SSUWM 2016, Monte Veritá, Switzerland 23 Aug 2016

A Framework for Real-Time Spatially Distributed Demand Estimation and Forecasting

Paulo José Oliveira[a], S. M. Masud Rana[a], Tian Qin[a],
Hyoungmin Woo[a], Jinduan Chen[b] and Dominic L. Boccelli[a]

[a]University of Cincinnati, Cincinnati OH, USA
[b]IDModeling, Arcadia, CA, USA

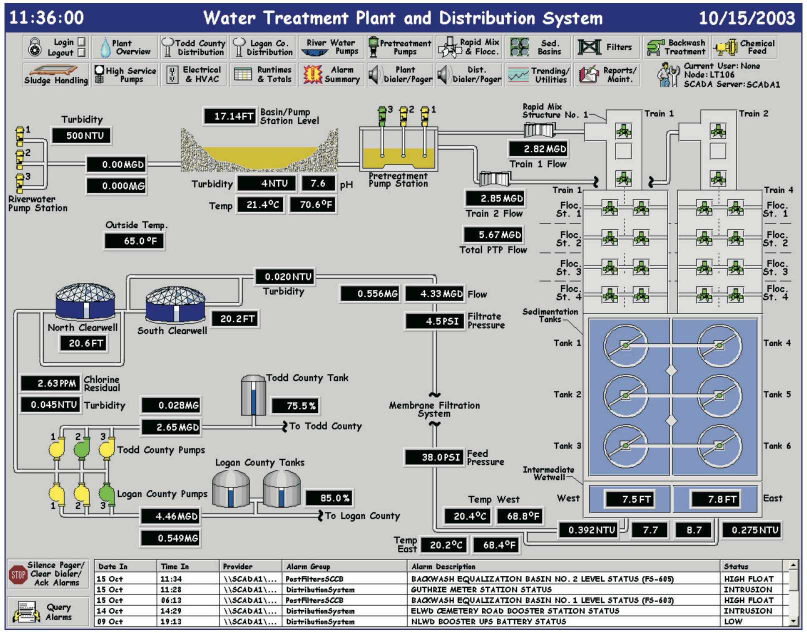
Introduction

- Water utilities must ensure potable water infrastructure are sustainable, robust and resilient to long- and short-term challenges
- Long-term challenges include
  - Climate change
  - Population shifts
  - Aging infrastructure
- Addressed through infrastructure design

http://trenchlessonline.com/new-south-carolina-water-main-provides-for-future-needs/

Introduction

- Water utilities must ensure potable water infrastructure are sustainable, robust and resilient to long- and short-term challenges
- Short-term challenges include
  - Energy management
  - Water quality maintenance
  - Response to (un)intentional intrusion events
  - Leak detection
- Addressed through real-time monitoring and decision support

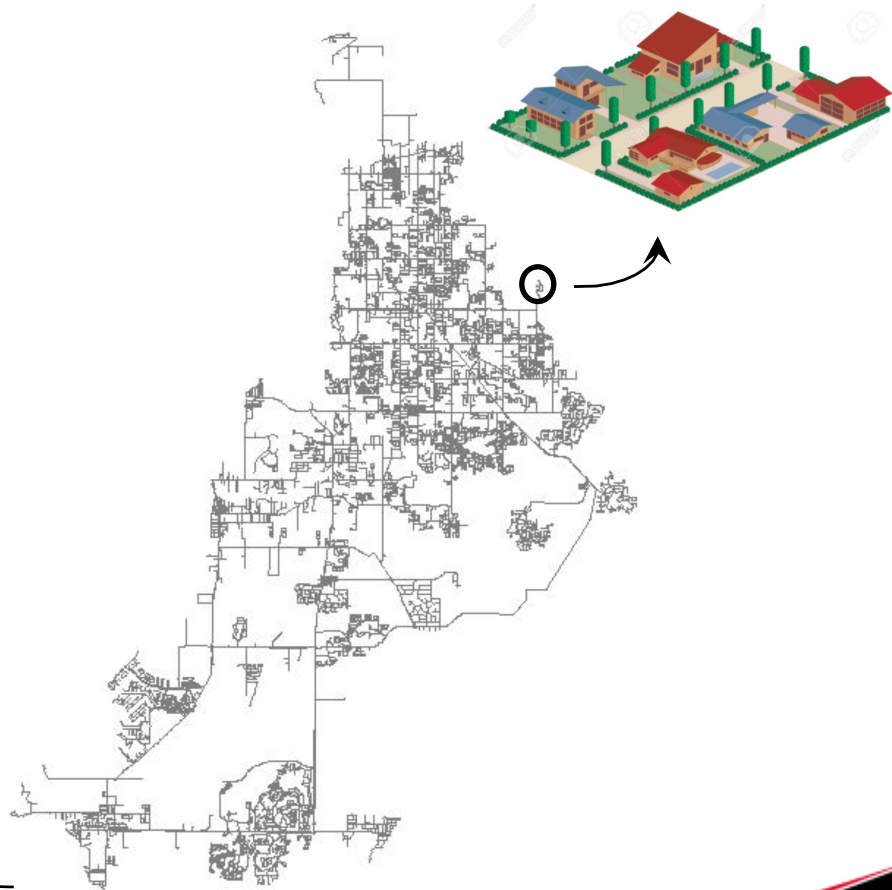


https://www.strand.com/services/water-supply/general-engineering

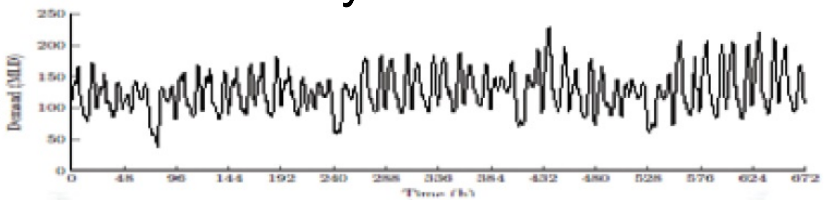
Scale of Interest

Single User Demands

- ... in demands somewhere between ...



Total System Demands



Real-Time Modeling Needs ...

- Network models that accurately represent the system infrastructure
- Solvers to simulate the hydraulics and water quality
- Ability to measure and forecast consumer demands
  - Drive the underlying hydraulics and water quality dynamics

- BUT ... consumer demands are usually not observed in real-time

Real-Time Modeling: Available Data

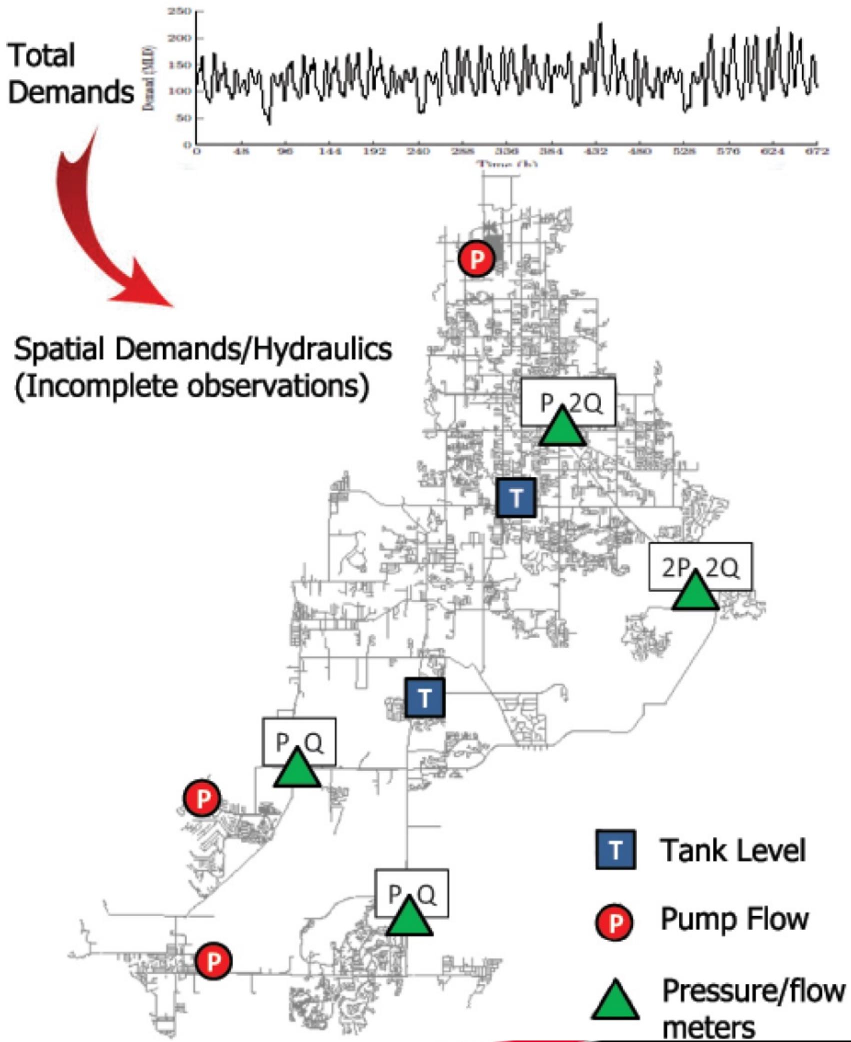
- Includes ...
  - System-wide (total) demands
  - Monthly/quarterly billing data
  - Limited, spatially distributed measurements of flow rates, pressures, tank levels at hourly (or shorter) time intervals
  - Demographic data associated with lot types, socio-economic information, etc

- How do we use this data to estimate and forecast demands?

Current Solution

- Developed a top-down approach to:
  - Estimate spatially distributed demands, and parameters of demand model
  - Using limited hydraulic observations

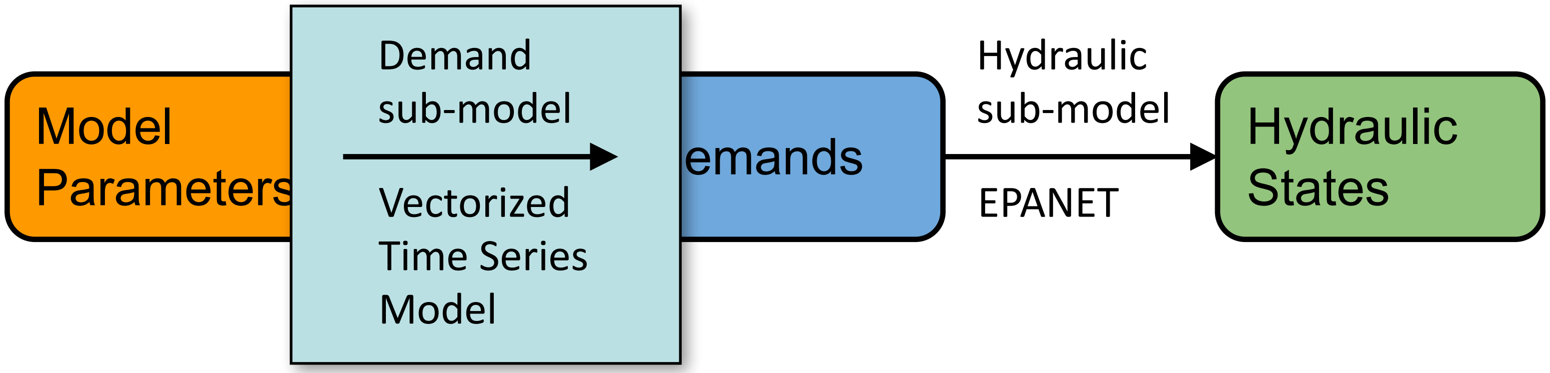
- Outcome is an algorithm to estimate and forecast:
  - Consumptive demands,
  - System states, and
  - Uncertainty characteristics



UNIVERSITY OF Cincinnati

Composite Demand-Hydraulic Model

- Developed the first approach to integrate
  - A vectorized time-series model for demands with
  - A hydraulic solver (e.g., EPANET)



- Formulated as a Dynamic Bayesian Network

Demand Sub-Model: Vectorized Time Series Model

- Capable of implementing any ARIMA model structure
- Focused on auto-regressive (AR) single- or double-seasonal models

Demands

$$\underbrace{\phi(B)\Phi_1(B^{24})\Phi_2(B^{168})}_{\text{Autoregressive parameters}}\underbrace{\nabla_1^d\nabla_{24}^{D_1}\nabla_{168}^{D_2}}_{\text{Differencing operators}}q_t = a_t$$

Gaussian error

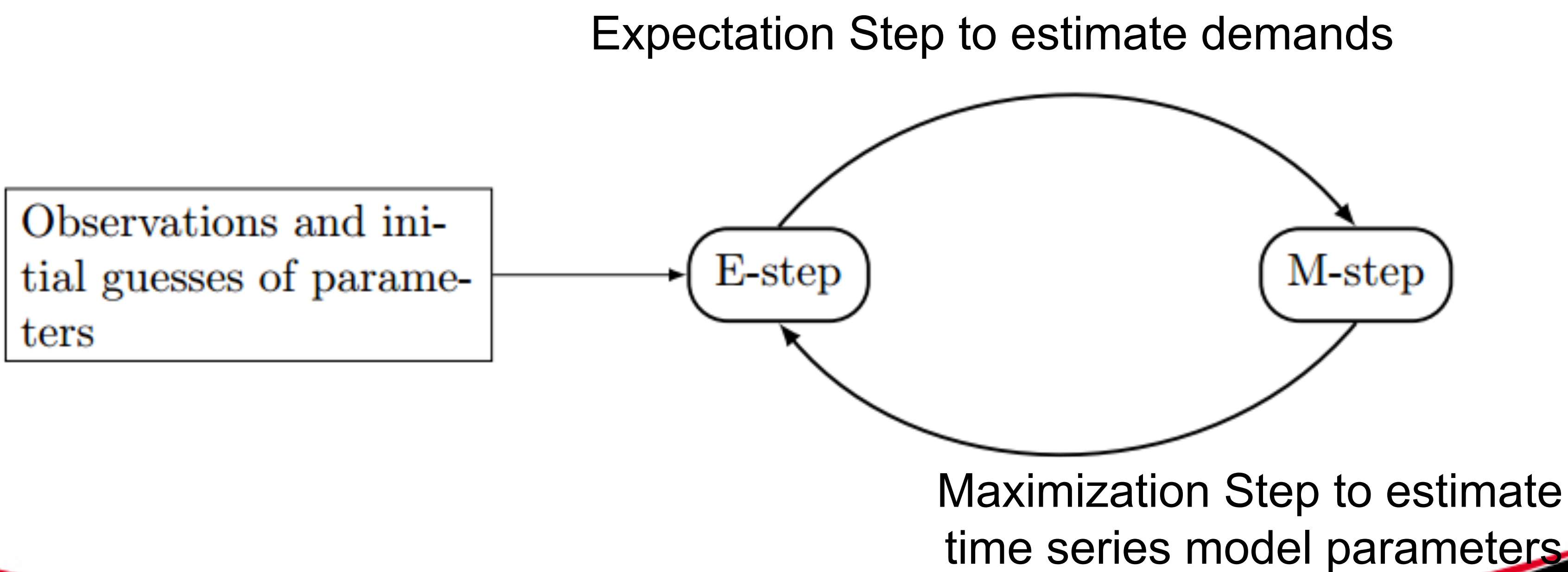
Autoregressive parameters

Differencing operators

Challenge: How do we estimate the unobserved demands and VARIMA model parameters using limited observed hydraulics?

Parameter/Demand Estimation

- Implemented an Expectation-Maximization (E-M) algorithm
  - An iterative approach used to estimate latent variables using observed data



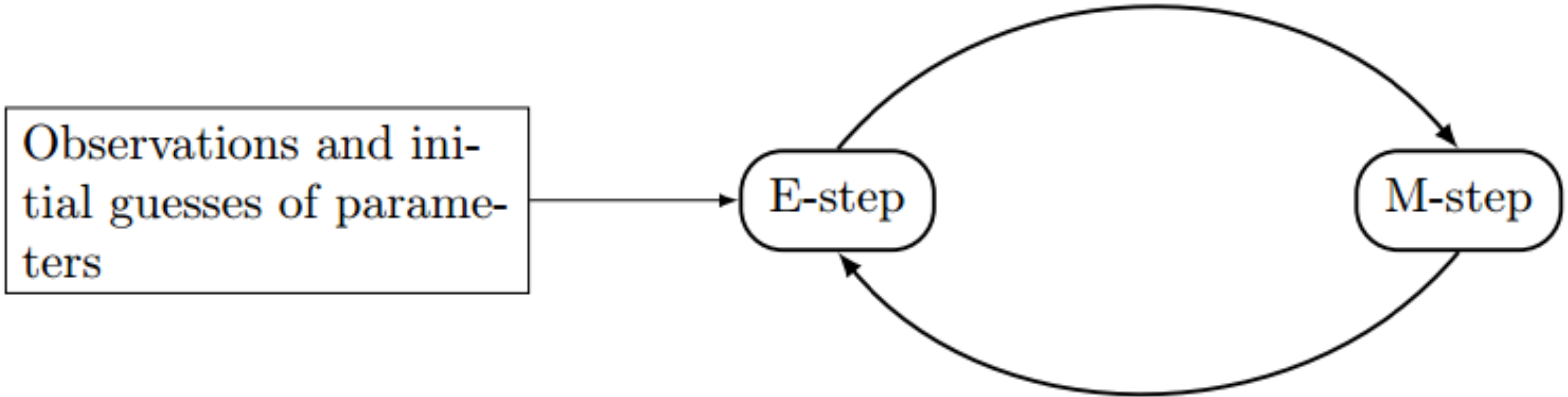
Dempster et al, 1997; Pasula et al, 1999

Expectation Step

- E-step: estimates water demands using a Markov chain Monte Carlo algorithm

Water demand estimates conditioned on the time series model and hydraulic observations

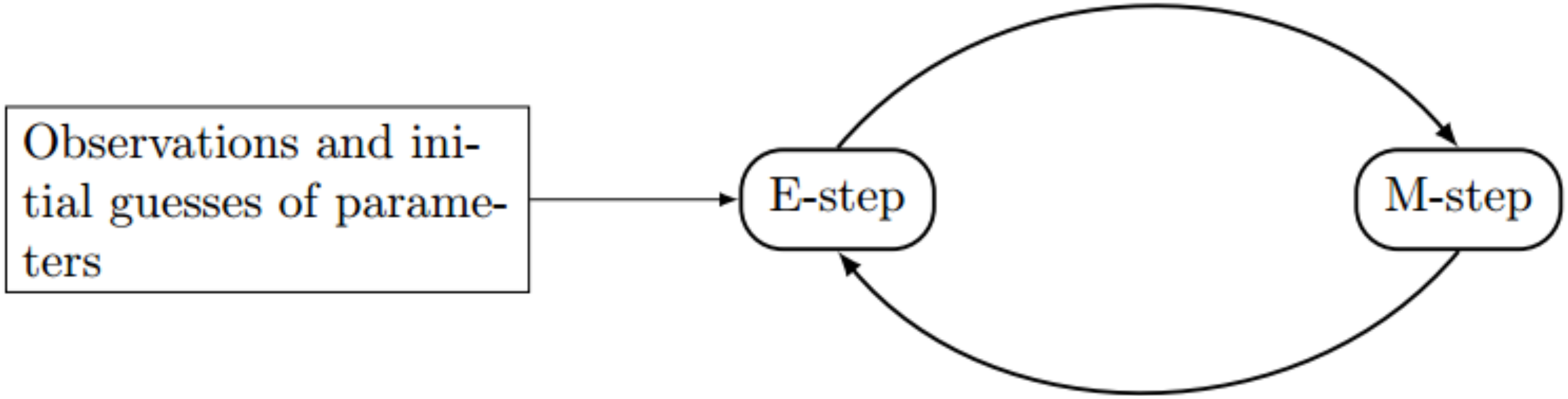
$$\overset{\text{Posterior}}{p(q|Y,\Xi)} \propto \overset{\text{Likelihood}}{p(Y|q)} \cdot \overset{\text{Prior}}{p(q|\Xi)}$$



Dempster et al, 1997; Pasula et al, 1999

Maximization Step

- M-step: non-linear parameter estimation for the time series model by minimizing the mean squared error (equivalent to maximum likelihood estimates)



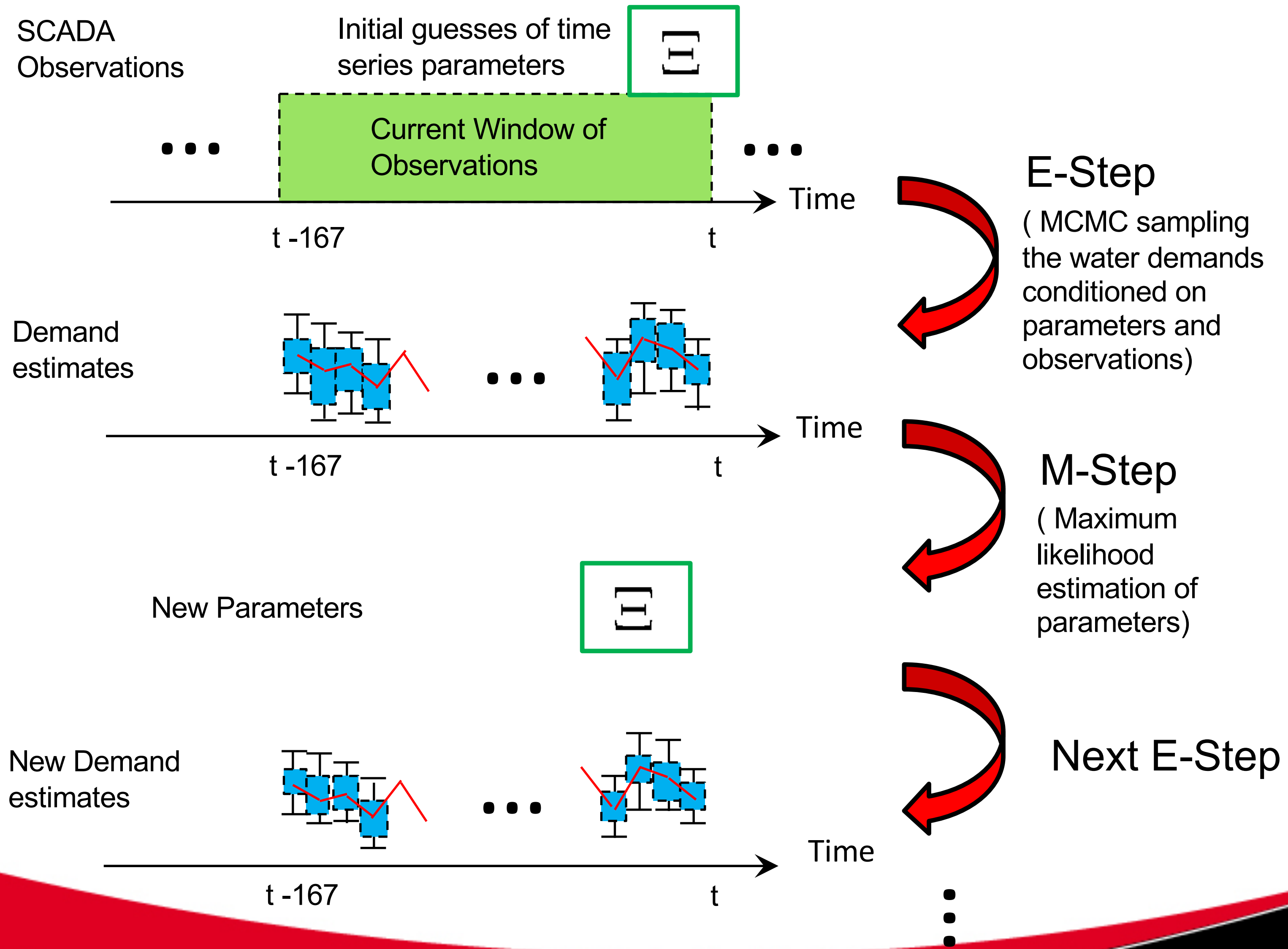
$$\max_{\Xi} \log L(\Xi; q, Y) = \log p(q, Y|\Xi)$$

Time series parameters updated using estimated demands

Dempster et al, 1997; Pasula et al, 1999

UNIVERSITY OF Cincinnati

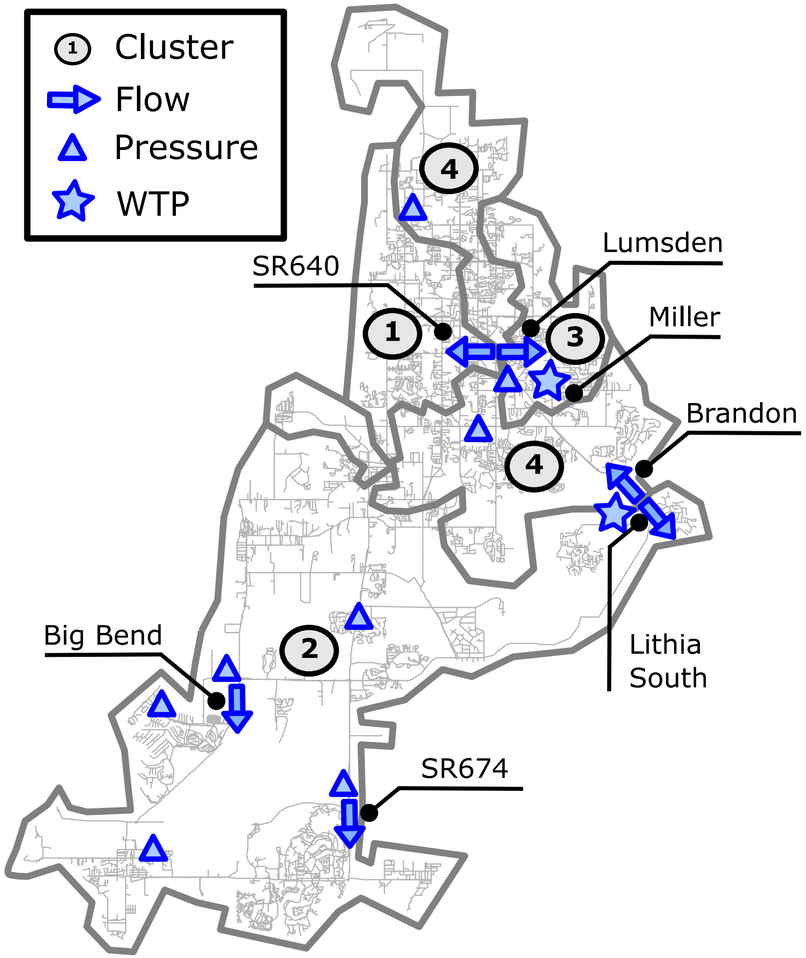
Graphical Concept



Real-World Network Study

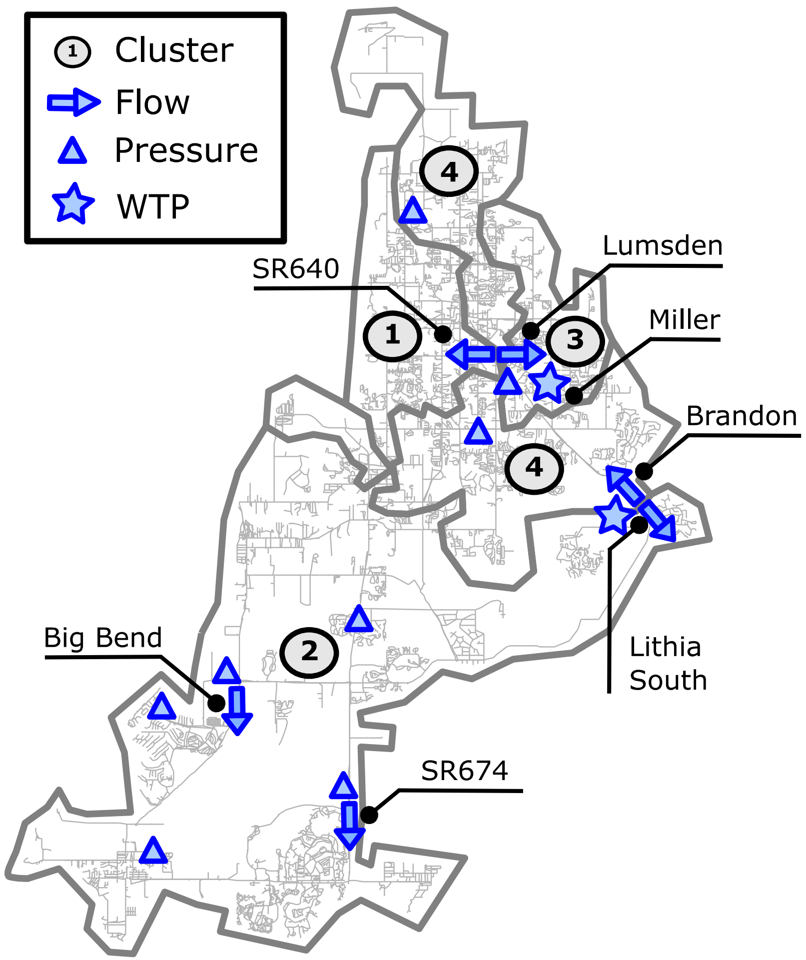
- Applied the composite demand-hydraulic model to a real-world case study to
  - Evaluate the overall performance
  - Identify challenges associated with a real-world application
- Intent was to identify additional needs to improve the integrated demand-hydraulic modeling approach

Case Study



- Real-world system with
  - Main treatment plant (Brandon and Lithia South) 56 – 170 MLD [15- 45 MGD]
  - Secondary treatment plant (Miller) 16 MLD [4.3 MGD]
  - Two tanks
  - Six flow measurements
  - Nine pressure measurements
  - Network has ~60,000 service connections represented by ~12,000 nodes
- Clustering
  - To reduce parameterization network was clustered into four regions based on flow path downstream from flow meters [modified from Qin and Boccelli, 2016 (under review)]

Case Study



- Time Series Model
  - Preliminary model used two auto-regressive and one seasonal term (24-hr)
  - Same model structure, not parameters, applied to each cluster

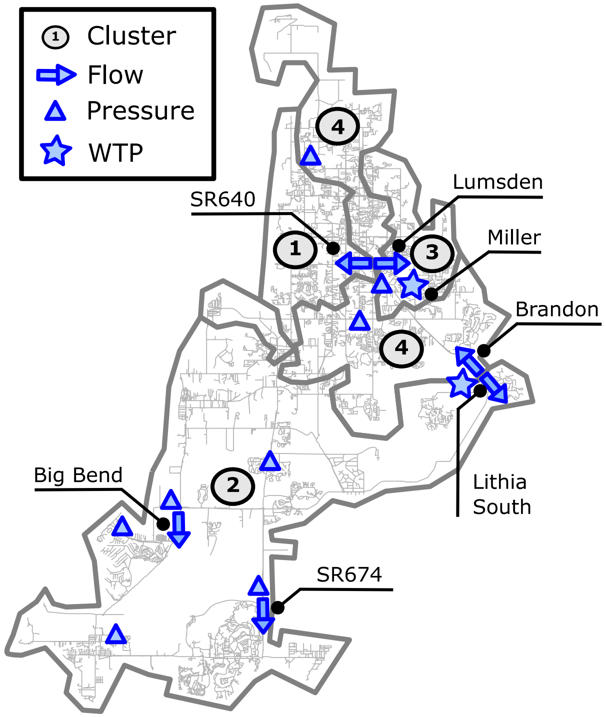
$$
\begin{aligned}
q_t &= \phi_1 q_{t-1} + \phi_2 q_{t-2} + \\
&\phi_{24} q_{t-24} + \phi_1 \phi_{24} q_{t-25} + \\
&\phi_2 \phi_{24} q_{t-26} + a_t
\end{aligned}
$$

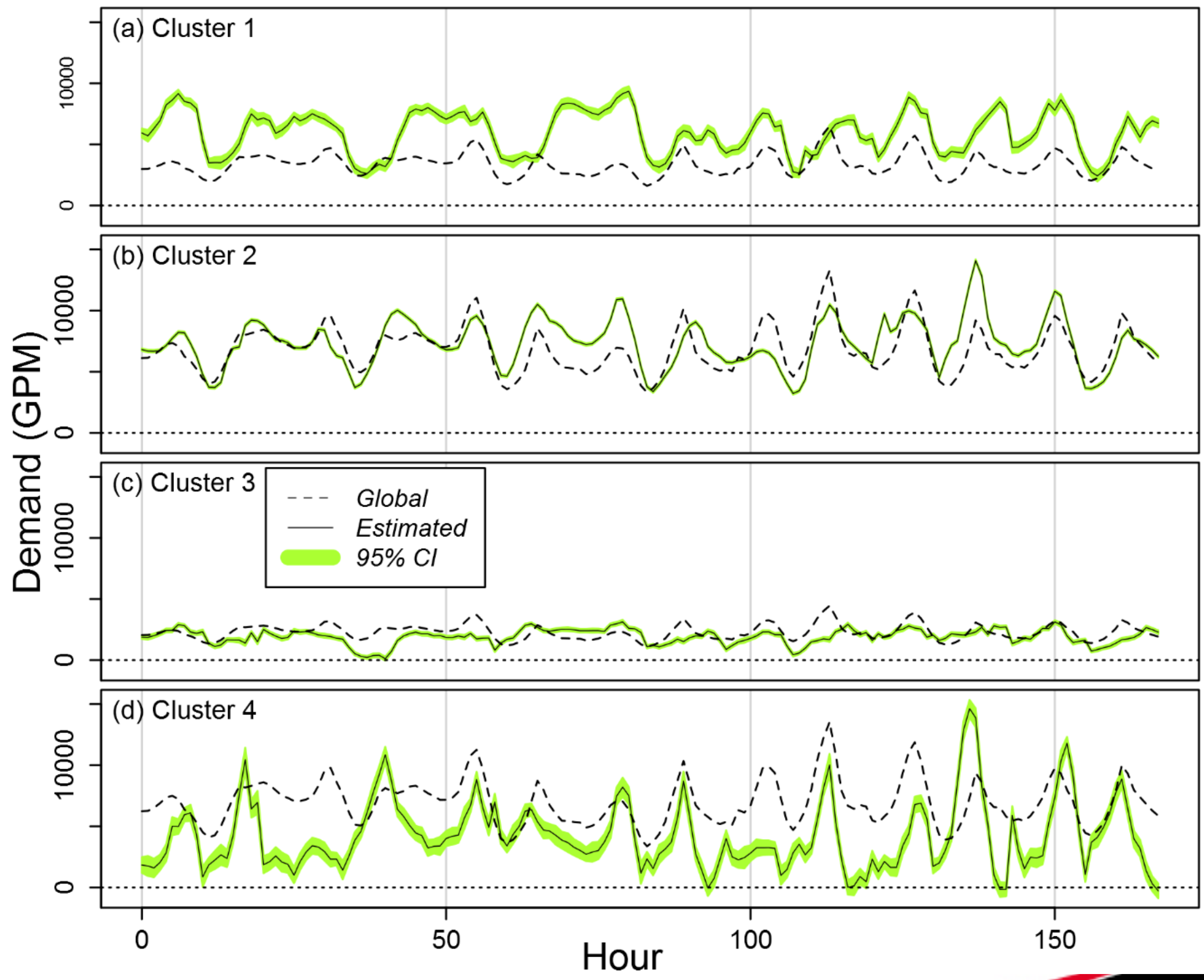
- Performed demand estimation with 168-hours

- Forecasted demands for an additional 24 hours

University of Cincinnati

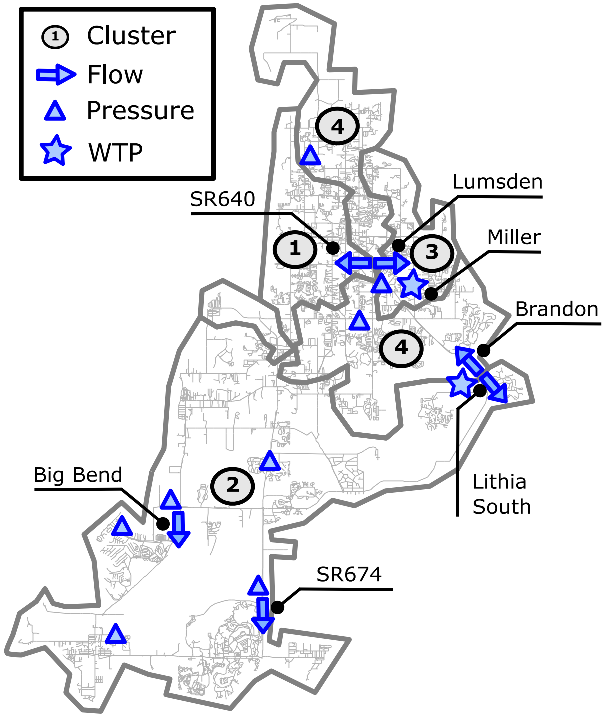
Results: Demand Estimates

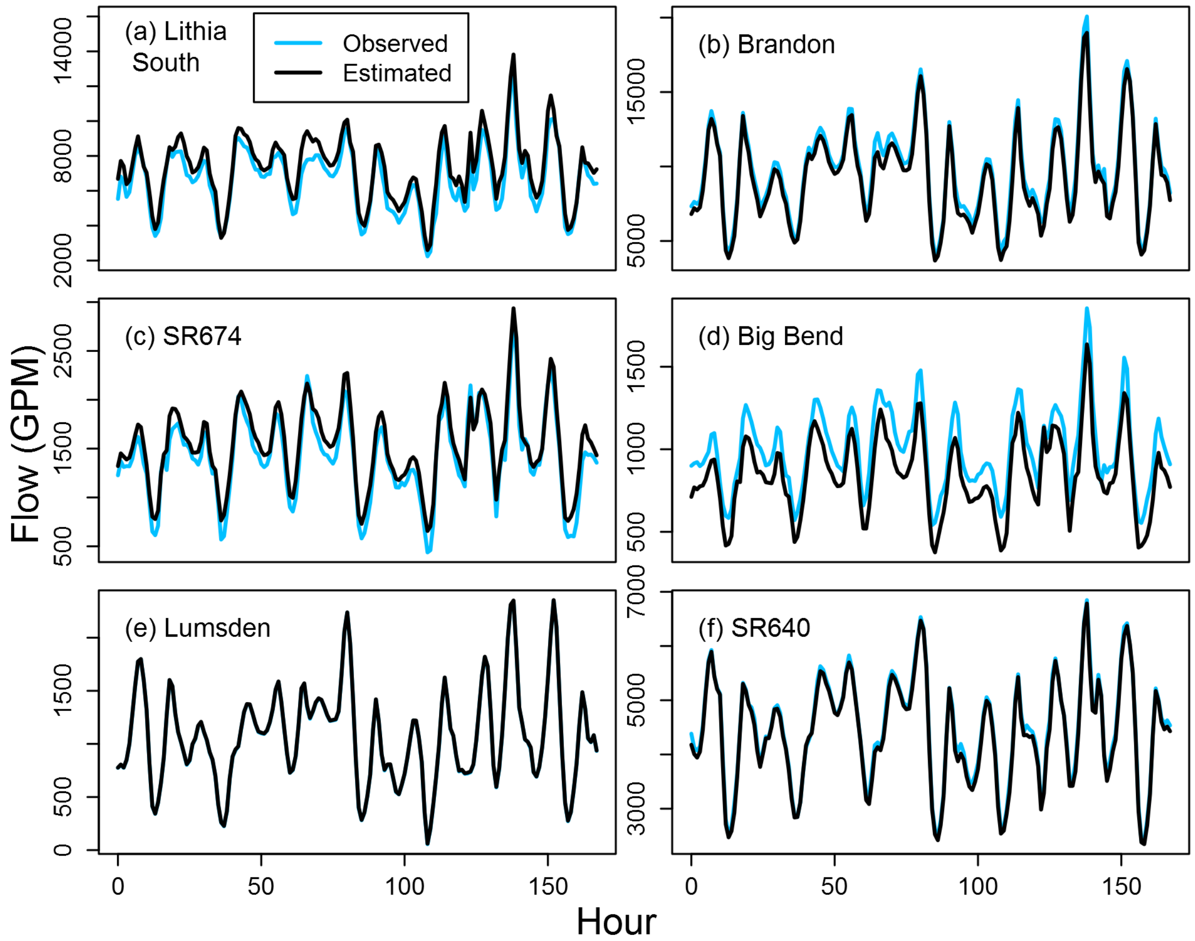




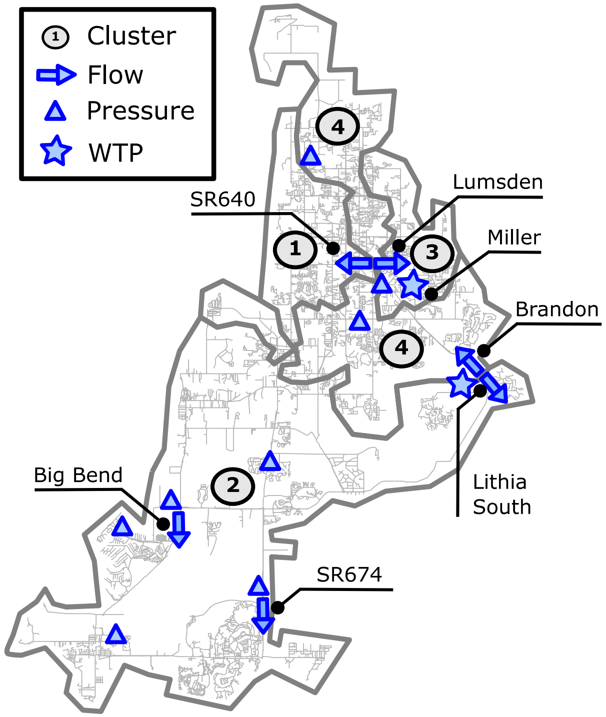
University of Cincinnati

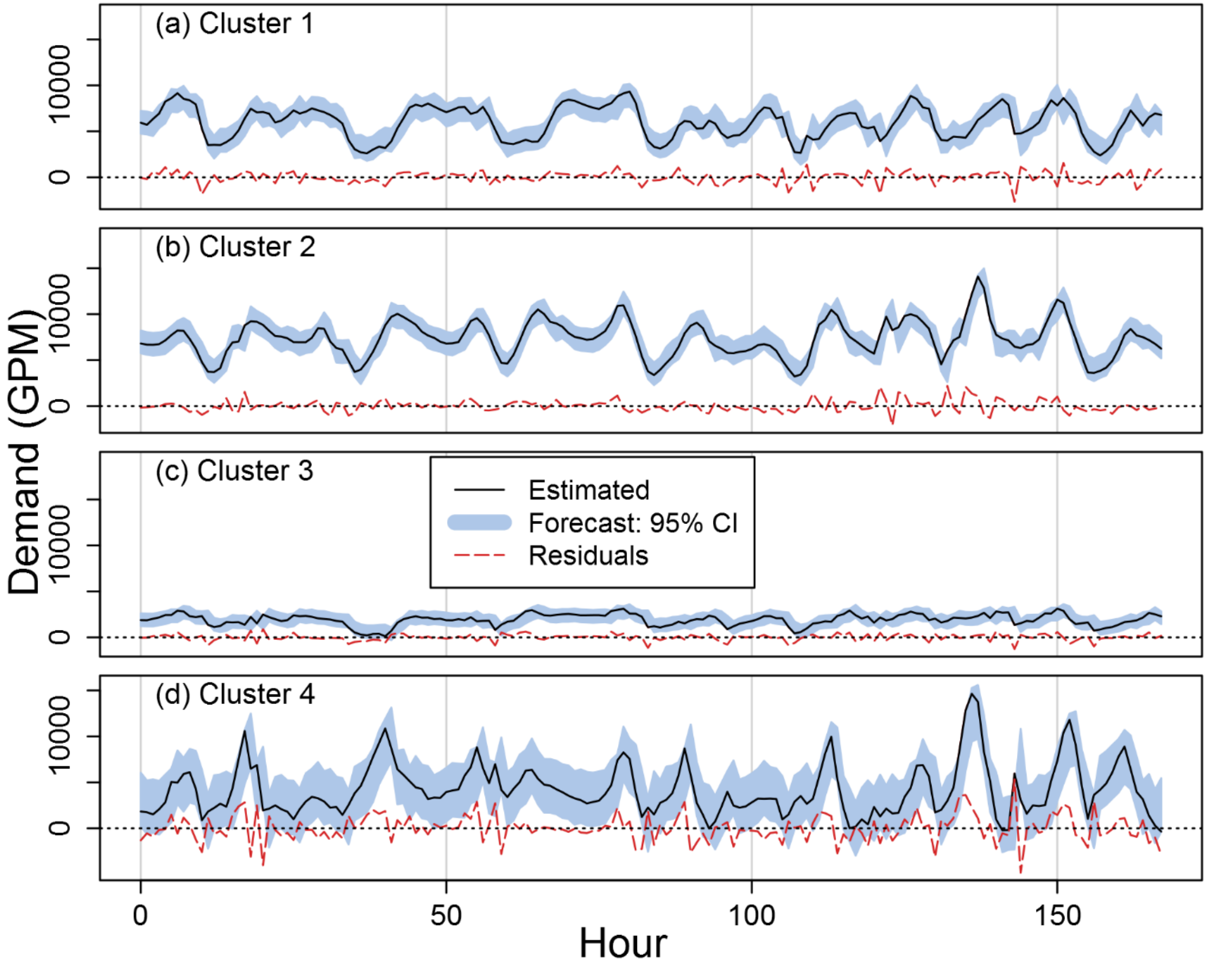
Results: Observed and Estimated Flows



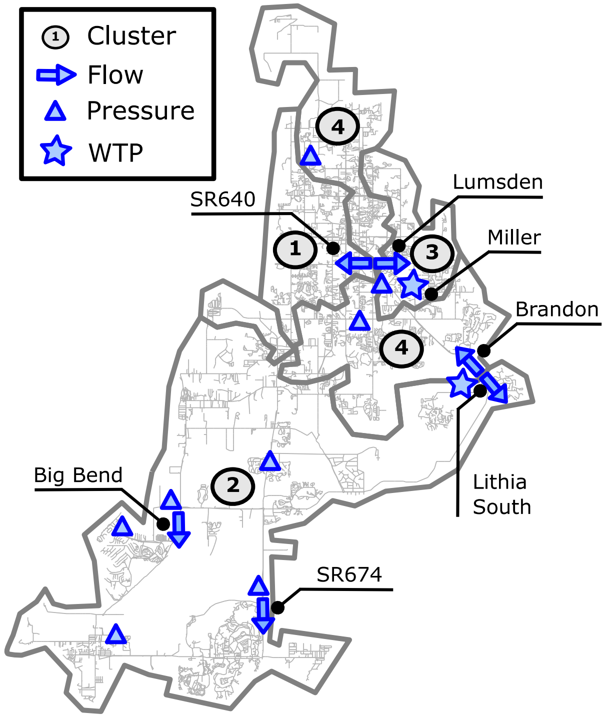


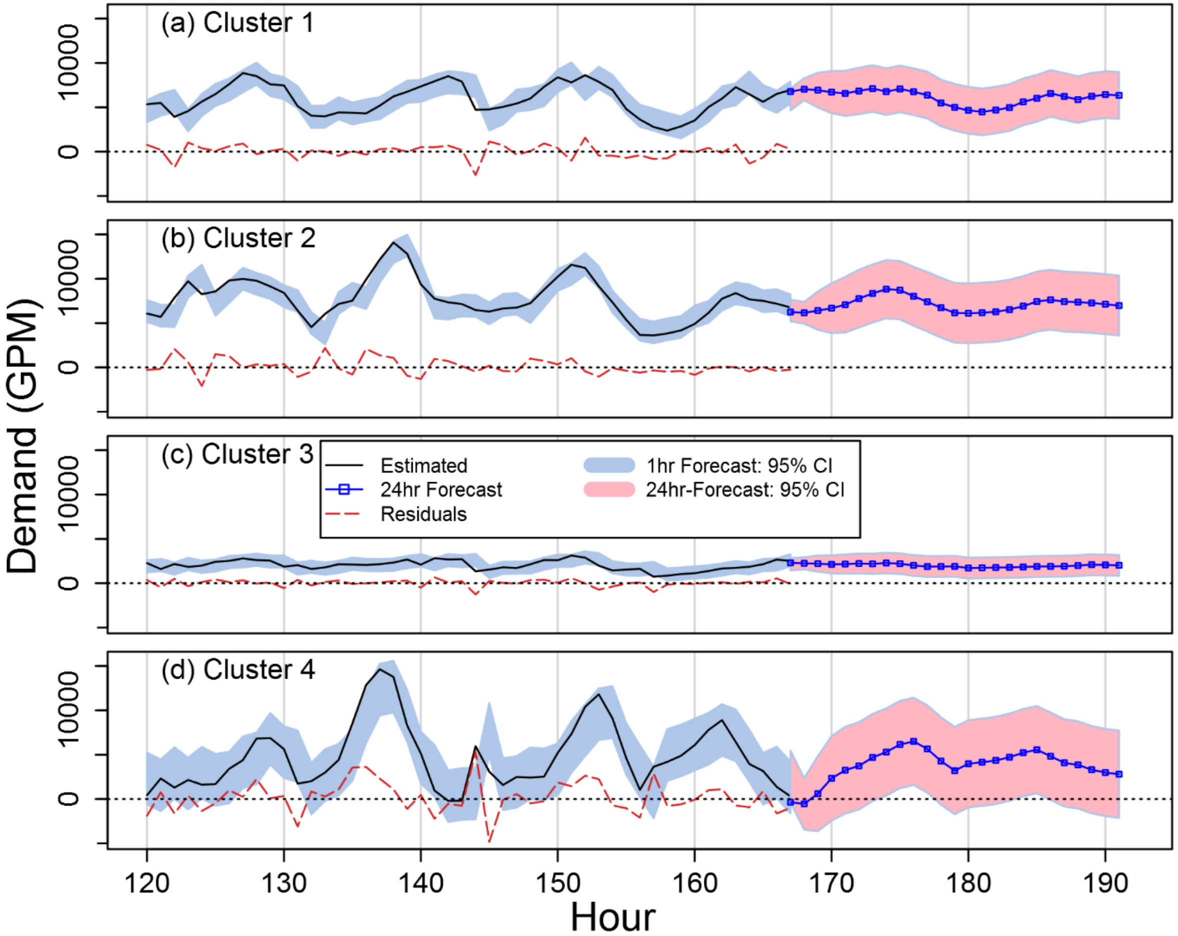
Results: 1-hr Ahead Forecasts





Results: 24-hr Forecasts

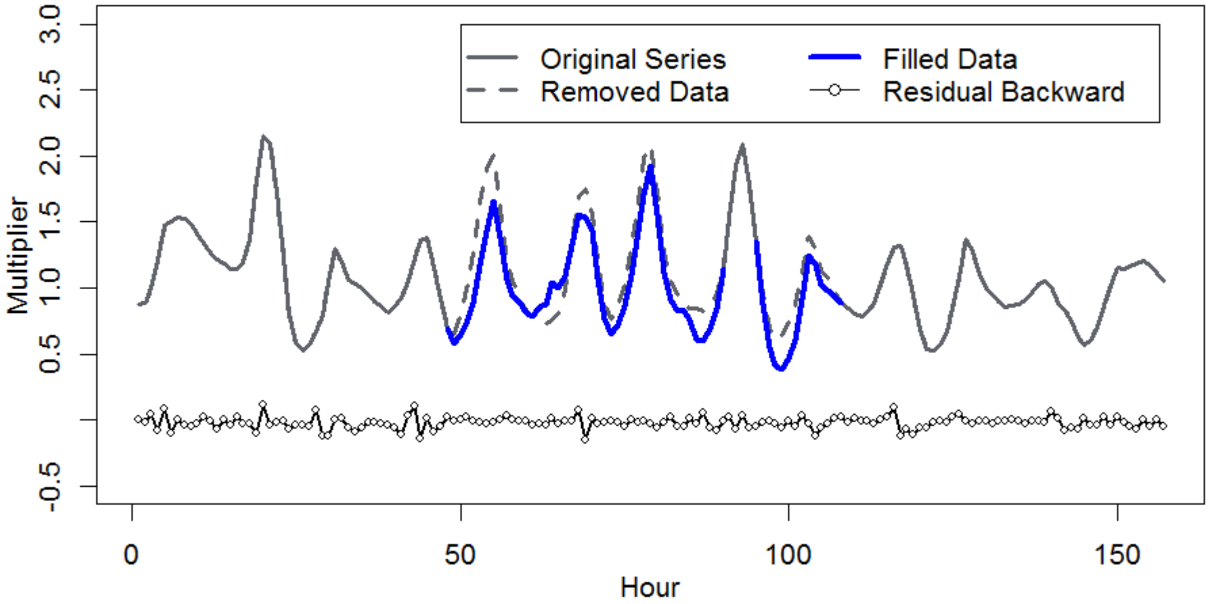




Lessons Learned: Missing Data

- One significant issue with SCADA data is incorrect and missing data
- Need approaches to identify and replace (or ignore) missing data when occurring

Have tested using time series models to represent observed flow data and filling in missing data



Lessons Learned: Clustering and Measurements

- The development of the clusters and/or location of the monitoring stations can effect the demand estimation process

- Observations [not shown] have demonstrated that for the same number of clusters, but using different approaches to cluster the network, can result in poor demand estimates
  - i.e., zero or negative demands

Lessons Learned: Physical Inacurracies

- Unknown/unobserved differences between reality and model representation
  - In particular, for this case study, there were significant challenges representing tank dynamics
    - Can adequately represent flows out of the tank through pumps, but typically overestimated the fill flow rate by 3 – 4 times the observed flow
    - Model was missing a pressure sustaining valve that physically existed

Summary and Conclusions

- This first real application of the composite demand-hydraulic model provided:
  - Good demand estimates and representation for observed hydraulics
  - Demand estimates routinely within the 1-hr ahead forecasting values
  - Long-term forecasting results in relatively large uncertainties
  - Implementation of a real-time model also requires significant investment into ensuring accurate representation of the physical system

Next/Future Steps

- Demand Estimation
  - Lognormal representation of the demands
  - Double seasonal times series models and additional model identification
- Demand Forecasting
  - Identifying model structures to improve forecasting not just estimation
  - Efficient approaches for forecasting demands and hydraulic states
- Real System Assessment
  - Work more closely with utility on physical representation
  - Comparison of performance with available tracer data to assess transport improvements

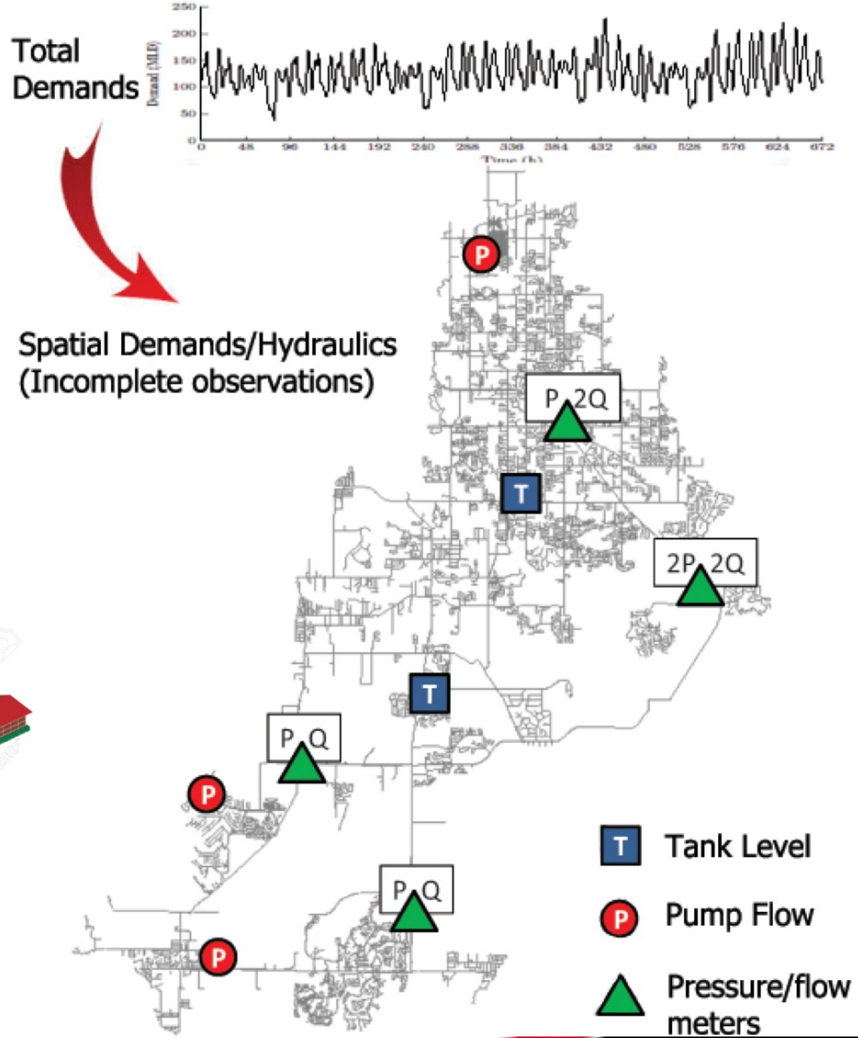
Acknowledgements

- Partial funding support from
  - National Science Foundation CMMI (#09000713)
  - Water Research Foundation (#04345)
  - National Science Foundation CBET (#1511959)
  - Ohio Water Resources Center (#60048647)

- Questions?

Scale of Interest

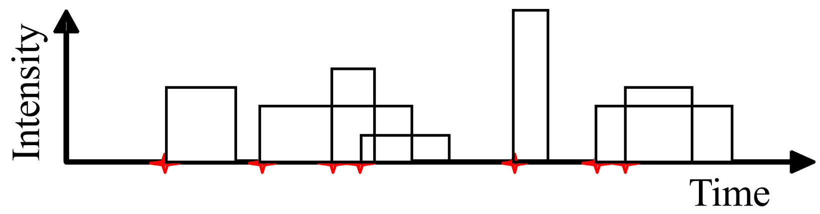
- Interested in demands between ...



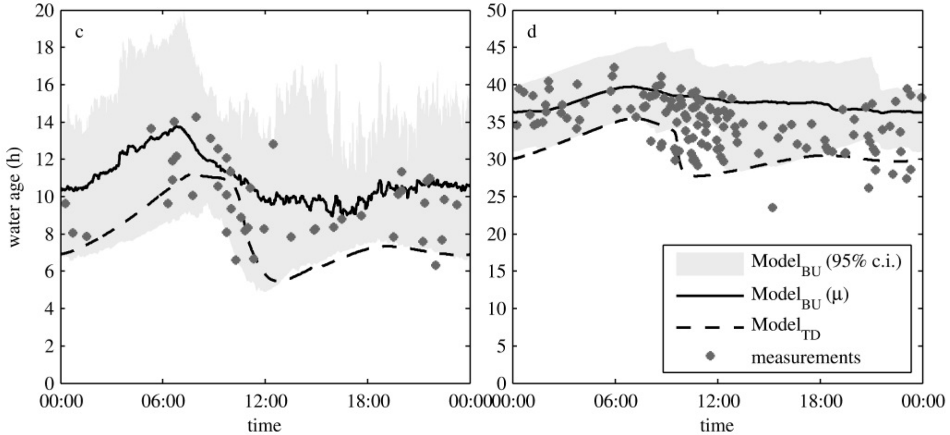
https://www.123rf.com

Demand Modeling: Bottom-Up Approach

- Stochastic modeling of demands at individual service connections
  - Includes arrival rates, and distributions of intensity and duration of individual water usage



  - Blokker et al used demographic information to estimate demands
  - Data intensive, challenging to keep up the data set



Li and Buchberger, 2006; Blokker et al, 2010

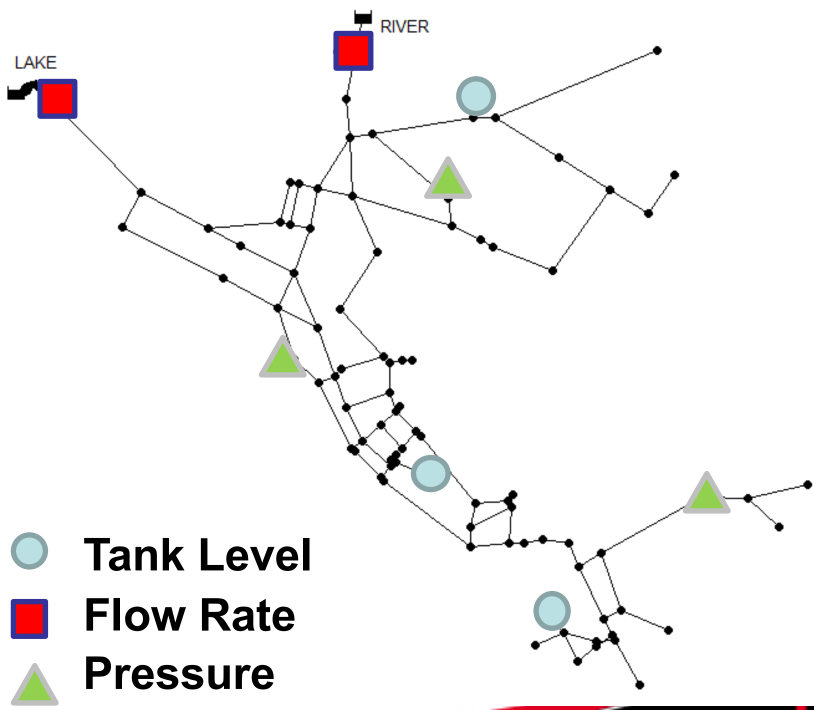
University of Cincinnati

Demand Modeling: Top-Down Approach

- Deterministic modeling with temporal/spatial demands representing an average/extreme demand scenario
  - Typically performed as “calibration” to match observations
  - Real-time approaches have used extended Kalman filters to estimate the demands

    Shang et al, 2006; Kang and Lansey, 2010

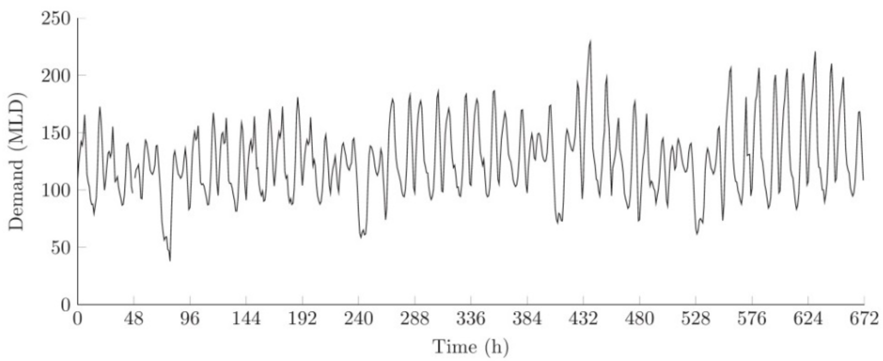
  - Capture spatial distribution, but not temporal relationships
  - No predictive ability

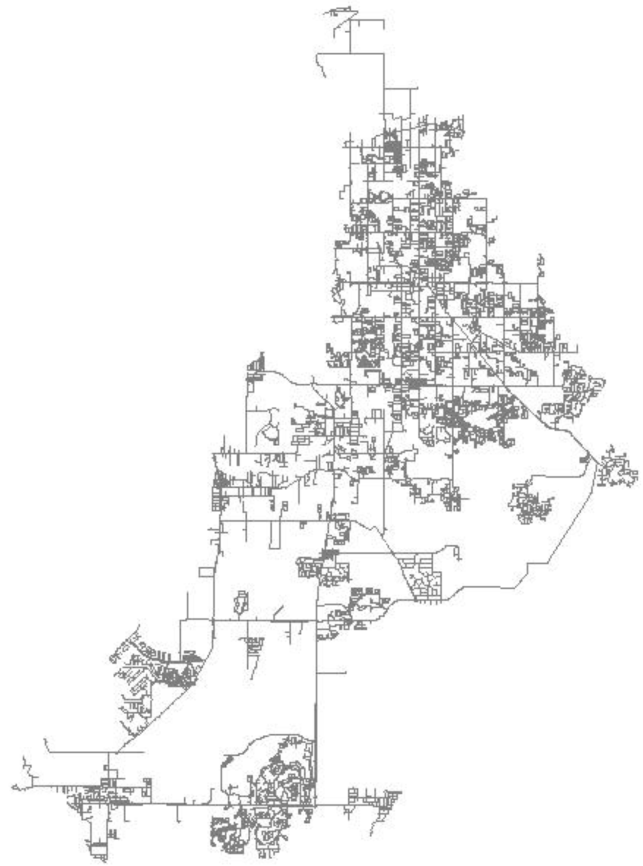


UNIVERSITY OF Cincinnati

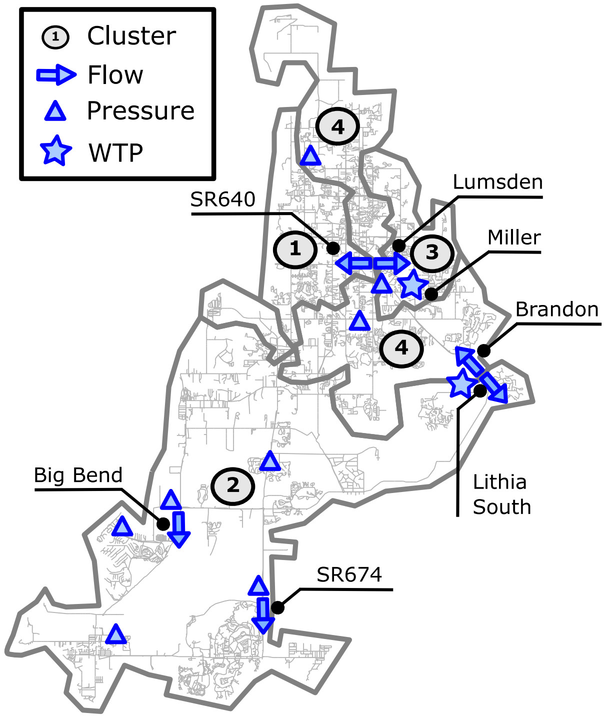
Demand Modeling: Temporal Correlations

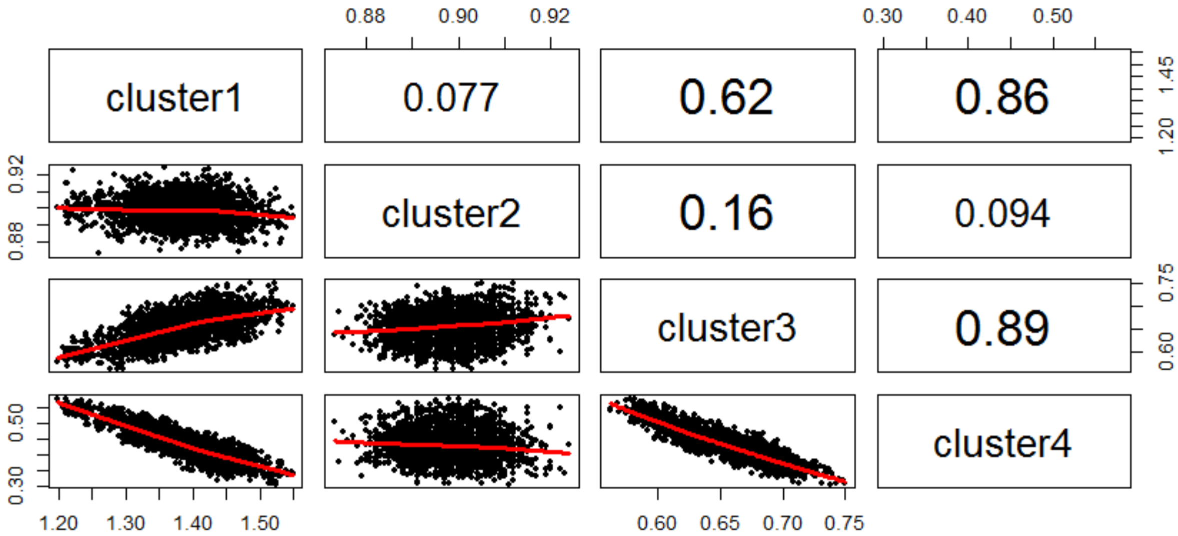
- Time series modeling applied to observed system-wide (total) demands
  - No spatial distribution





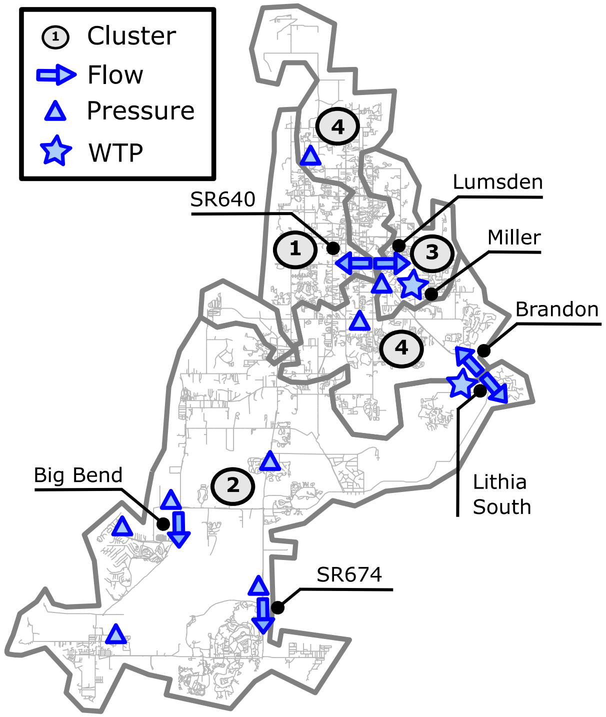
Results: Scatter Plots Demands

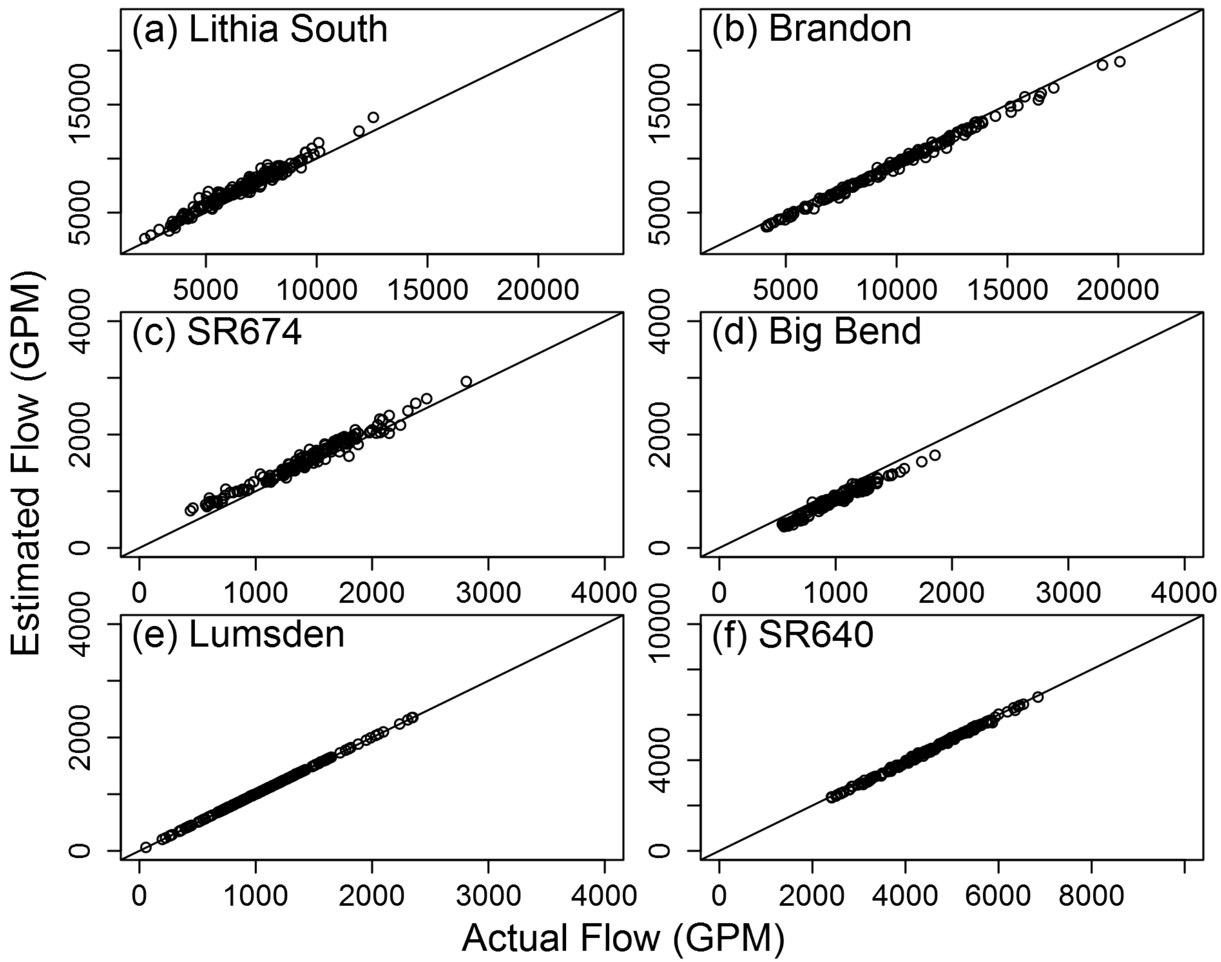




UNIVERSITY OF Cincinnati

Results: Scatter Plots Flows





UNIVERSITY OF Cincinnati

Real-Time Modeling

- Requires real-time demand estimates and forecasts
- Challenge: How to estimate and forecast demands using:
  - System-wide (total) demands
  - Monthly/quarterly billing data (i.e., base demands)
  - Spatially limited measurements of flow rates, pressures, tank levels at hourly (or shorter) time intervals
  - Potentially inaccurate model representations of the physical network

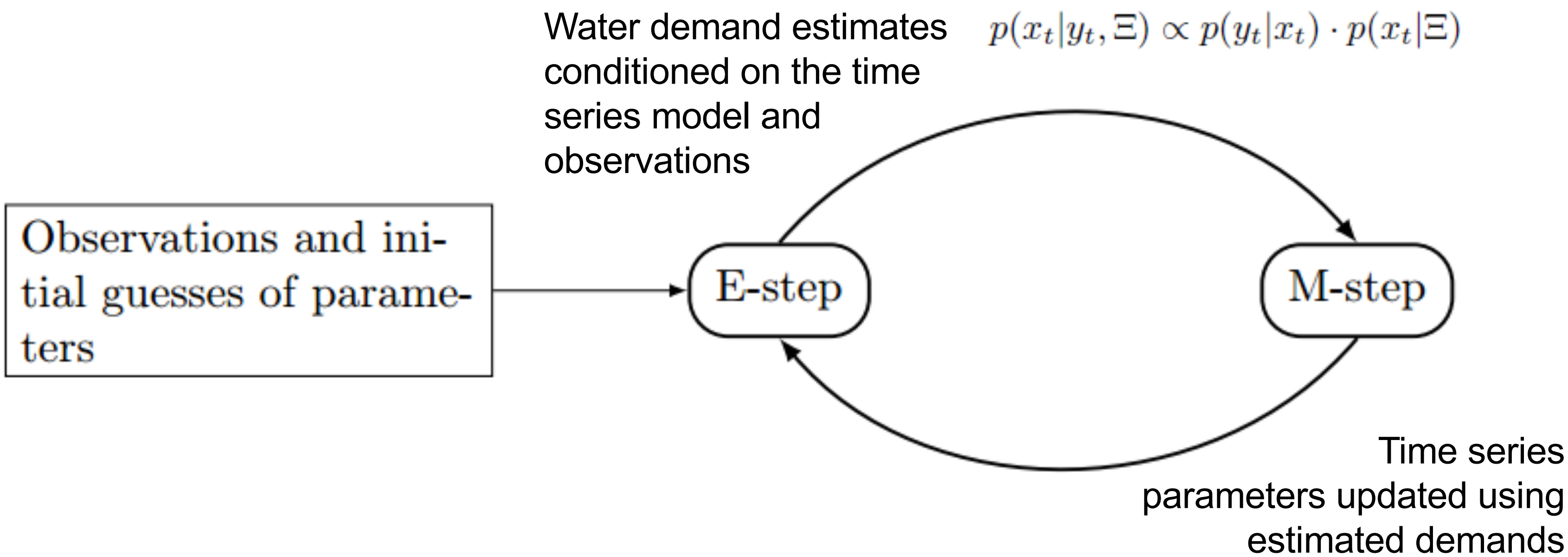
Parameter/Demand Estimation

- Implemented an Expectation-Maximization (E-M) algorithm

- The E-M algorithm is used to
  - Estimate latent variables
    - demands and time series parameters
  - Using observed data
    - i.e., flows, pressures

Parameter/Demand Estimation

- Implemented an Expectation-Maximization (E-M) algorithm
  - An iterative approach used to estimate latent variables using observed data



Dempster et al, 1997; Pasula et al, 1999

Expectation (E)-step

- Estimate the posterior distribution of demands using likelihood function using
  - Time series model as a prior, and
  - Observed data

$$p(\boldsymbol{q}|\boldsymbol{Y},\Xi) \propto p(\boldsymbol{q},\boldsymbol{Y},\Xi) = p(\boldsymbol{Y}|\boldsymbol{q},\Xi)\cdot p(\boldsymbol{q}|\Xi) = \boxed{p(\boldsymbol{Y}|\boldsymbol{q})\cdot p(\boldsymbol{q}|\Xi)}$$

q: demand esitmates
Y: hydraulic observations
Ξ: time series model parameters

Known with hydraulic sub-model

Known with demand sub-model

- Use a Markov chain Monte Carlo estimation approach to estimate demands

Maximization (M)-step

- Given the estimated demands
  - Estimate the parameters of the VARIMA demand model using mean squared error (equivalent to maximum likelihood estimates)

Using likelihood principle

$$\log L(\Xi; \boldsymbol{q}, \boldsymbol{Y}) = \log p(\boldsymbol{q}, \boldsymbol{Y}|\Xi)$$

$$= \log p(\boldsymbol{q}|\Xi) + \boxed{\log p(\boldsymbol{Y}|\boldsymbol{q}) + C}$$

q: demand esitmates
Y: hydraulic observations
Ξ: time series model parameters

Independent of Ξ;
estimated in E-step

UNIVERSITY OF Cincinnati

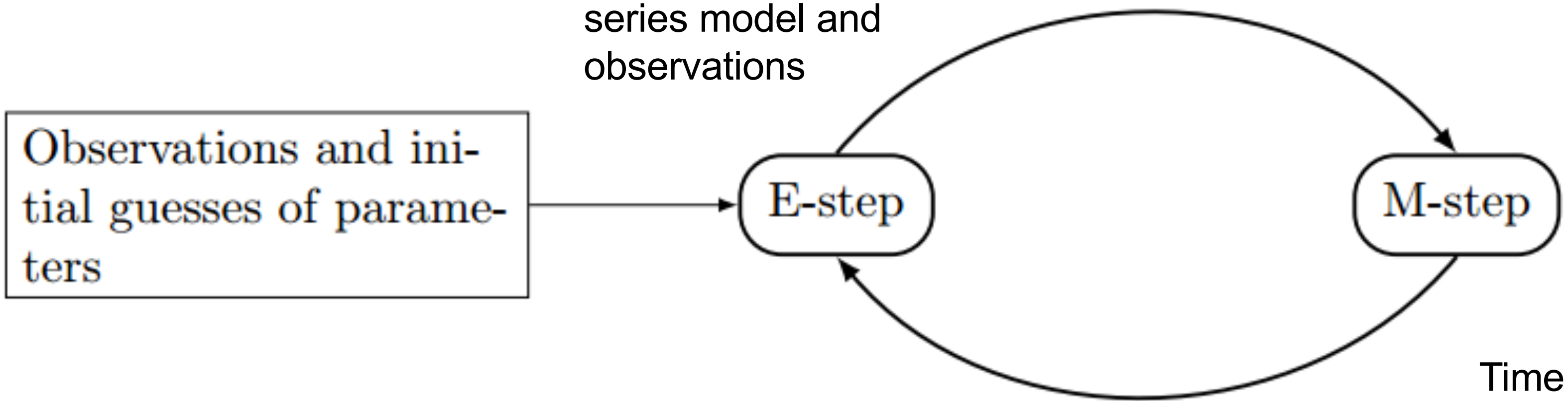
E-M Algorithm

- E-step: estimates water demands using a Markov chain Monte Carlo approach with a
  - Time series model as a prior and
  - Likelihood of observed data

Posterior Likelihood Prior

$$p(x_t|y_t,\Xi) \propto p(y_t|x_t) \cdot p(x_t|\Xi)$$

Water demand estimates conditioned on the time series model and observations



Time series parameters updated using estimated demands

- M-step: estimates the parameters of the time series model by minimizing the mean squared error (equivalent to maximum likelihood estimates)

Dempster et al, 1997; Pasula et al, 1999

Demand Sub-Model: Vectorized Time Series Model

- Example: single-seasonal model

$$\left.\begin{aligned} x_t - A_1 x_{t-1} - \cdots - A_P x_{t-P} - \mu &= a_t \\ \phi(B)\Phi_1(B^s)\nabla^d \nabla_s^{D_1} x_t &= a_t \end{aligned}\right\} A(B)x_t = a_t$$

x_t is the vector of water demands at time t

Challenge: How do we estimate the unobserved demands and VARIMA model parameters using limited observed hydraulics?

covariance matrix Σ.

UNIVERSITY OF Cincinnati

Real-Time Modeling

- Available information for demand estimation
  - System-wide (total) demands
  - Monthly/quarterly billing data
  - Demographic data associated with lot types, socio-economic information, etc.
  - Spatially limited measurements of flow rates, pressures, tank levels at hourly (or shorter) time intervals
- How do we use this data to estimate and forecast demands?

Outline

- Background
- Motivation
- A statistical demand-hydraulic model
- The Expectation-Maximization (E-M) algorithm
- Case study
- Results and discussions
- Future work

Introduction

- Water utilities must ensure potable water infrastructure are sustainable, robust and resilient to long- and short-term challenges

- Long-term challenges include
  - Climate change
  - Population shifts
  - Aging infrastructure
- Addressed through infrastructure design

http://trenchlessonline.com/new

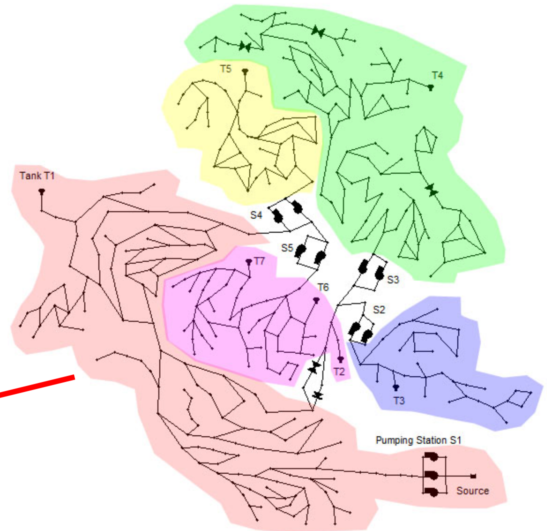
Hydraulic model: framework

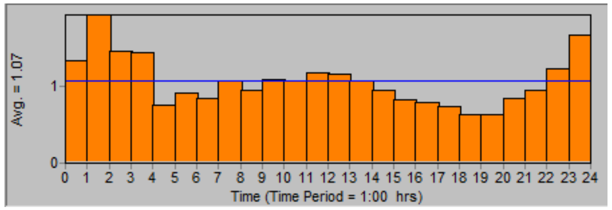
	Type	Data	Data Source/Measurements
Inputs	Network	Network connectivity, pipe diameters/roughness, tank geometries, etc. (Static during EPS)	GIS; Asset Management System (AMS)
Inputs	"Controls"	On/off statuses of pumps/control valves, speed settings of VFPs, tank levels, etc.	Control rules or results from previous time steps (in EPS), historic actions are available in SCADA DB
Inputs	**Demands**	Short-term water demands for individual customers	Automatic Meter Reading (AMR) system; monthly water bills; empirical patterns ?
Outputs	Hydraulics	Nodal pressures, pipe/pump flows	SCADA system (however, typically only partial coverage for a network)

UNIVERSITY OF Cincinnati

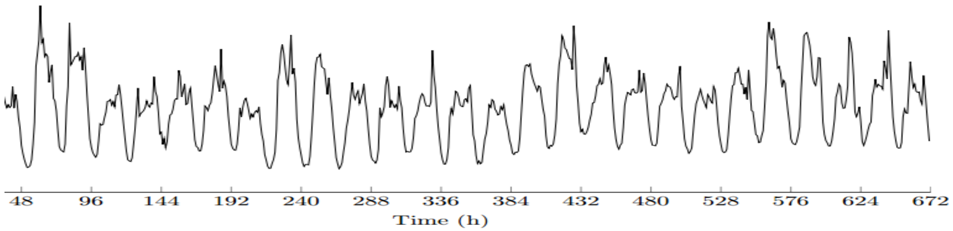
The models of water demands

Traditional approach:
demand group-demand pattern





……



…

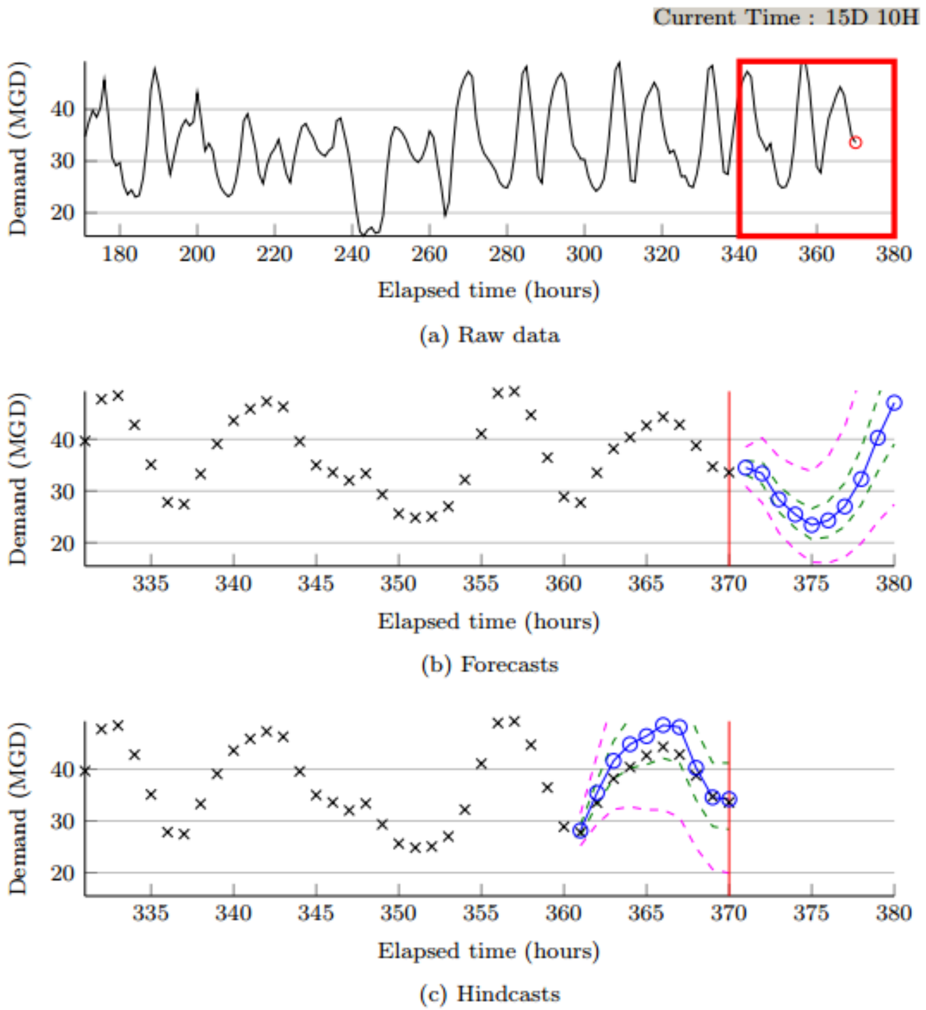
Improvements: explicitly model the two characteristics :
(1) periodicity and (2) short-term (auto-) correlations

UNIVERSITY OF Cincinnati

Benefits of using time series models

- Using seasonal (periodic) time series model is expected to improve the forecasts of **system-wide demands** (Chen and Boccelli, 2013)
- Forecasts are updated as real-time observations are received
- Varied forecasting horizons
- Quantification of uncertainties

Can we use (vector) time series model for spatially distributed demands?



UNIVERSITY OF Cincinnati

Motivation

$$\boxed{\text{Time series model}} \xrightarrow{\text{Water demands}} \boxed{\text{EPANET}}$$

- We would like to use SCADA data to estimate the parameters of the (multivariate) time series model
  - Extension to the methodology for univariate water demands

- The composite model will have the capabilities provided by the time series model
  - Better online forecasting of demands and hydraulics
  - Uncertainty quantification

The demand-hydraulic model

Demand model: $\boxed{x_t} = \boxed{A_1} x_{t-1} + \cdots + \boxed{A_P} x_{t-P} + \boxed{\mu} + a_t$ $\qquad a_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \boxed{\Sigma})$

Vector of water demands at t

Linear parameter matrices of the time series model

White noise

Covariance matrix

Hydraulic model: $\boxed{y_t} = H(v, u_t, \boxed{x_t}) + e_t$ $\qquad e_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \Sigma_e)$

Vector of monitored hydraulic variables

Network data

Control data

Measurement errors

Our objective is to estimate water demands and model parameters given hydraulic observations

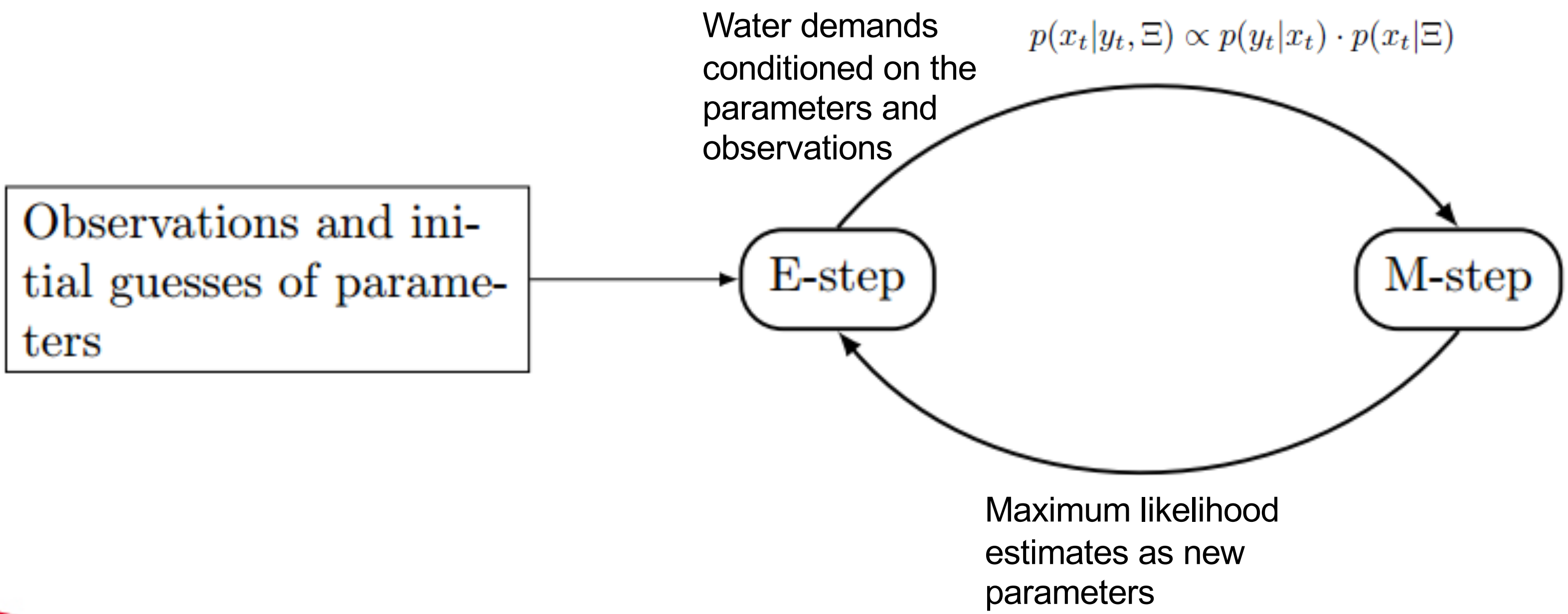
$\boxed{y_t} \xrightarrow{?} \boxed{x_t}$, $\boxed{A_1 \; A_P \; \Sigma}$

(n*n) matrices

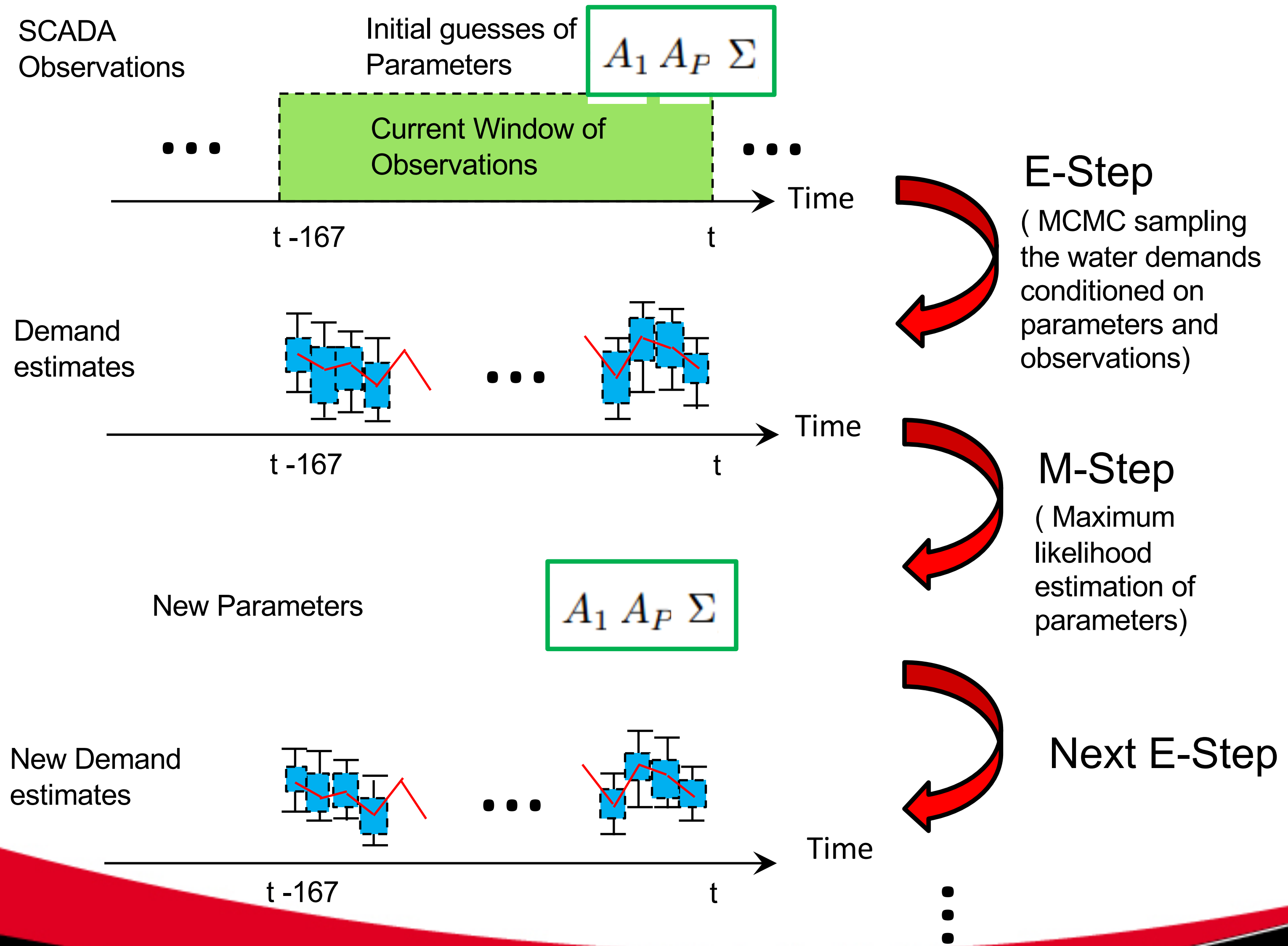
UNIVERSITY OF Cincinnati

The EM algorithm (Pasula et.al., 1999)

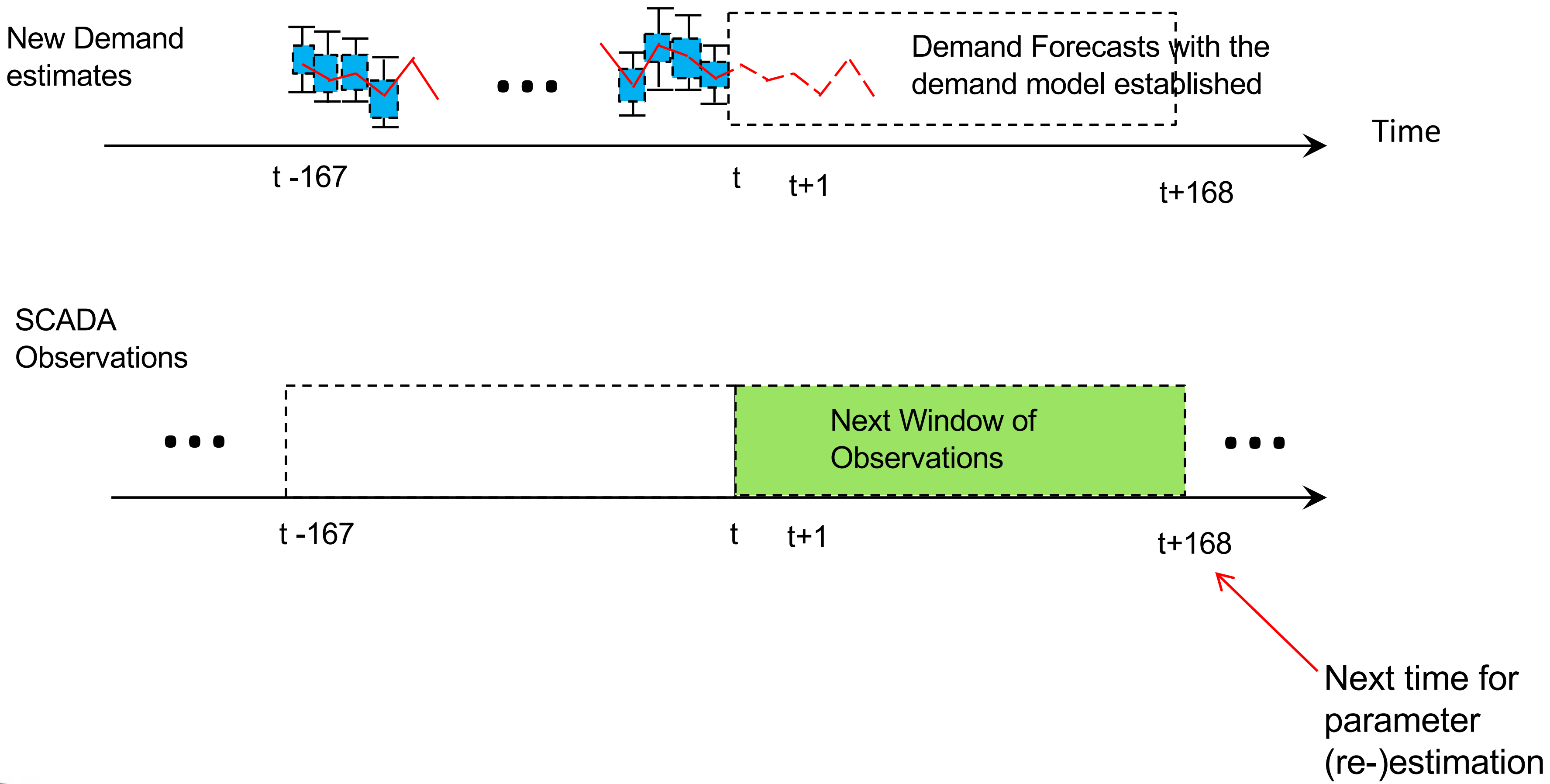
- Expectation-Maximization
- Iteratively update point estimates of parameters and distribution estimates of latent variables (demands)
- E-step: Markov chain Monte Carlo



Concept: E-M algorithm in demand estimation



Concept: E-M algorithm in demand estimation

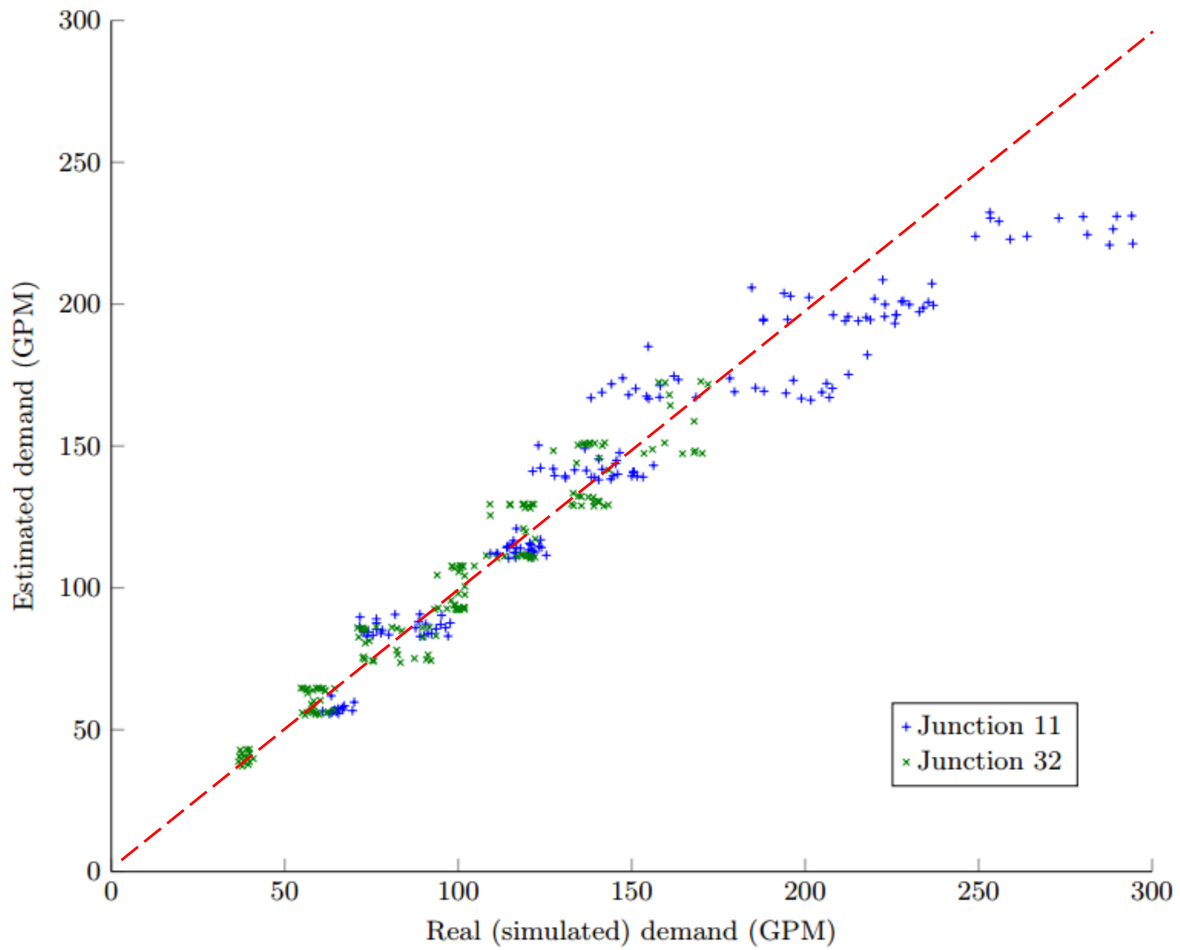


UNIVERSITY OF Cincinnati

Demand estimates

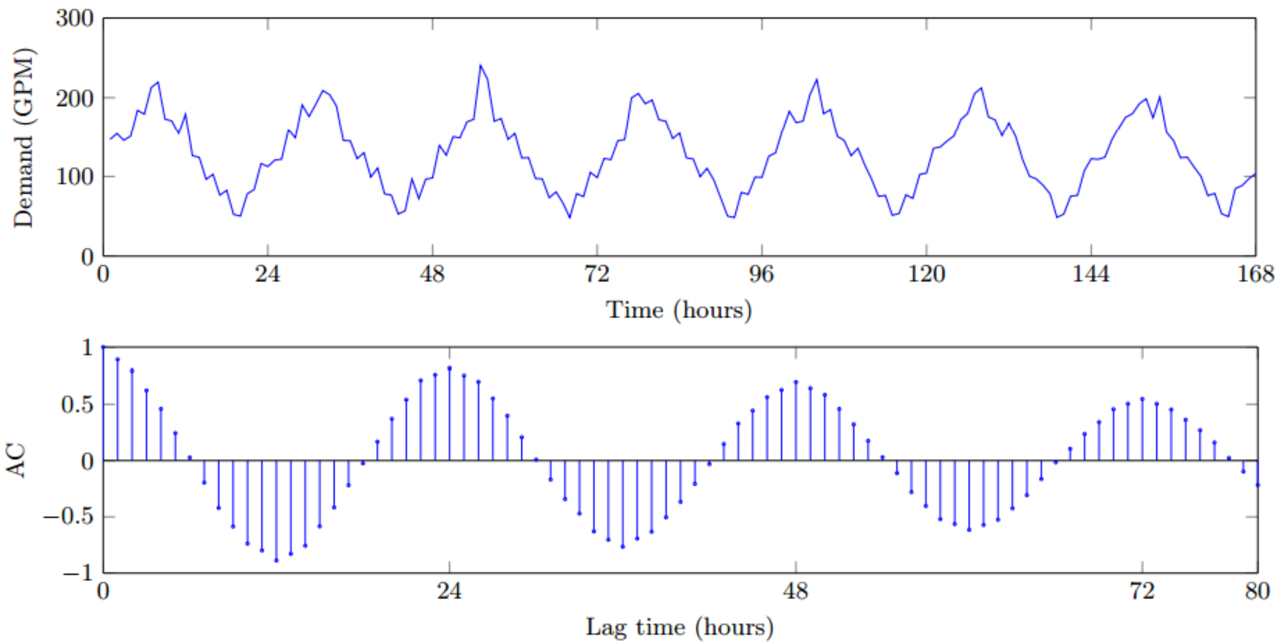
Customer	R^2	MAPE*
Junc. 11	0.88	10.4%
Junc. 12	0.92	8.4%
Junc. 13	0.93	7.1%
Junc. 21	0.91	7.3%
Junc. 22	0.91	8.0%
Junc. 23	0.93	8.1%
Junc. 32	0.94	7.1%

- Demand estimates showing good match for small-to-medium values
- Underestimated the high demands for Junc. 11



Temporal correlations of demand estimates

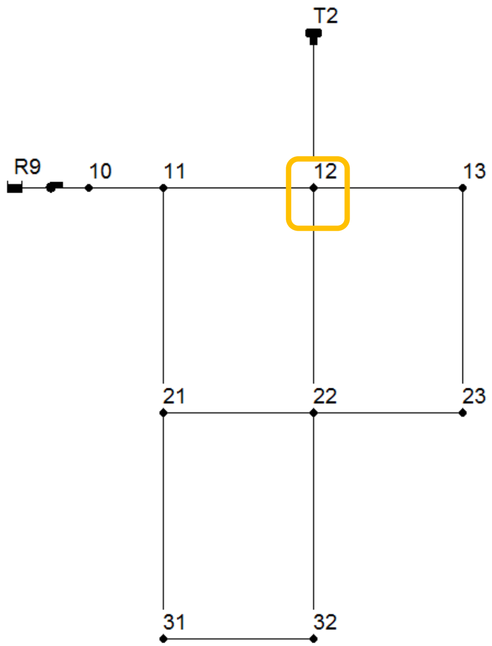
- Junction 11 water demands and autocorrelations

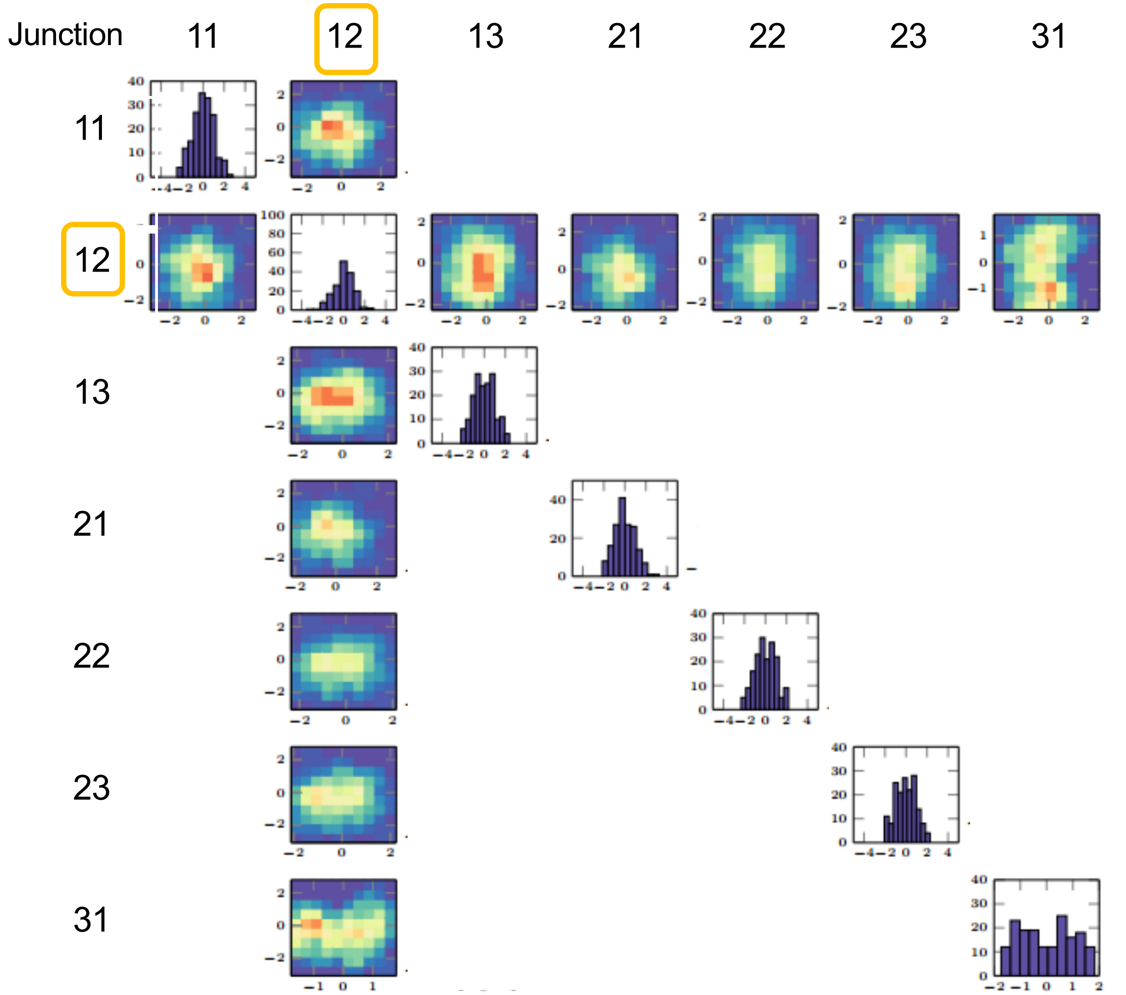


- Structure of autocorrelations similar to previous results on univariate water demands

UNIVERSITY OF Cincinnati

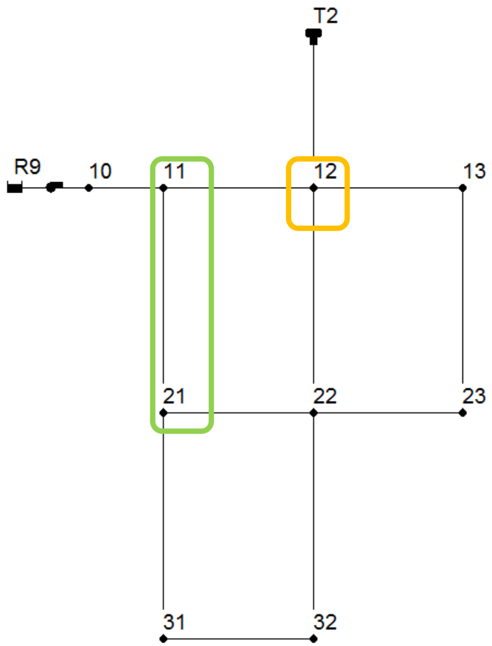
Spatial correlations of demand estimates

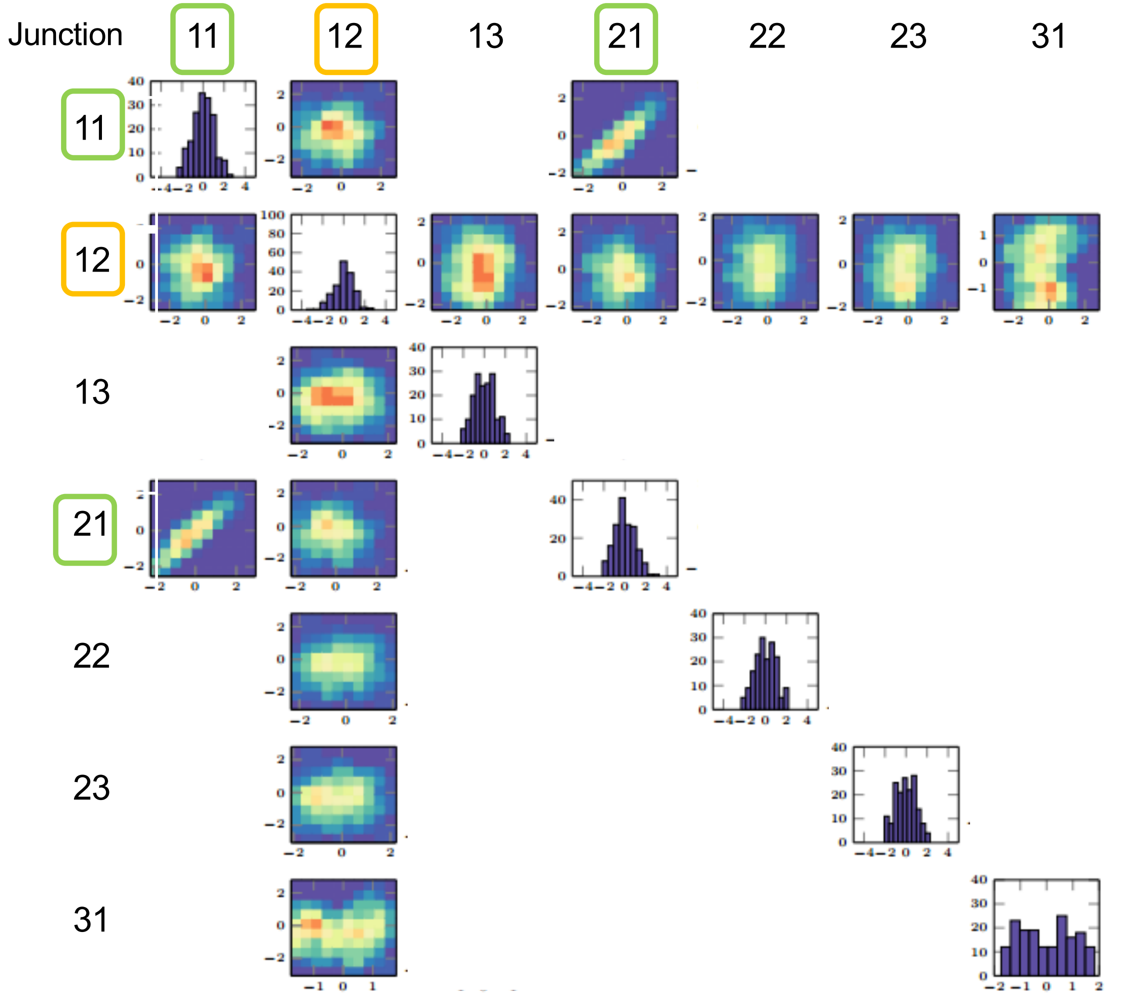




- Junction 12
- Not correlated

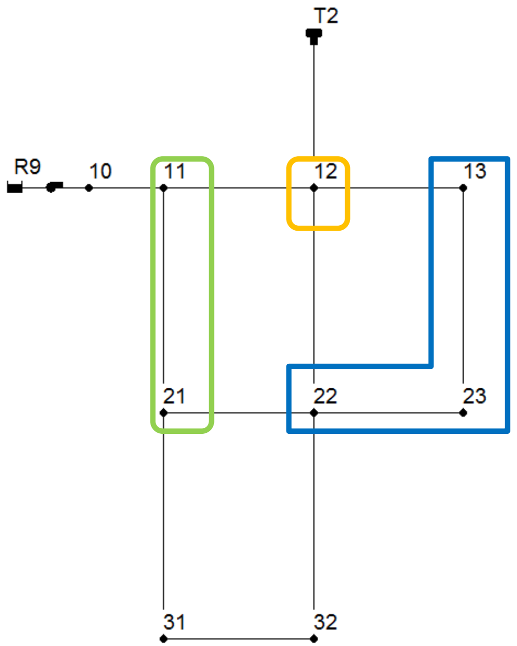
Spatial correlations of demand estimates

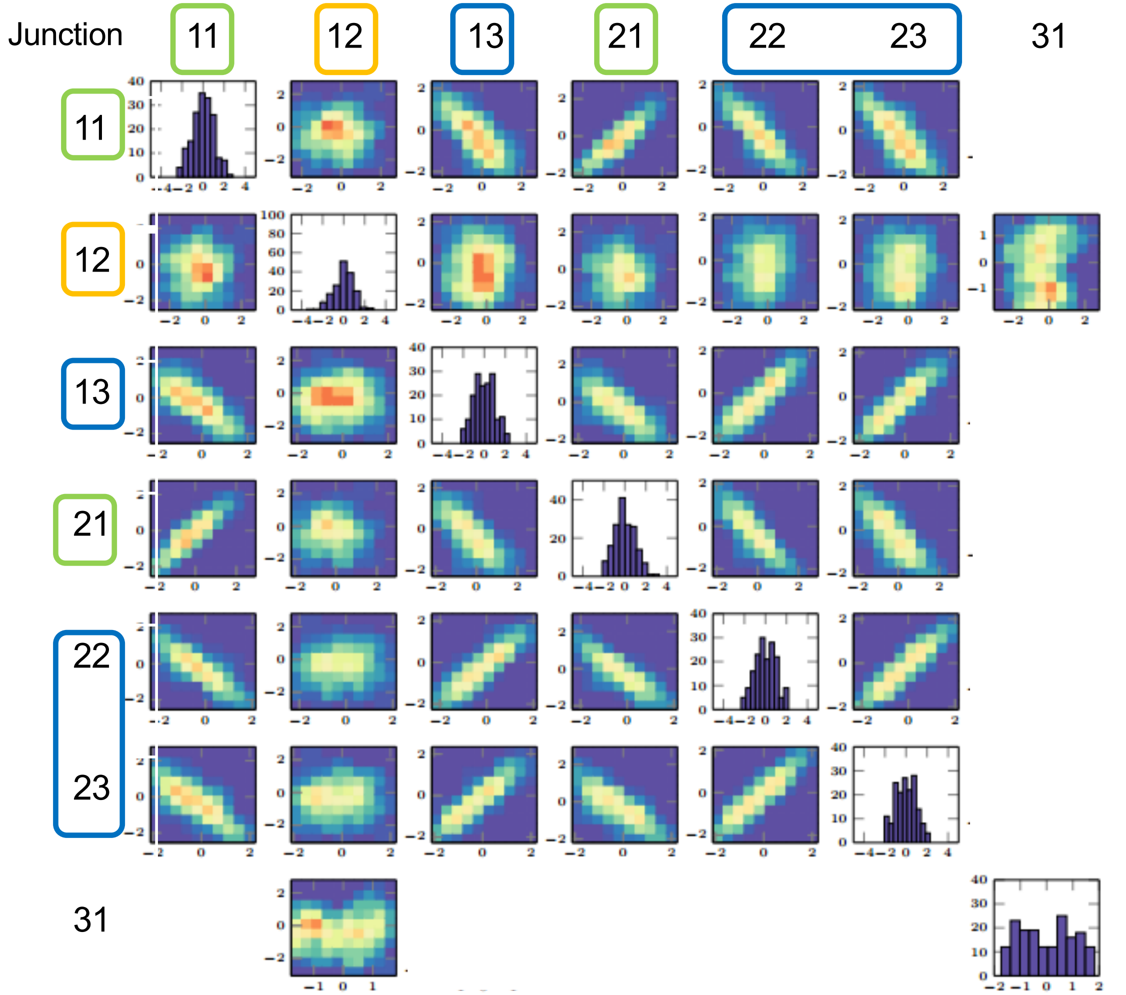




- Junction 11 and 21

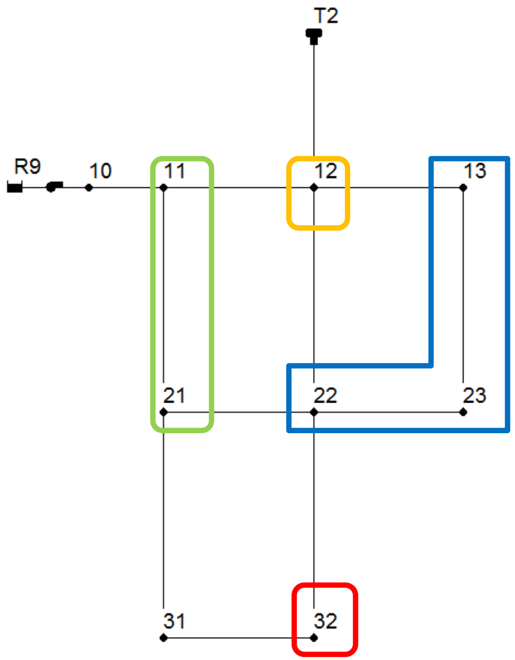
Spatial correlations of demand estimates

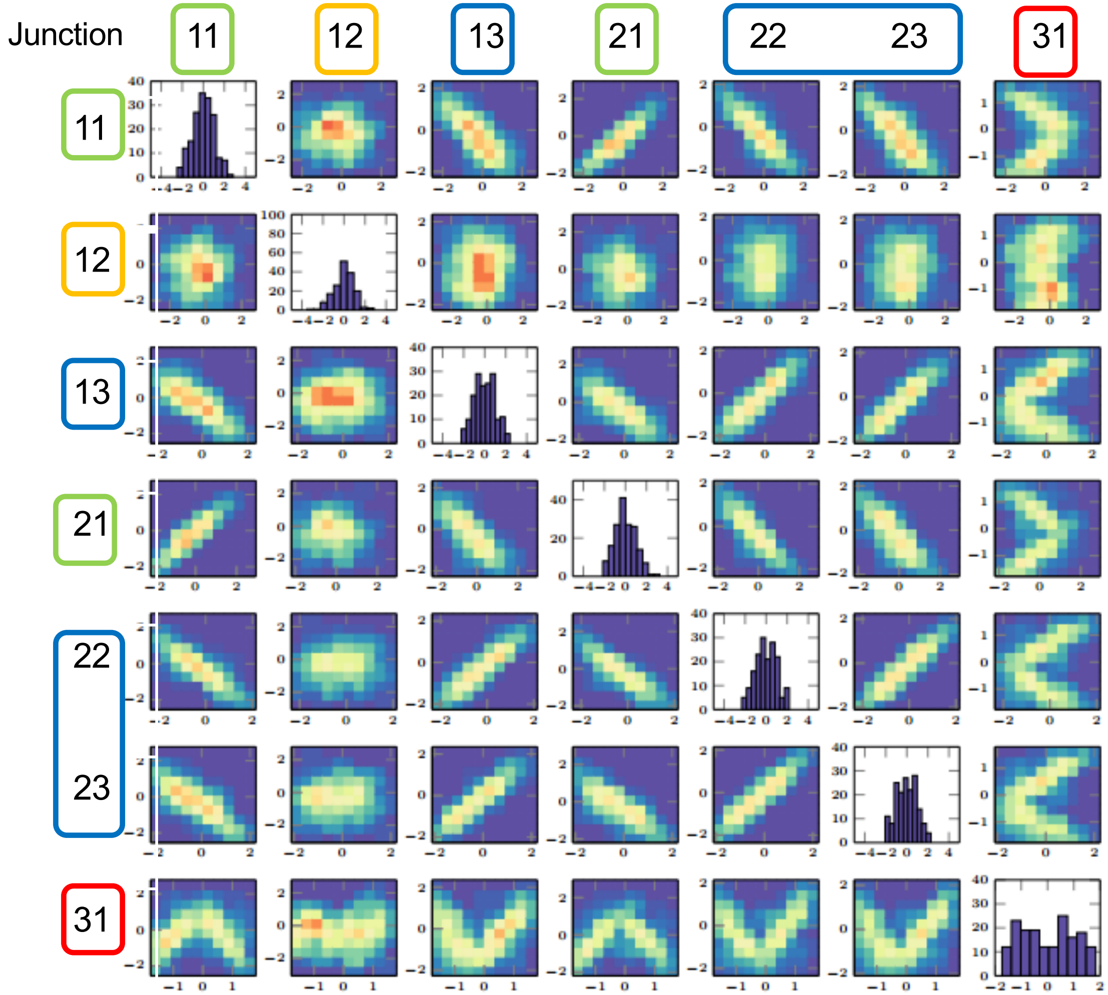




- Junction 13, 22, and 23

Spatial correlations of demand estimates





- Combinational results of intrinsic uncertainty of demands and the layout of SCADA sensors

Conclusions

- The EM algorithm is effective in estimating the parameters and demands in a proof-of-concept study case
- Spatial and temporal correlations of water demands can be quantified
- Lots of computational resources consumed
  - 60-80 minutes to assimilate 1-week worth of SCADA data
  - Applicable for small network
  - Large network may need simplification/consumer grouping

Future work

- Use the demand model with estimated parameters for short-term forecasting
  - Prediction of demands and hydraulics
- Investigate the impact of different layouts of SCADA sensors to the uncertainty of demand estimates
- Potential new method of customer grouping based on spatial correlations
- EM algorithm may be applicable in other problems with the “time series model + non-linear model” structure

Thanks!

Flowchart of the EM algorithm