A Case Study on Water Demand Forecasting in a Coastal Tourist City

Antoniel Kleber Stefaniak

Pablo Andretta Jaskowiak

Lucas Weihmann

November 21, 2024

Centro Tecnológico de Joinville (CTJ)

Universidade Federal de Santa Catarina (UFSC)

Joinville – Santa Catarina – Brasil

PPGESE

UFSC

Outline

- Motivation
- Case Study
- Experimental Setup
- Results and Discussion
- Final Remarks

Motivation

- Water demand is succeptible to variations
  - Economic, demographic and climatic

- Different forecasting windows and goals
  - Short-term (hours, days)
    - Optimize water treatment / reservouir levels
  - Long-term (months, years, decades)
    - Infrastructure Planning

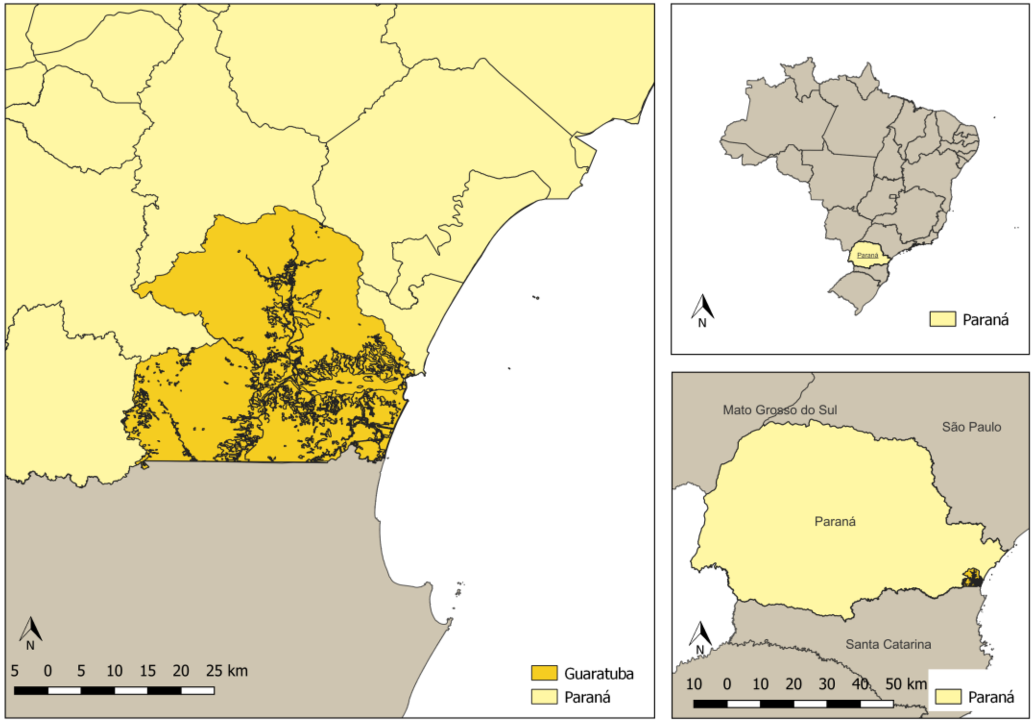
Motivation

- In touristic cities such a planning becomes more urgent
- We evaluated statistical methods (ARIMA / SARIMA)
  - One week ahead, daily forecasts
- Comparison with previous study based on regression models
- Related work employed ARIMA/SARIMA as baselines, mostly with auto tunned parameters (which we believe is not fair)

Vieira et al.; Machine learning for water demand forecasting: Case study in a Brazilian coastal city. Water Practice and Technology 1 May 2024; 19 (5): 1586–1602. doi: https://doi.org/10.2166/wpt.2024.096

Case Study

- Guaratuba - Paraná



Vieira et al.; Machine learning for water demand forecasting: Case study in a Brazilian coastal city. Water Practice and Technology 1 May 2024; 19 (5): 1586–1602. doi: https://doi.org/10.2166/wpt.2024.096

PARANÁ GOVERNO DO ESTADO | AGÊNCIA ESTADUAL DE NOTÍCIAS

Inicial | Últimas notícias | Rádio | Vídeos | Imagens | Editorias | Arquivo de Notícias | Séries Especiais | Contato

1 milhão de pessoas passaram a virada no Litoral do Paraná, estima Polícia Militar

As praias do Litoral receberam cerca de 1 milhão de pessoas na virada de 2023 e 2024, segundo estimativas de público da Polícia Militar do Paraná divulgadas nesta segunda-feira (1). A prainha de água doce de Porto Rico, na Costa Noroeste, recebeu aproximadamente 30 mil turistas na virada do ano.

Publicação
01/01/2024 - 13:50

Editoria
Verão Maior Paraná

COMPARTILHE:

Confira o áudio desta notícia

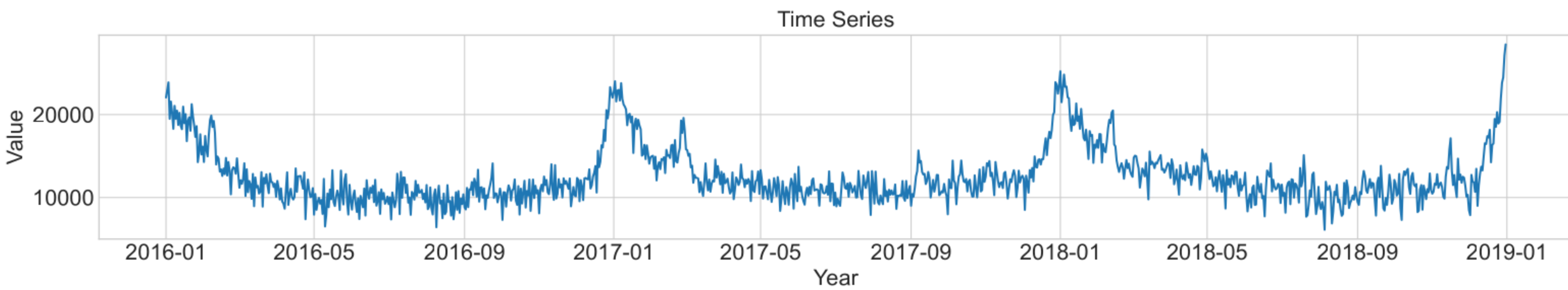
0:00 / 3:17

1 milhão de pessoas passaram a virada no Litoral do Paraná, estima Polícia Militar
Foto: PMPR

https://www.aen.pr.gov.br/Noticia/1-milhao-de-pessoas-passaram-virada-no-Litoral-do-Parana-estima-Policia-Militar

Case Study

- Daily water consumption in m^3



- Four year dataset from SANEPAR (2016 – 2019)

- We actually considered produced water, not consumed
  - Correlated, but not the same

Case Study – Exploratory Analysis

Determination of ARIMA and SARIMA parameters

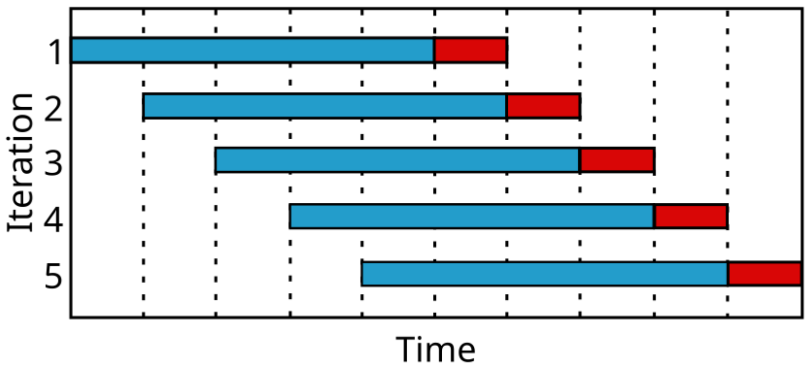
- Performed time-series decomposition
- Autocorrelation Function (ACF)
- Partial Autocorrelation Function (PACF)

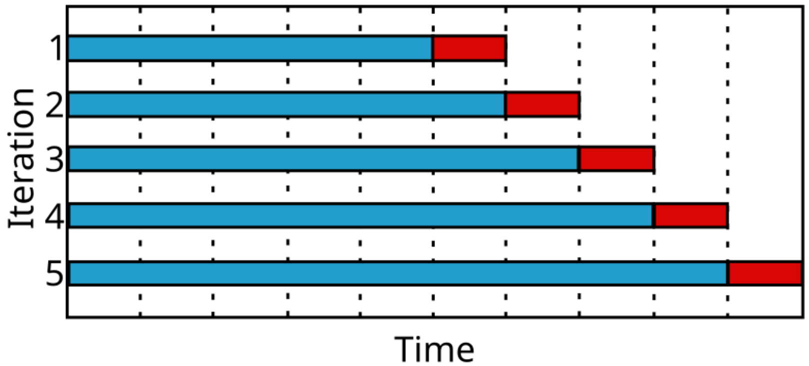
Experimental Setup

- Data was split into
  - Train (2016 and 2017)
  - Validation (2018)
  - Test (2019)

Model	Hyperparameter	Search Space
ARIMA	p	(2, 3, 8)
	d	(0, 1)
	q	(3, 4, 28, 38)
SARIMA	p	(2, 3, 6)
	d	(0, 1)
	q	(3, 4, 6)
	P	(2, 3)
	D	(0, 1)
	Q	(1, 3, 4)
	s	(7)

Experimental Setup





Window Type	Window Size		Sliding Steps
	Train	Test	
Sliding Window (SW1Y)	365 days	7 days	7 days
Sliding Window (SW2Y)	730 days	7 days	7 days
Expanding Window (EW)	*	7 days	7 days

Results – Model Selection

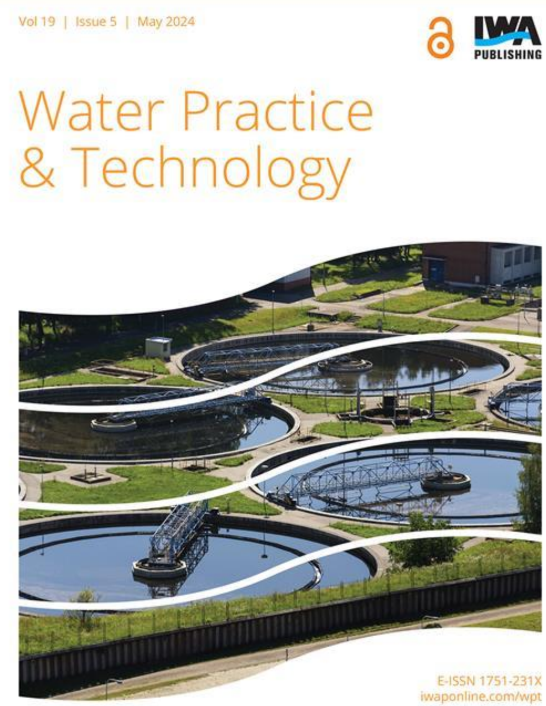
- Regarding the validation set
- For each method, we observed
  - Best / Average / Worst
- Check the whether parameter tunning matters
- Best model selected for further comparison

Results – Model Selection

		Model	RMSE $\pm\ \sigma$	MAE $\pm\ \sigma$	MAPE(%) $\pm\ \sigma$(%)
EW	ARIMA	Best (8,0,38)	1480.51 ± 805.15	1261.14 ± 711.8	9.91±4.24
		Worst (2,0,3)	1955.36 ± 1172.69	1682.67 ± 1091.59	12.96±5.88
		Mean (All)	1741.45 ± 926.27	1494.52 ± 828.09	11.81±5.61
	SARIMA	Best (6,0,6),(2,0,3,7)	1618.21 ± 893.33	1353.34 ± 754.17	10.66±4.9
		Worst (2,1,3),(2,1,4,7)	2032.53 ± 1063.01	1786.65 ± 953.22	14.19±7.14
		Mean (All)	1776.79 ± 935.03	1525.95 ± 806.72	12.09 ±5.51
SW1Y	ARIMA	Best (8,1,38)	1606.31 ± 1081.16	1385.81 ± 959.71	10.78±5.96
		Worst (2,0,4)	1923.57 ± 986.1	1651.42 ± 919.82	13.05 ±5.59
		Mean (All)	1764.58 ± 1022.13	1517.97 ± 909.44	11.91 ±5.92
	SARIMA	Best (6,1,3),(3,0,4,7)	1630.01 ± 1070.06	1374.61 ± 932.09	10.61 ± 5.4
		Worst (3,1,3),(3,1,1,7)	2053.38 ± 974.55	1761.8 ± 852.4	13.97 ± 6.01
		Mean (All)	1810.17 ± 1027.26	1546.68 ± 868.25	12.15 ± 5.51
SW2Y	ARIMA	Best (8,0,38)	1425.35 ± 918.88	1217.98 ± 807.72	9.50 ± 5.08
		Worst (2,0,3)	1994.85 ± 1216.67	1722.07 ± 1140.02	13.17 ± 6.08
		Mean (All)	1737.82 ± 961.27	1494.01 ± 857.24	11.78 ± 5.76
	SARIMA	Best (6,1,4),(3,0,4,7)	1601.85 ± 885.79	1376.80 ± 772.35	10.8 ± 5.13
		Worst (3,1,3),(3,1,1,7)	2023.58 ± 1089.58	1760.96 ± 943.46	13.90 ± 6.6
		Mean (All)	1785.27 ± 958.85	1533.13 ± 824.63	12.14 ± 5.6

Results – Test Set

- Evaluation on 2019 data
  - Not seen by any of the models before



- Compared w.r.t. Vieira et al. 2024
  - Best ML methods

- In both studies, same windows, data*, pre-processing

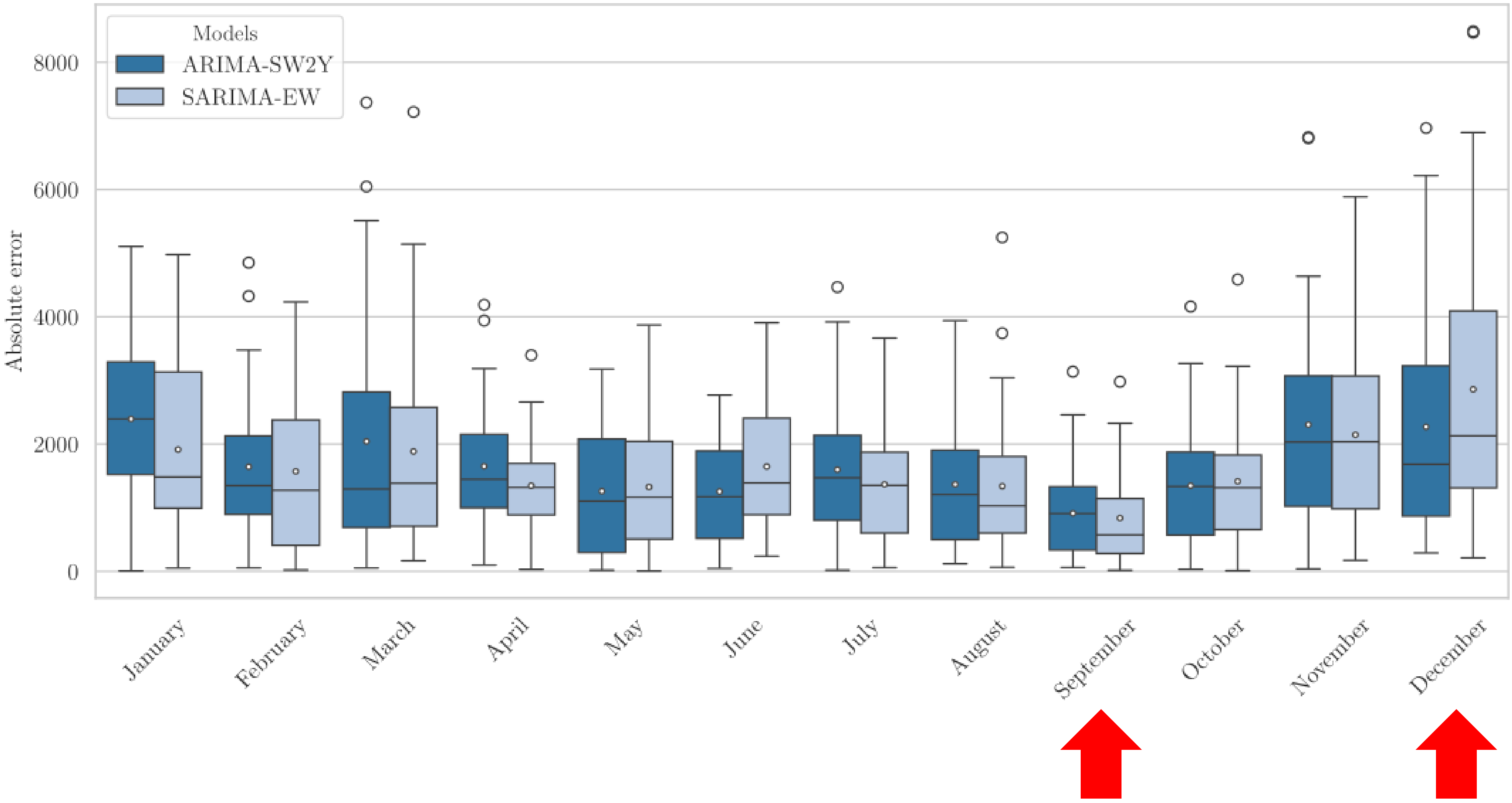
Vieira et al.; Machine learning for water demand forecasting: Case study in a Brazilian coastal city. Water Practice and Technology 1 May 2024; 19 (5): 1586–1602. doi: https://doi.org/10.2166/wpt.2024.096

Results – Test Set

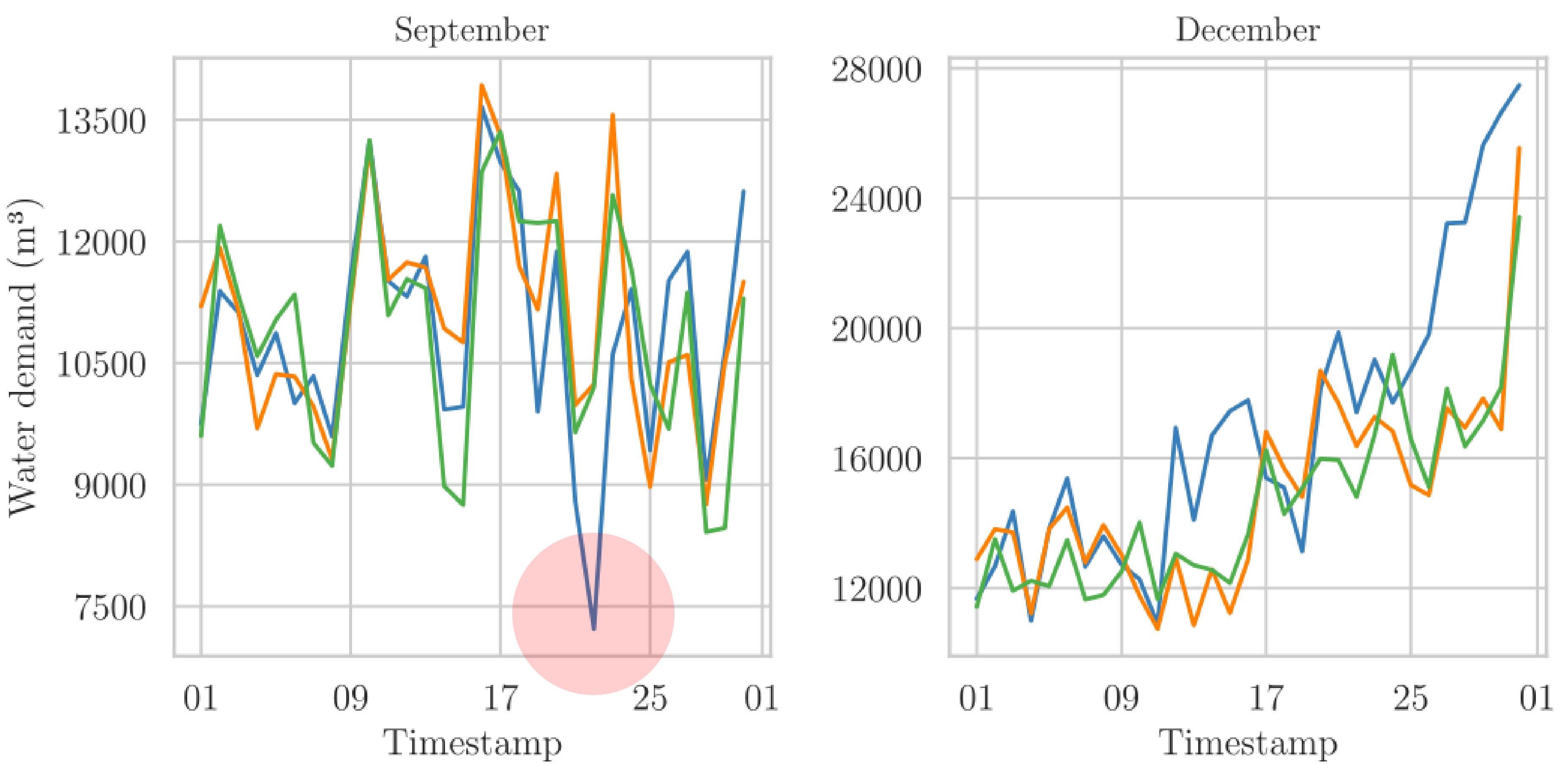
Model	Evaluation Metrics					
	RMSE	σ	MAE	σ	MAPE(%)	σ(%)
ARIMA-EW	1771.39	1044.06	1523.65	928.79	11.99	6.38
SARIMA-EW	**1870.03**	962.6	**1630.14**	893.12	**12.81**	5.9
MLP-EW	1709.70	593.104	1408.97	515.158	11.81	4.707
ARIMA-SW1Y	1745.08	1165.27	1499.44	999.01	11.52	6.21
SARIMA-SW1Y	2096.22	1096.88	1799.32	991.68	14.0	6.12
KNN-SW1Y	1906.99	969.900	1618.19	954.202	12.47	5.055
ARIMA-SW2Y	**1652.67**	1078.71	**1419.4**	975.04	**11.15**	6.65
SARIMA-SW2Y	1947.23	884.83	1667.8	807.33	13.04	4.93
MLP-SW2Y	1927.41	732.005	1620.15	630.532	13.08	4.916

Best ARIMA and SARIMA models in bold

Results – Test Set



Results – Test Set



Final Remarks

- Statistical models provided sound results

- If these are employed as baselines
  - A rigorous analysis of parameters needed

- Previous study from Vieira et al. 2024
  - Regression on same time-series + meteorological data

- Future work considering LSTM and ensemble models

Thanks for the attention

Questions?

pablo.andretta@ufsc.br
pajaskowiak.github.io