 6. Barán B, Von Lücken C, Sotelo A. Multi-objective pump scheduling optimisation using evolutionary strategies. *Adv Eng Softw* 2005; 36: 39–47.
- 7. Abkenar SMS, Stanley SD, Miller CJ, et al. Evaluation of genetic algorithms using discrete and continuous methods for pump optimization of water distribution systems. *Sustain Comput Informatics Syst* 2015; 8: 18–23.
- 8. Odan FK, Ribeiro Reis LF, Kapelan Z. Real-Time Multiobjective Optimization of Operation of Water Supply Systems. *J Water Resour Plan Manag* 2015; 141: 04015011.
- 9. Wang JY, Chang TP, Chen JS. An enhanced genetic algorithm for bi-objective pump scheduling in water supply. *Expert Syst Appl* 2009; 36: 10249–10258.
- 10. Kentucky Water Resources Research Institute. EPANET file of KY14, http://www.uky.edu/WDST/KYEPAzip/ky14 EPANET.zip (2014, accessed 4 May 2020).
- Kuritza JC, Camponogara G, Marques MG, et al. Dimensionless curves of centrifugal pumps for water supply systems: development and case study. *Brazilian J Water Resour*; 22. DOI: 10.1590/2318-0331.0217170018.
- 12. CAISO. CAISO real-time price of node 0096WD_7_N001, Apr 15th, 2019, http://oasis.caiso.com/mrioasis/logon.do (2019, accessed 7 May 2020).
- energyonline.com. ERCOT real-time price of Houston load zone, Jul 23rd, 2019, http://www.energyonline.com/Data/GenericData.aspx?DataId=4 (2019, accessed 24 July 2019).
- 14. CAISO. CAISO real-time price of node 0096WD_7_N001, Jul 15th, 2019, http://oasis.caiso.com/mrioasis/logon.do (2019, accessed 7 May 2020).