This presentation premiered at WaterSmart Innovations

watersmartinnovations.com

watersmart
INNOVATIONS

Comparison of Pattern Recognition and Auto Regressive Models for Short-Term Urban Water Demand Forecasting

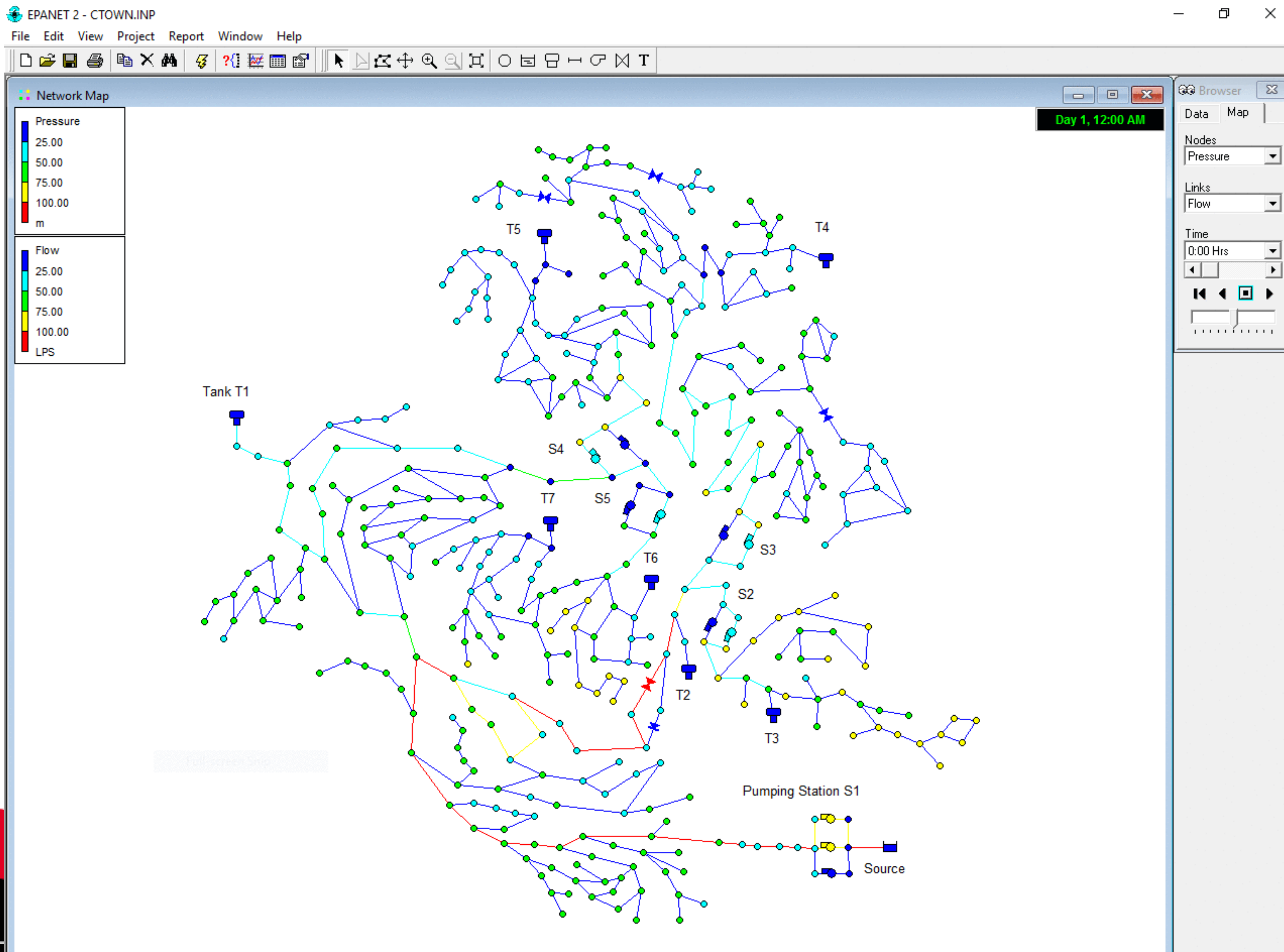
Paulo José Oliveira and Dominic L. Boccelli

University of Cincinnati, Cincinnati OH, USA

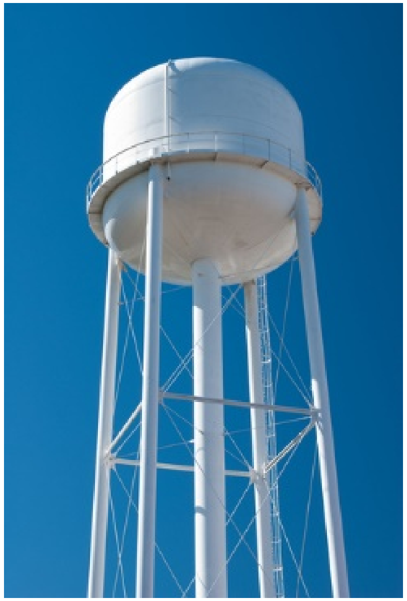
Water distribution operation

- Energy management
- Water quality maintenance
- Response to intentional/accidental intrusion events
- Leak detection

EPAnet simulation



Optimal control



θ : unknown parameter

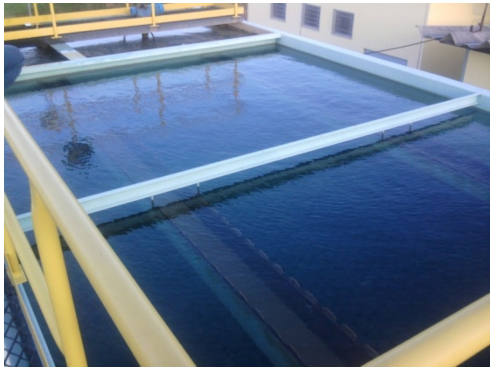
a : one possible action

$L(a, \theta)$: loss function

$\pi(\theta)$: pdf of θ

Expected loss

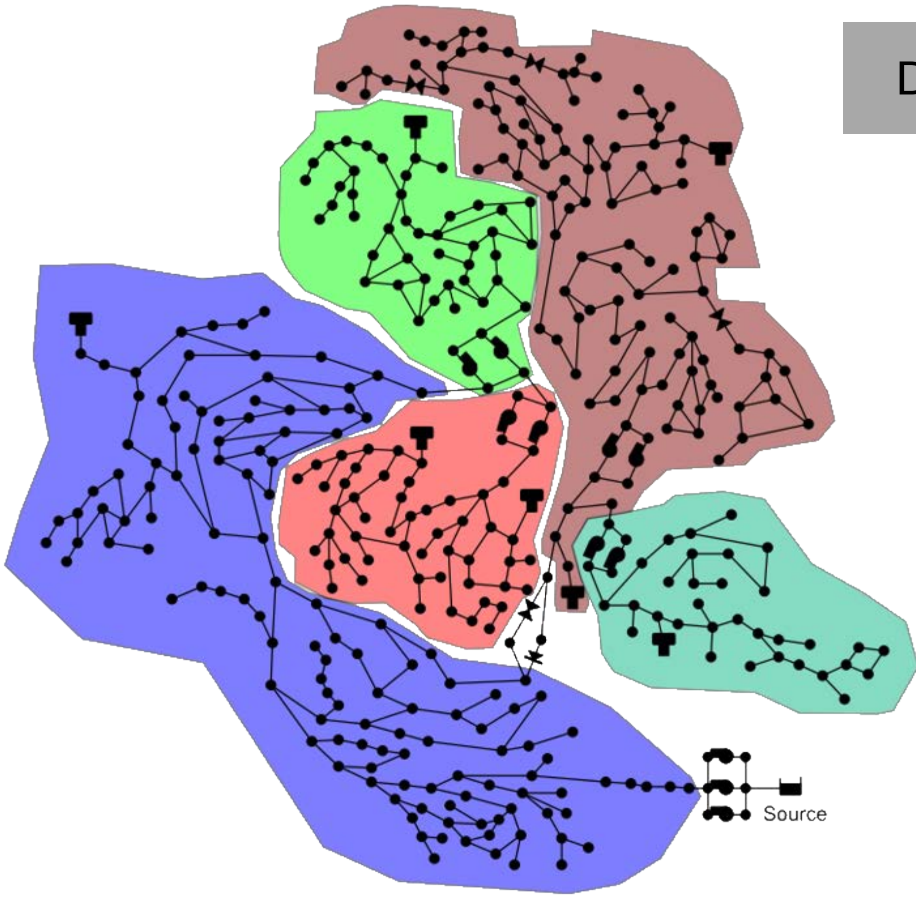
$$r(a) = \int L(a, \theta)\pi(\theta)d\theta$$



Best action

$$a^* = \arg\inf r(a)$$

Real-time challenges



Demand Estimation

Past water demands

Time Series

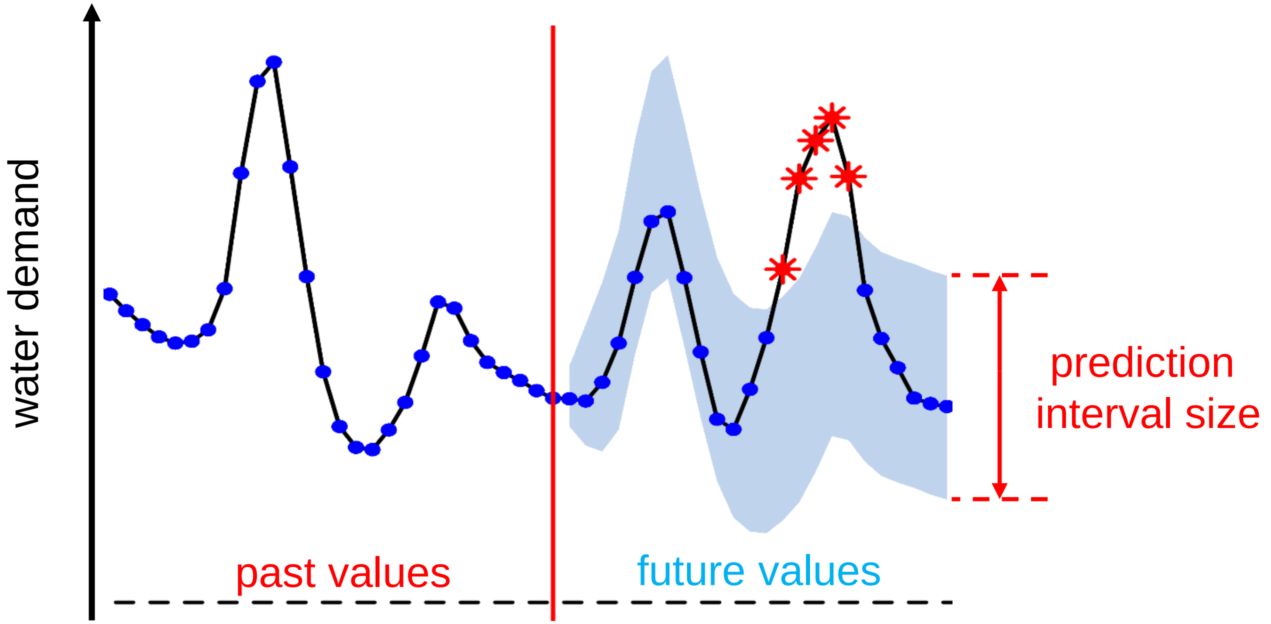
~~Future water demands~~

Optimization

Future demands

Best operational decision

Short-term forecasting (24h)

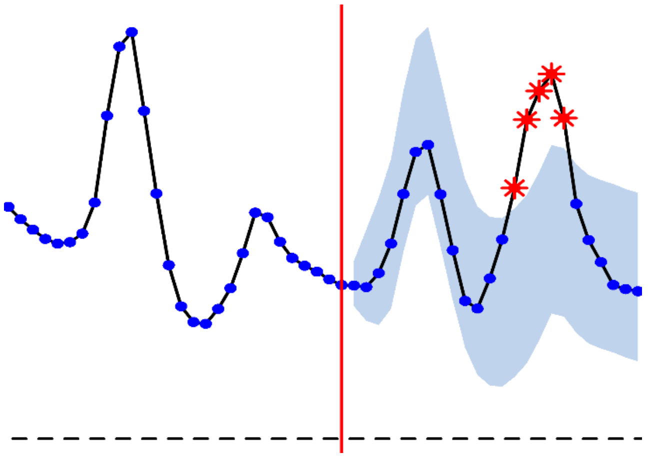


Sharpness: Average size of a given prediction interval

Reliability: Percentage of observations that fall within the forecasted prediction bounds

Short-term forecasting (24h)

- Prediction Accuracy
  - Small sharpness values
- Uncertainty
  - Large
- Explanation
  - Simple
- Adaptive Model



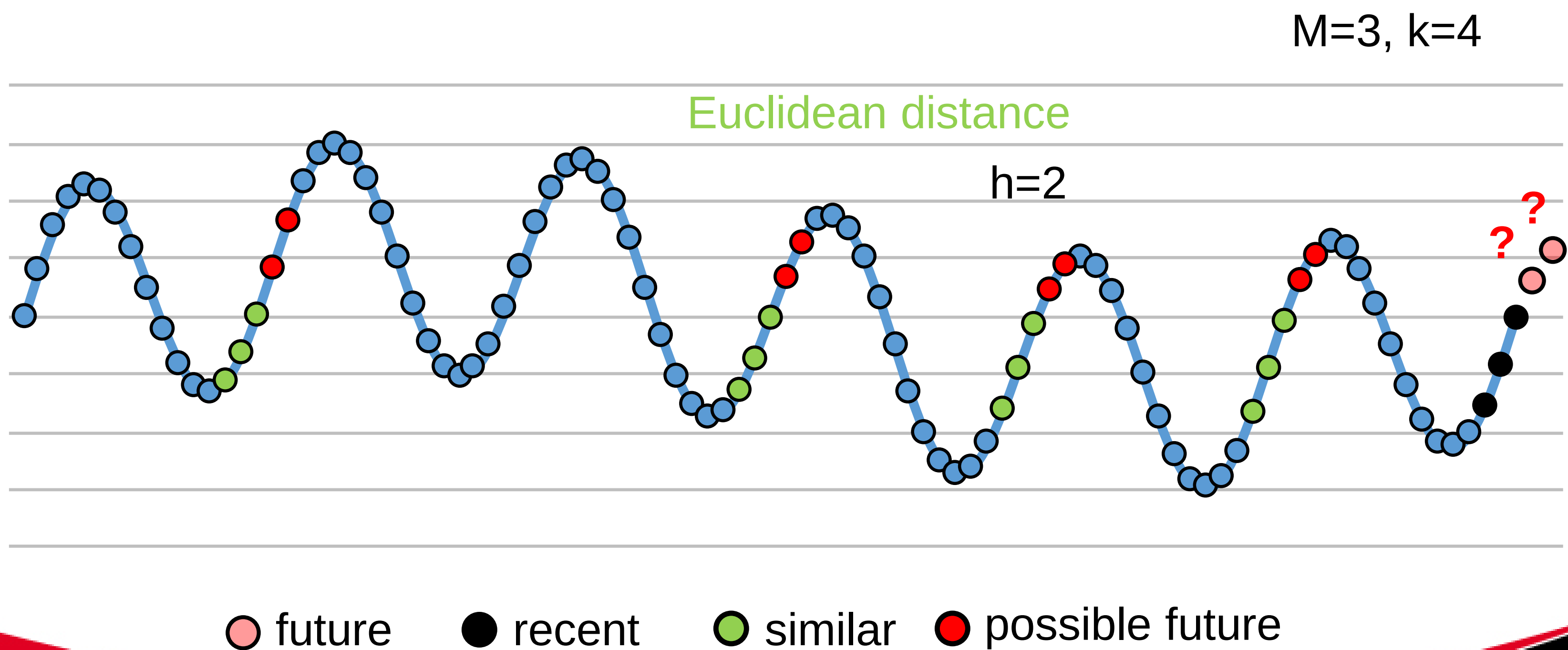
Evaluation of
2 Methods

- knn
- SAR

K-nearest neighbor (KNN)

- Pattern recognition approach that makes the prediction based on the most similar past observations

(Yakowitz 1987)

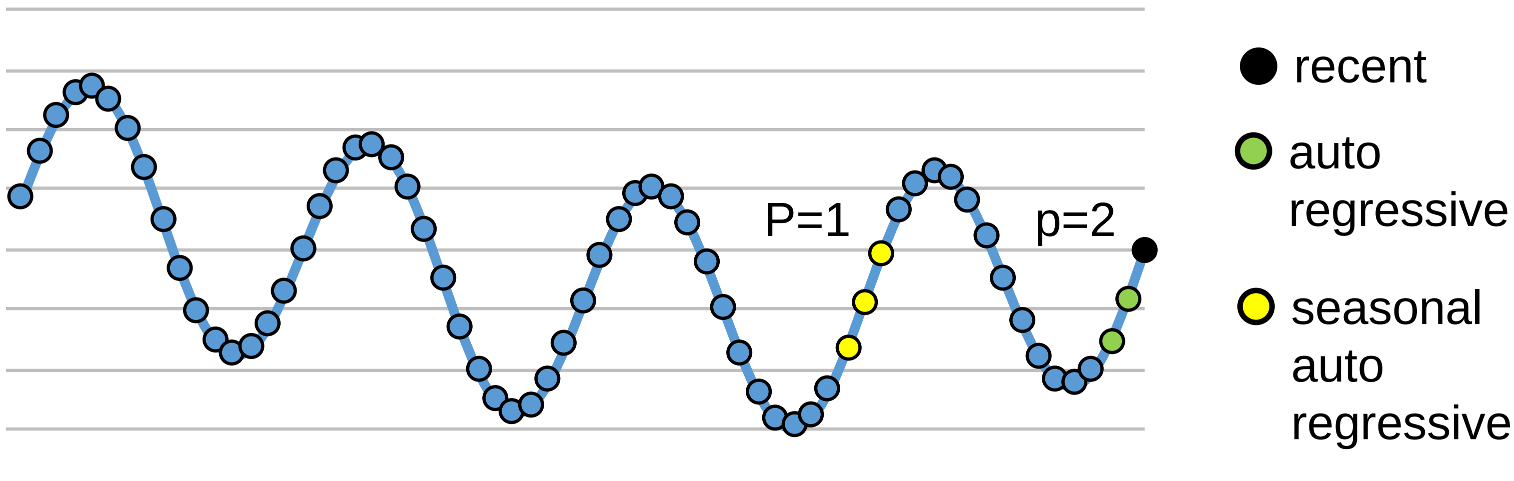


UNIVERSITY OF Cincinnati

Seasonal autoregressive (SAR)

$$\phi_p(B)\Phi_P x_t = a_t$$ (Box and Jenkins)

$$\left(1 - \phi_1 B^1 \ldots - \phi_p B^p\right)\left(1 - \Phi_1 B^S \ldots - \Phi_P B^{PS}\right) x_t = a_t$$



for S=24h

$$(1 - \phi_1 B^1 - \phi_2 B^2)(1 - \Phi_1 B^{24}) x_t = a_t$$

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \Phi_1 x_{t-24} - \phi_1 \Phi_1 x_{t-25} - \phi_2 \Phi_1 x_{t-26} + a_t$$

Prediction Intervals

- SAR

$$\sigma_{\hat{F}} = \left(1 + \sum_{j=1}^{l-1} \psi_j^2\right)^{1/2} s_a$$

(Box and Jenkins)

- knn

$$\sigma_{\hat{F}} = \left\{\frac{1}{k-1} \sum_{j=1}^{k} \left[F(t_j) - \mu_{\hat{F}}\right]^2\right\}^{1/2}$$

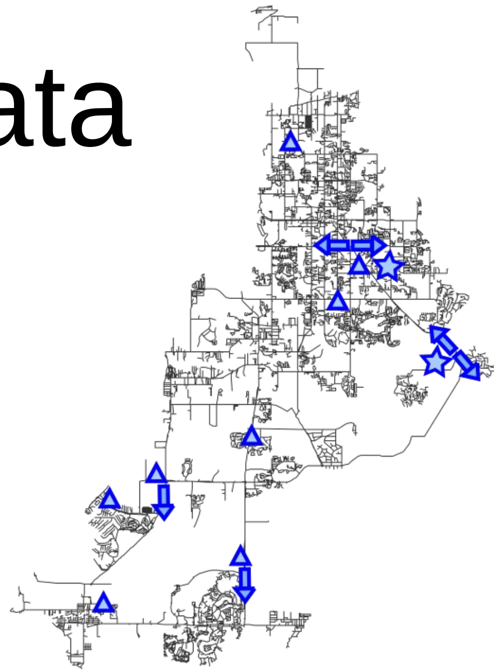
Prediction interval

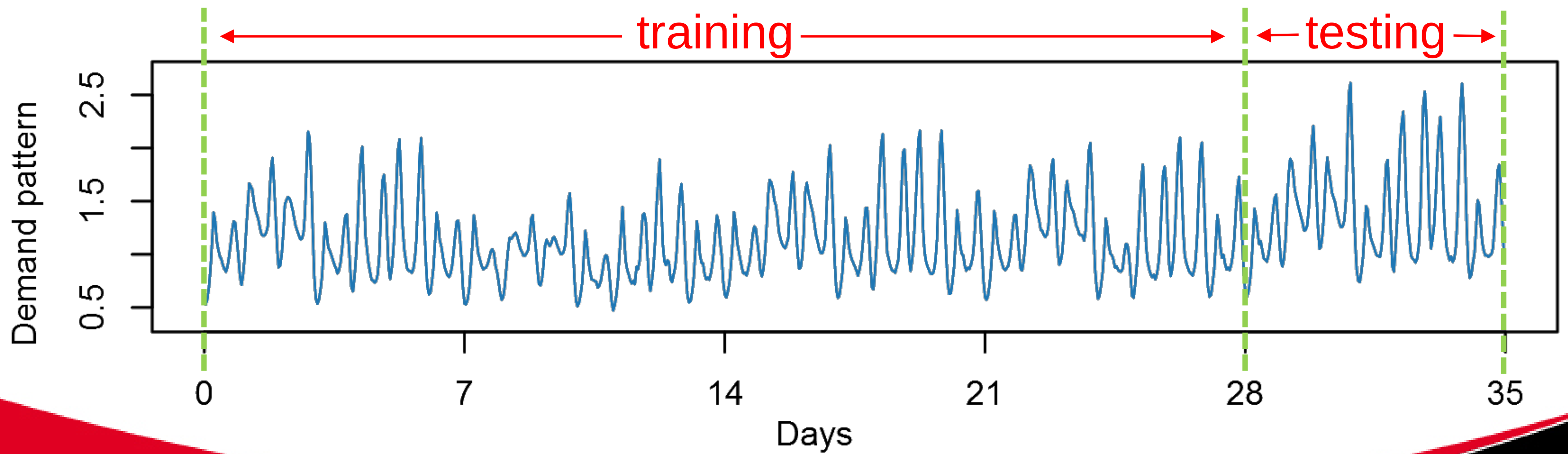
$$x_{\hat{F}} - z_{\alpha/2}\, \sigma_{\hat{F}} < \hat{x} < x_{\hat{F}} + z_{\alpha/2}\, \sigma_{\hat{F}}$$

Normality assumption

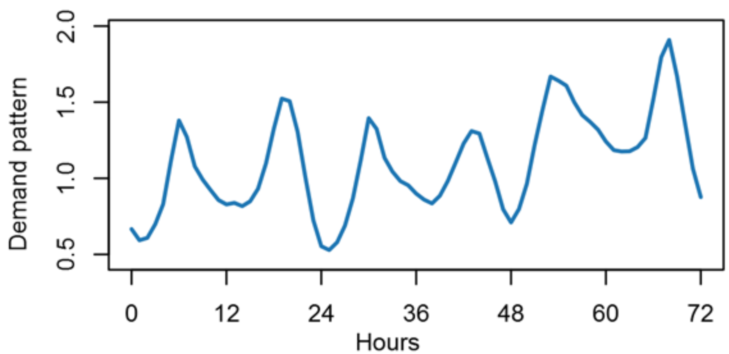
Water demand data

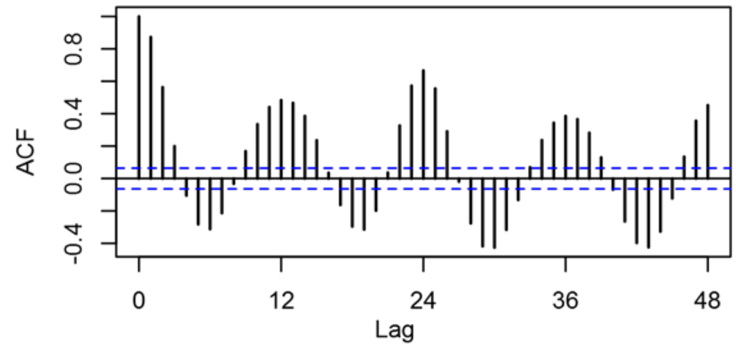
- Global city demand pattern
- 5 weeks of data
  - 4 weeks for training
  - last week for testing





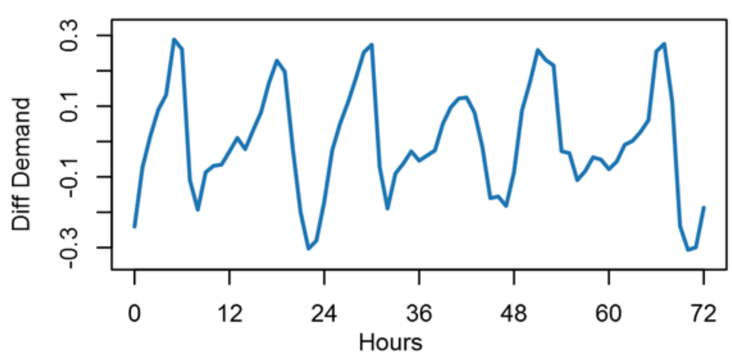
Time series transformation

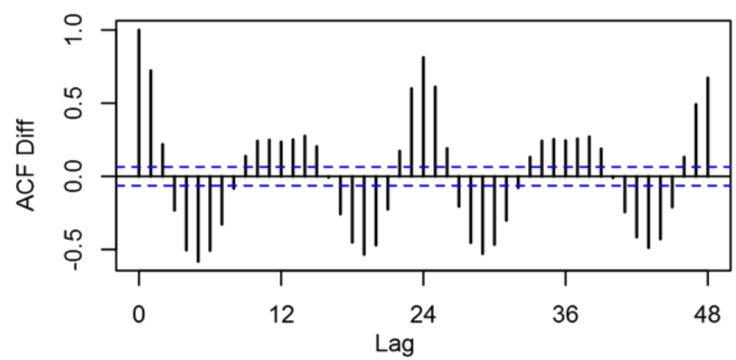




<u>Difference operator</u>

$$\nabla x_t = (1 - B)x_t = x_t - x_{t-1}$$



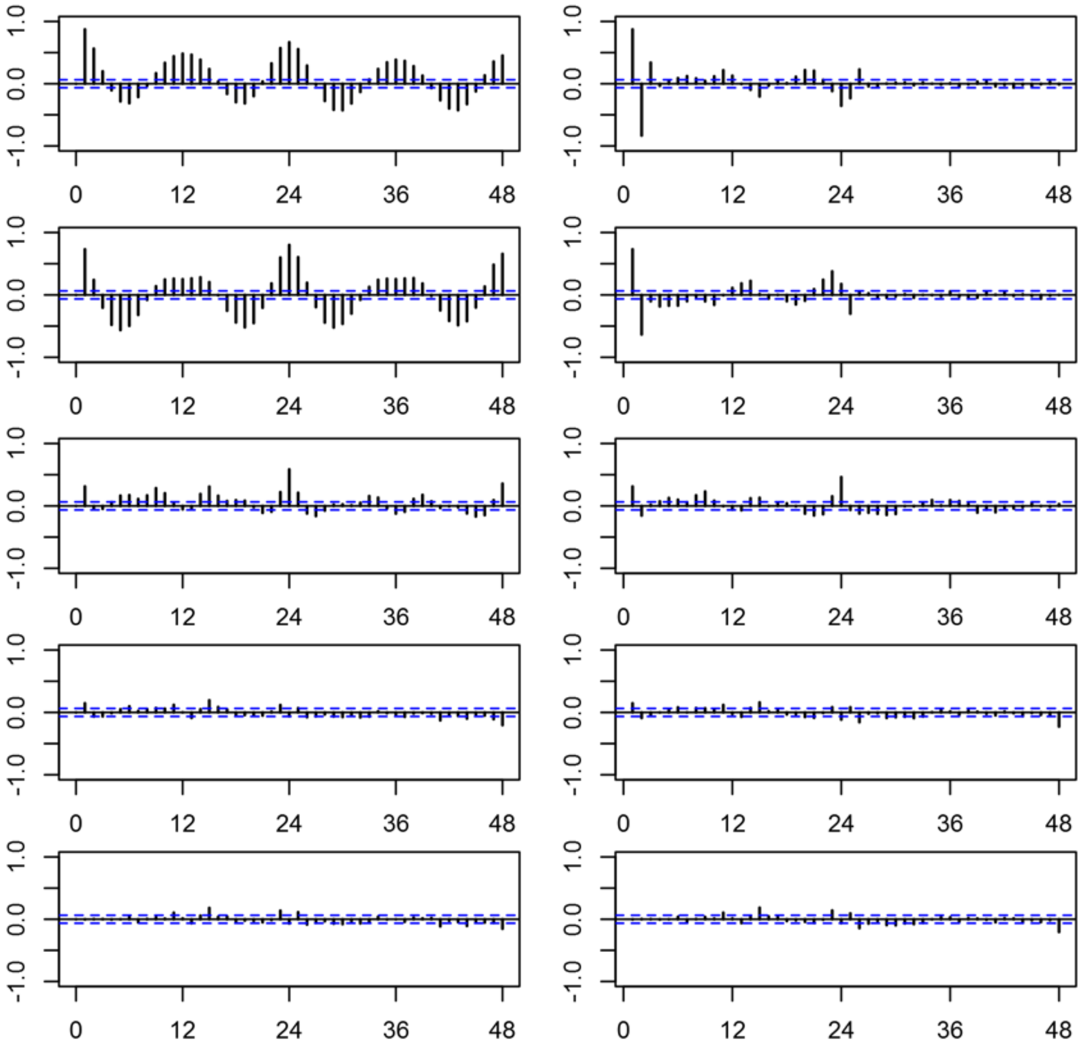


Evaluation steps

1) SAR identification (training)
   - How many AR terms are needed?
   - Stepwise selection based on AIC criteria
2) Knn identification (training)
   - What is the best k value?
   - What is the best M value?
   - Exaustive search for all combinations
   - Choice based on best Sharpness & Reliability
3) Forecasting evaluation (testing)
   - With demand serie (SAR, knn)
   - With differenciated demand serie (SAR_diff, knn_diff)

SAR model building

ACF PACF



Original

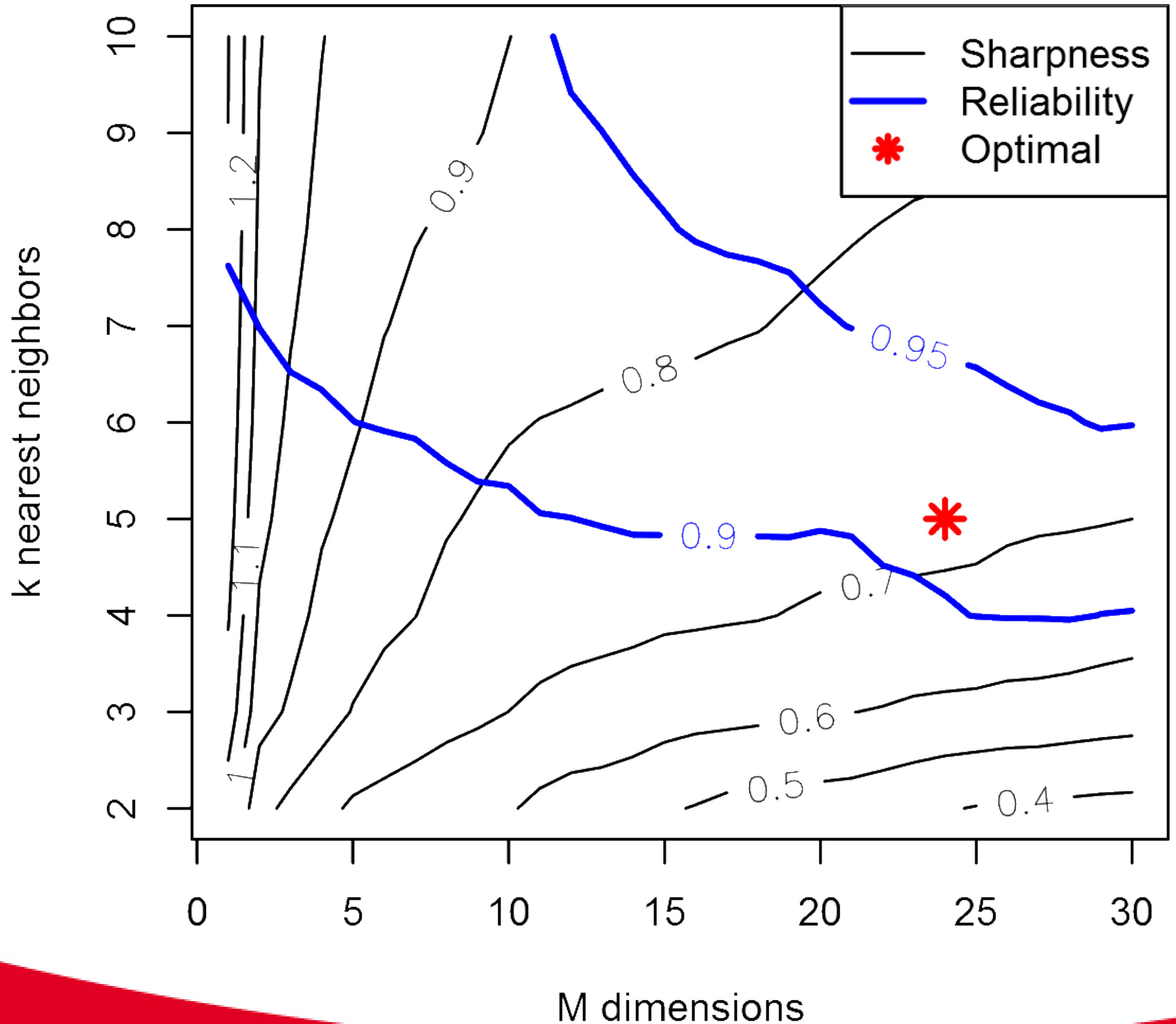
AIC = -2.63 BIC = -3.62 σ^2 = 0.02640
$\phi_1 = 0.87$

AIC = -3.85 BIC = -4.83 σ^2 = 0.00779
$\phi_1 = 1.61$ $\phi_2 = -0.84$

AIC = -4.41 BIC = -5.39 σ^2 = 0.00443
$\phi_1 = 1.58$ $\phi_2 = -0.71$ $\Phi = 0.71$

AIC = -4.45 BIC = -5.43 σ^2 = 0.00424
$\phi_1 = 1.74$ $\phi_2 = -1.07$ $\phi_3 = 0.22$
$\Phi = 0.66$

Knn parameters



Best point

M = 24
K = 5

Forecasting evaluation

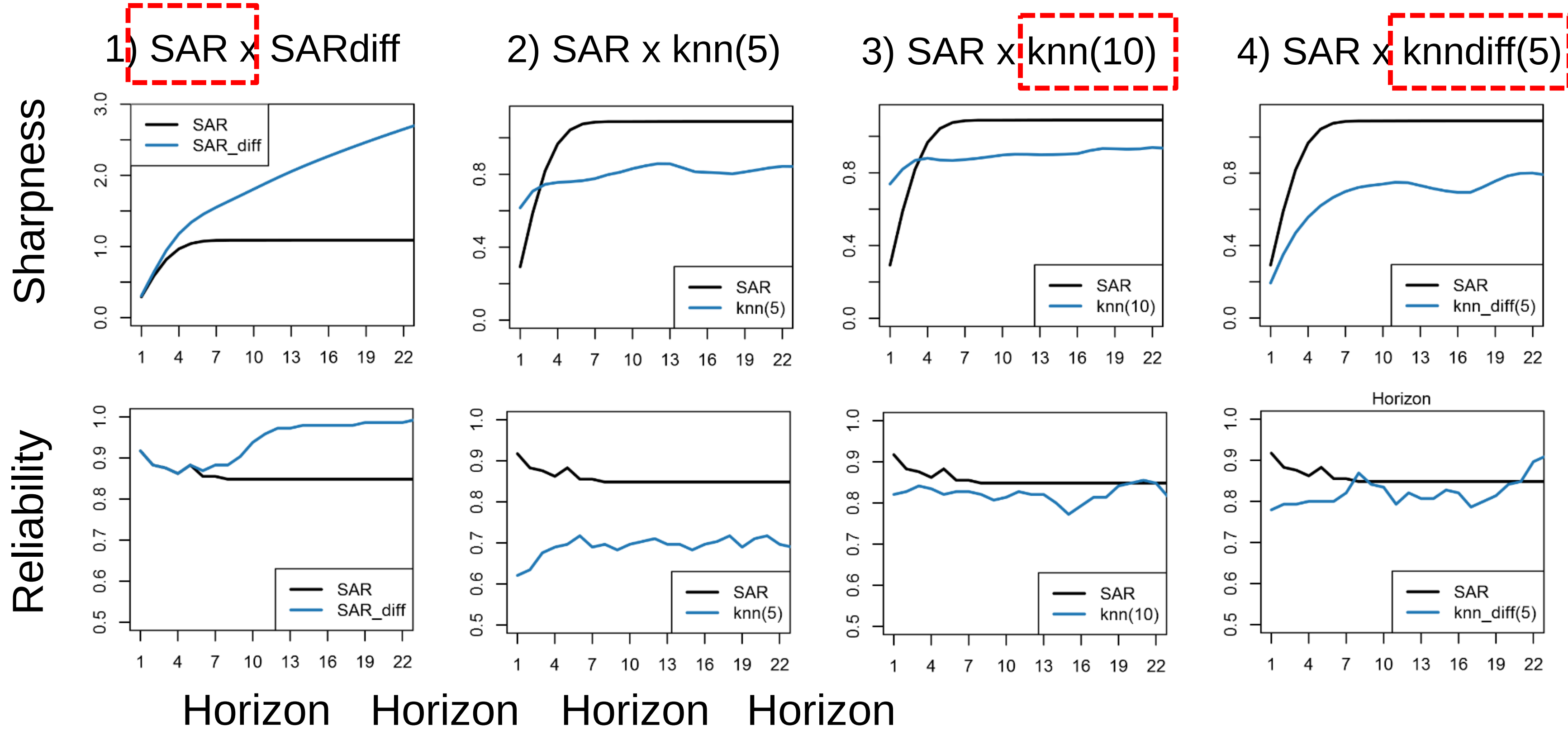
1) Sharpness
2) Reliability
3) MAPE - Mean absolute percentage error

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right|$$

4) RMSE - Root-mean-square error

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{x}_t)^2}$$

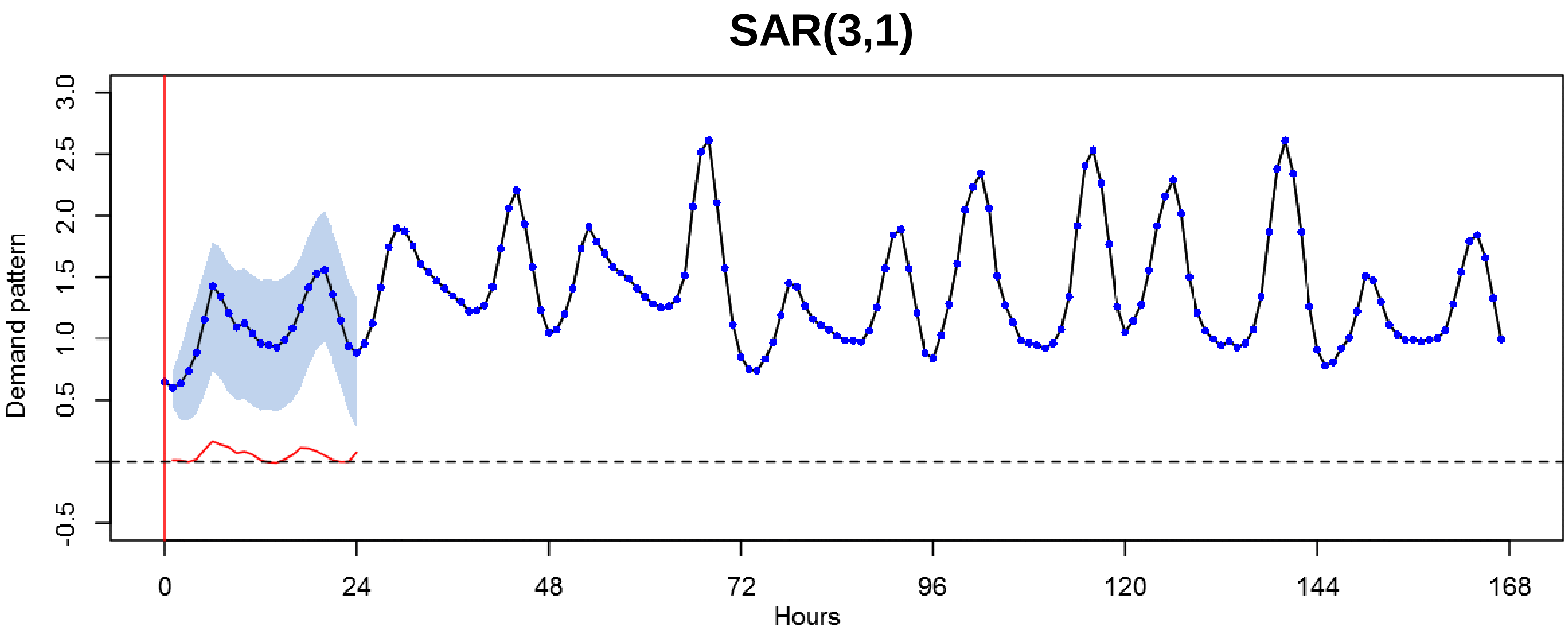
Forecasting comparison



<u>Sharpness</u> => Small

<u>Reliability</u> ≅ 0.95

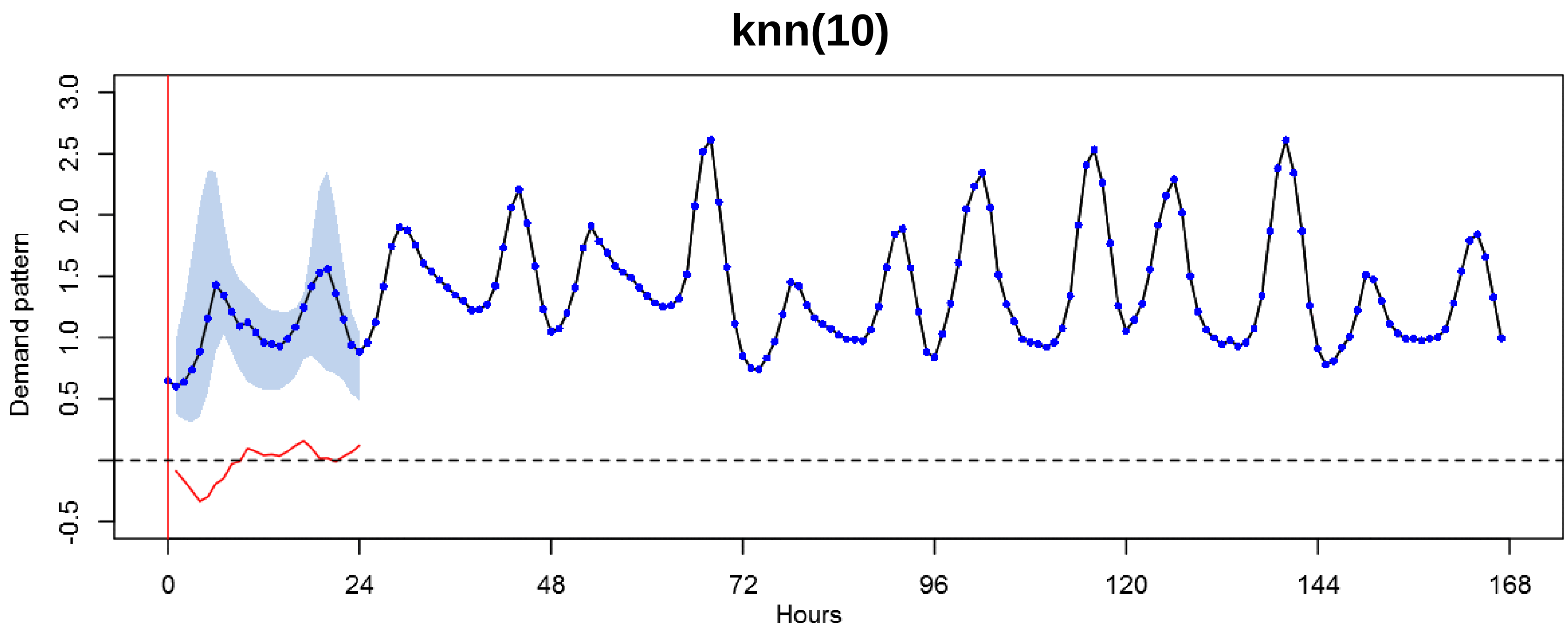
Forecasting comparison



SAR(3,1)

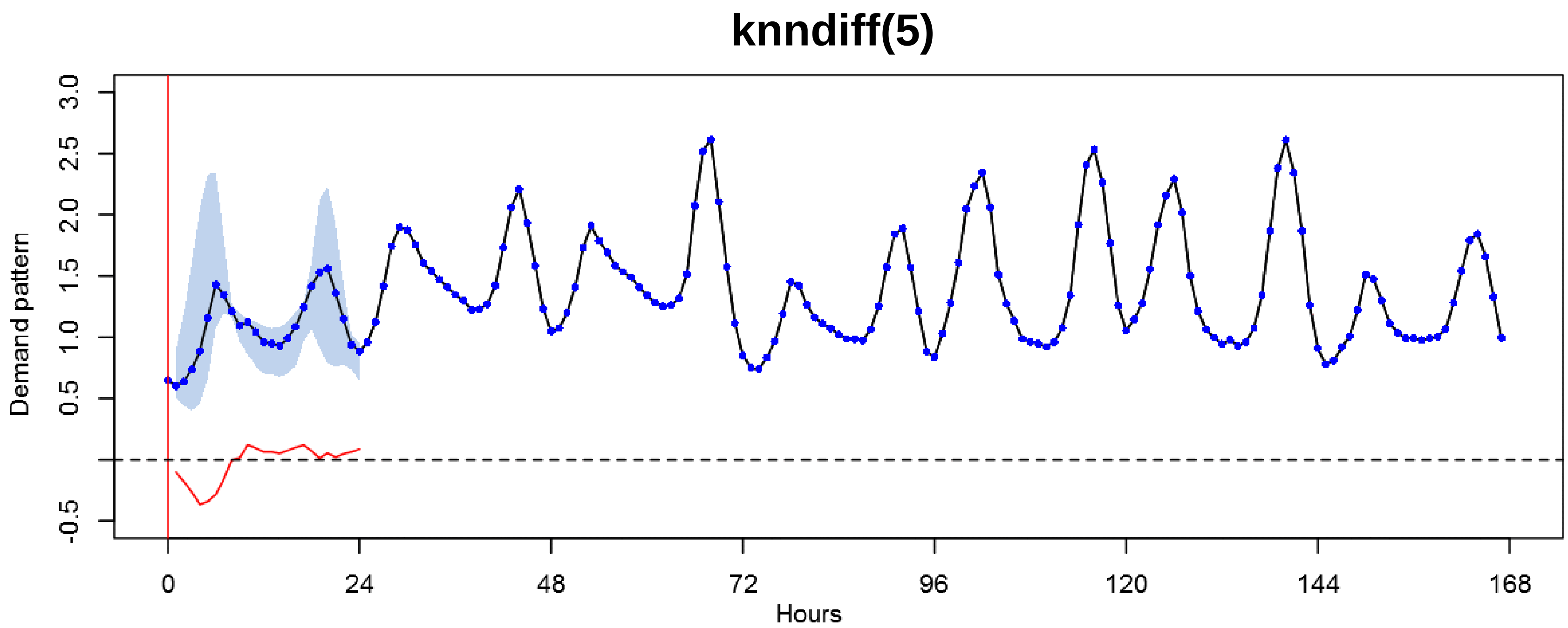
MAPE = 17%
RMSE = 0.34
Sharpness = 1.02
Reliability = 0.86

Forecasting comparison



SAR(3,1)	knn(10)
MAPE = 17%	MAPE = 16%
RMSE = 0.34	RMSE = 0.32
Sharpness = 1.02	Sharpness = 0.89
Reliability = 0.86	Reliability = 0.82

Forecasting comparison



SAR(3,1)	knn(10)	knndiff(5)
MAPE = 17%	MAPE = 16%	MAPE = 13%
RMSE = 0.34	RMSE = 0.32	RMSE = 0.23
Sharpness = 1.02	Sharpness = 0.89	Sharpness = 0.68
Reliability = 0.86	Reliability = 0.82	Reliability = 0.83

Summary and Conclusions

1. Reasonable predictions were obtained by both SAR and knn
2. The knn performance needs to be evaluated according with the amount of available data
3. The SAR, in general, is more stable than the knn which cannot predict values beyond the training dataset
4. The knn identification (k and M) needs to be more carefully evaluated
5. Single lagged differences can be beneficial for the knn which outperformed the SAR for all predicted horizons

Acknowledgements

Partial funding support from

– National Science Foundation CBET (#1511959)

Scholarship

– University Council of Water Resources (UCWR)

Questions?

UNIVERSITY OF Cincinnati