IJCNN 2021
The International Conference on Neural Networks
Virtual Conference 18-22 July 2021

Seasonal-trend and multi-objective ensemble learning model for water consumption forecasting

M.H.D.M. Ribeiro, R.G. da Silva, J.H.K. Larcher, J. D. de Lima, V.C. Mariani, L.S. Coelho

Industrial & Systems Engineering Graduate Program (PPGEPS)
Pontifical Catholic University of Parana (PUCPR), Parana, Brazil

SCIENTIA, VITA ET FIDES
PUCPR
GRUPO MARISTA

Agenda

- Introduction
- Objectives and Contribution
- Datasets
- Methodology
- Results
- Conclusion
- References

Source: https://www.fostercity.org/publicworks/page/water-consumption-customer-portal

Introduction

- Water demand plays a key role in the urban planning;
- Seasonal and trend are the main features of water demand;
- These features make the forecasting process hard;
- Ensemble methods can deal with it and enhance the forecasting accuracy.

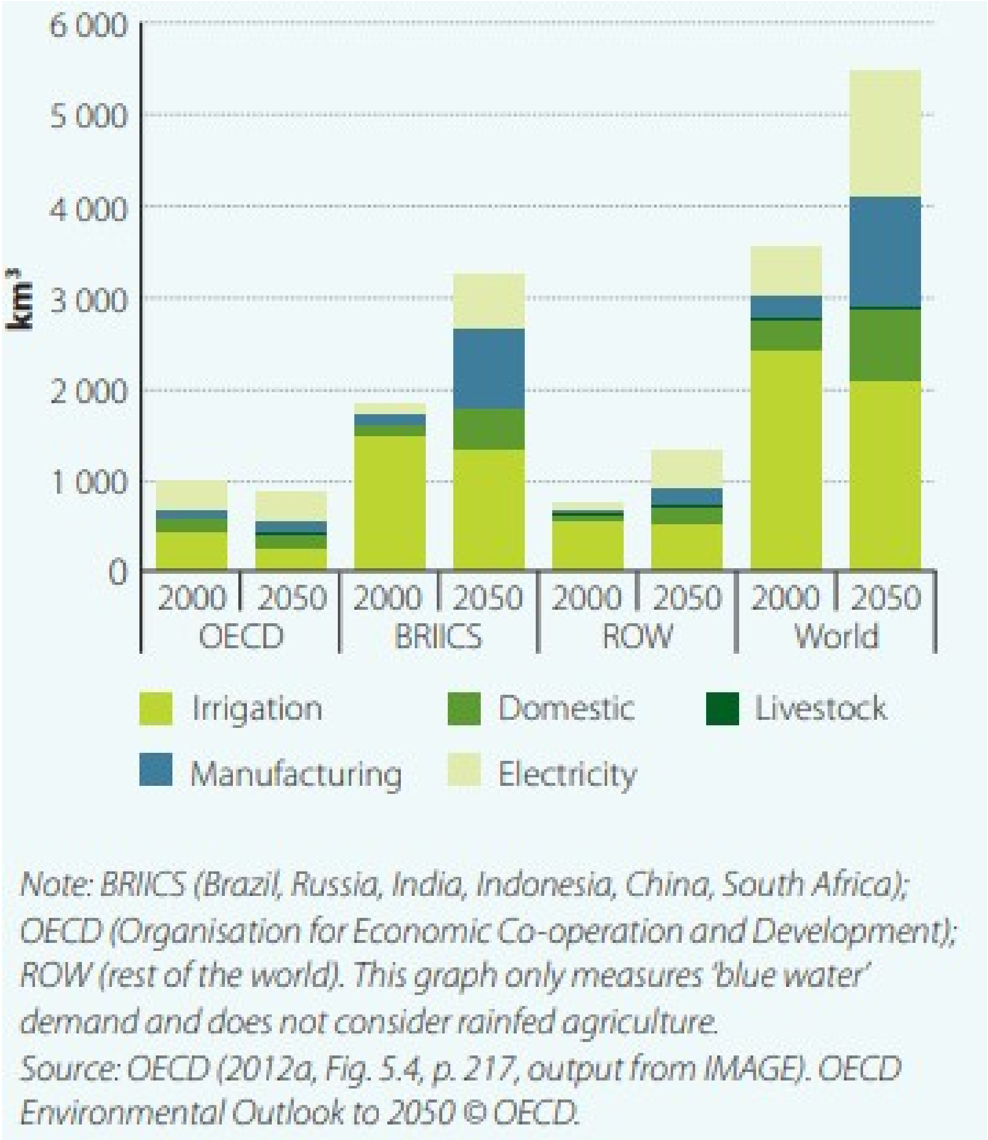


Figure 1: Global Water demand (km³) in 2000 and 2050

Source: https://unesdoc.unesco.org/ark:/48223/pf0000231823

Introduction

The objective is to propose an ensemble learning model to forecasting multi-step-ahead, one up to three-months-ahead water demand for Brazilian states;

Proposed model is composed by

- season and trend decomposition,
- some machine learning models, and
- **multi-objective** optimization;

The main contribution is the combination of different techniques including time-series pre-processing, non-linear forecasting models, and multi-objective optimization to develop an efficient forecasting model.

Datasets – Water demand (m³) for Brazilian Cities

- Datasets for two Brazilian cities, Palmas and Pato Branco;
- There are 152 monthly observations from January 2000 to August 2012 provided from the Water and Sanitation Company of Paraná;
- The data was split into Training and Testing sets in the proportion of 70% and 30%, respectively;
- Time series cross-validation is applied in training set to tune the models’ hyperparameters.

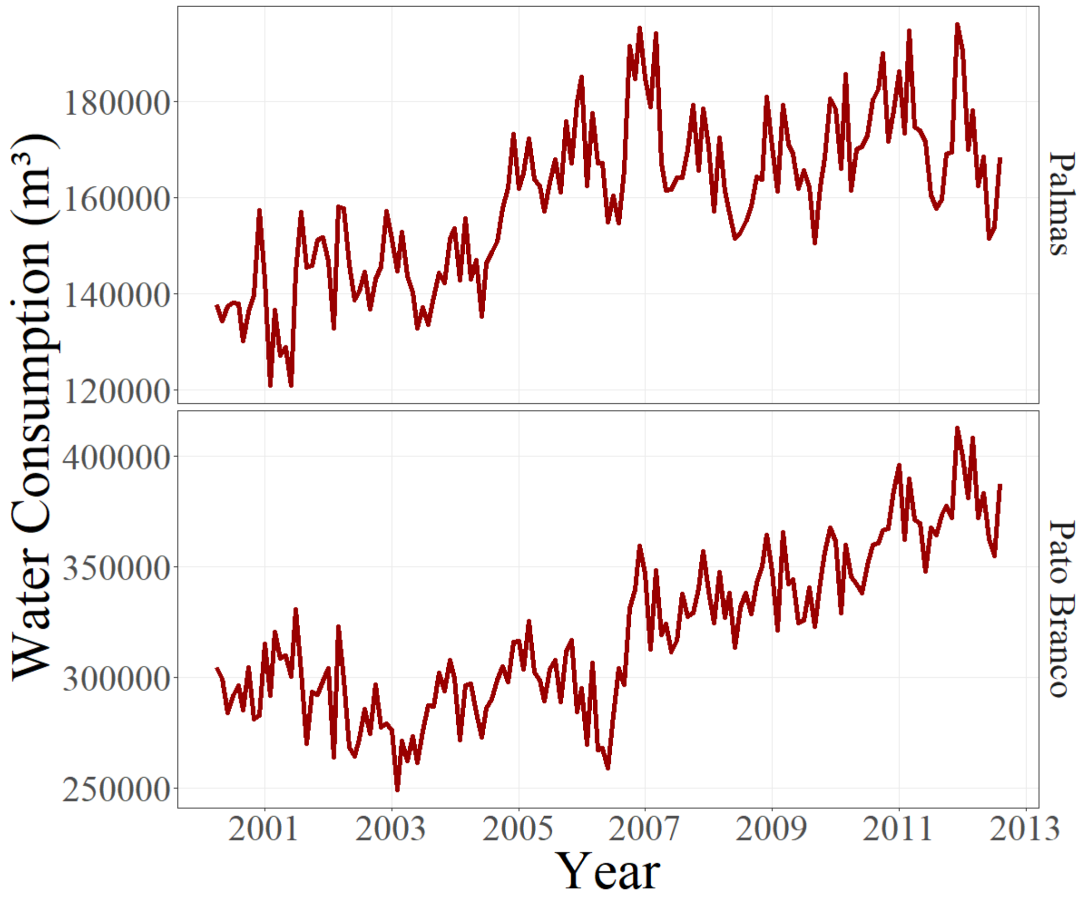


Figure 2: Water demand (m³) time series.

Methodology

- **STL** (Seasonal and Trend decomposition using Loess);
- Three different algorithms:
  - **GP** (Gaussian Processes)
  - **RIDGE** Regression;
  - **SVR** (Support Vector Regression with Linear kernel);
- Multi-Objective Procedure;
  - **NSGA-II** (Non-Dominated Sorting Genetic Algorithm – version II);
  - **BIAS-VARIANCE** trade-off;
  - **TOPSIS** (Technique for Order of Preference by Similarity to Ideal Solution);
- Performance measures:
  - **RMSE** (Root mean square error);
  - **MAPE** (Mean absolute percentage error);
  - **MAE** (Mean absolute error);

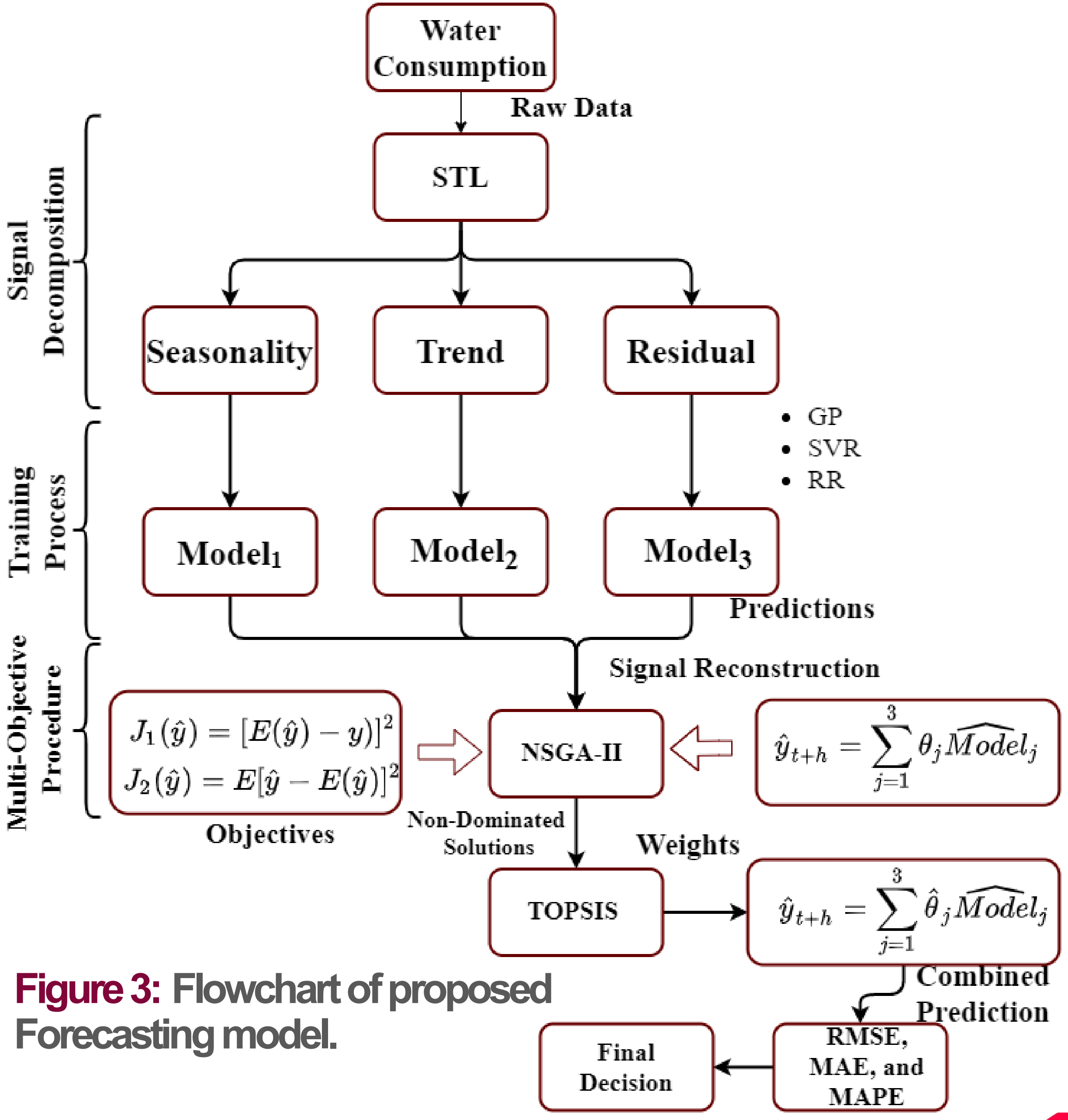


Figure 3: Flowchart of proposed Forecasting model.

PUCPR
GRUPO MARISTA

Decomposed water demand time series

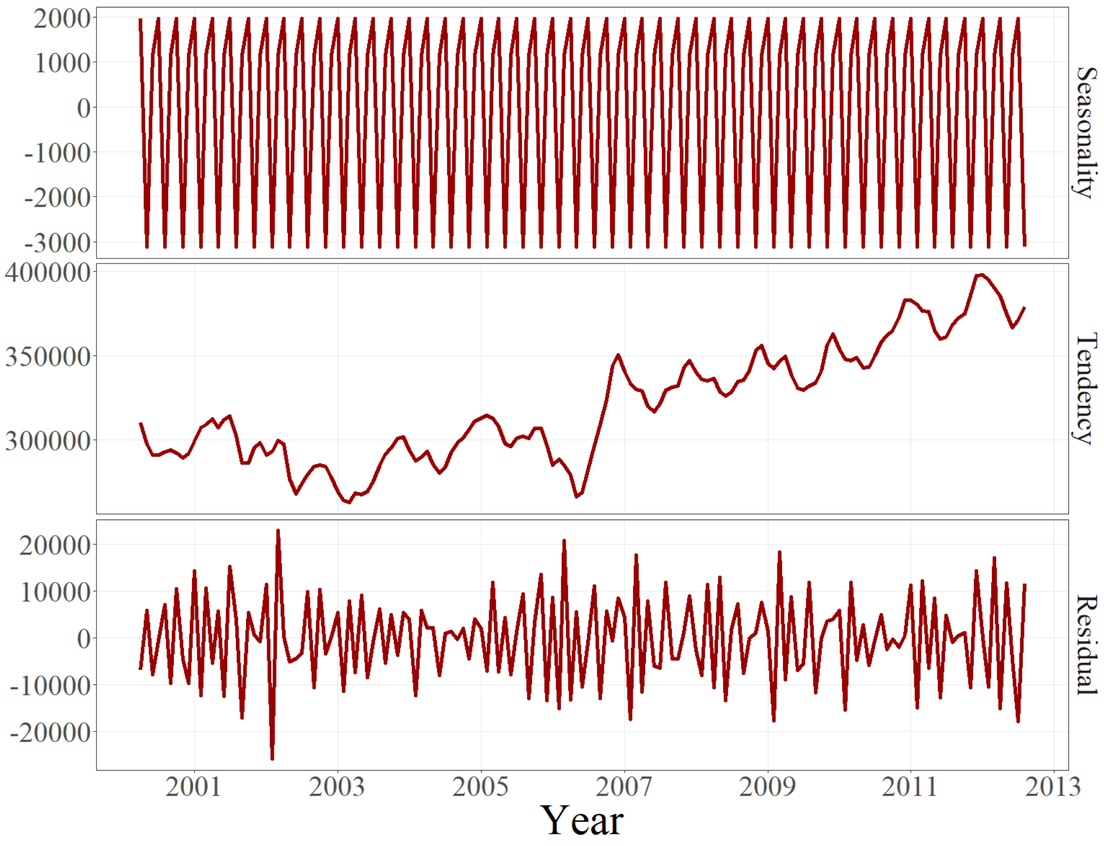


Figure 4: Pato Branco city.

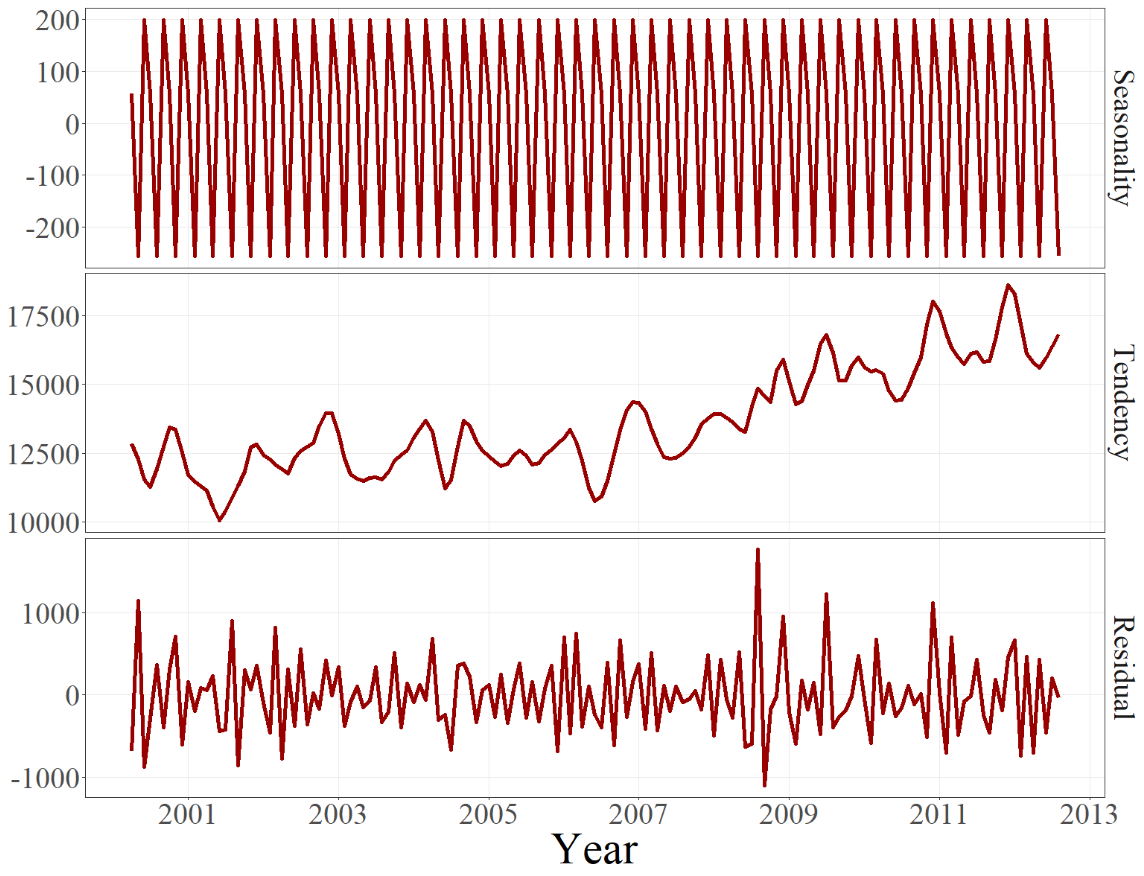


Figure 5: Palmas city.

Results

PERFORMANCE MEASURES OF FORECASTING METHODS.

City	Forecasting Horizon	Criteria	STL-MOO-GP	STL-MOO-SVR	STL-MOO-RR	STL-MOO-HTE	GP	SVR	RR
Palmas	One-month-ahead	RMSE	18205.24	3147.36	3088.05	**3034.19**	10635.46	10701.30	10644.82
		MAE	17960.02	2688.29	2661.23	**2592.85**	8651.98	8636.67	8670.79
		MAPE	10.56%	1.56%	1.54%	**1.51%**	4.97%	4.95%	4.99%
	Two-months-ahead	RMSE	12040.91	7282.46	6165.96	**6063.26**	11590.54	11691.17	11596.86
		MAE	10642.49	5853.94	4587.25	**4486.72**	9867.35	9707.77	9901.96
		MAPE	6.34%	3.49%	2.66%	**2.61%**	5.68%	5.57%	5.70%
	Three-months-ahead	RMSE	10323.38	171811.47	171822.19	**7490.07**	11209.06	11534.52	11190.45
		MAE	8463.93	171207.30	171573.15	**6093.70**	8861.69	9206.96	8850.50
		MAPE	4.82%	99.63%	100.04%	**3.52%**	5.06%	5.25%	5.05%
Pato Branco	One-month-ahead	RMSE	6008.38	36214.36	30162.89	**5119.59**	19224.80	19807.01	17902.26
		MAE	5045.90	35880.22	29834.84	**4331.76**	15879.23	16238.37	14847.18
		MAPE	1.37%	9.96%	8.28%	**1.18%**	4.31%	4.41%	4.05%
	Two-months-ahead	RMSE	10401.89	10475.07	9504.00	**8838.67**	20791.17	21398.96	19101.40
		MAE	7625.56	7435.73	**6651.20**	6725.26	17267.64	17737.44	15671.86
		MAPE	2.07%	2.02%	**1.81%**	1.87%	4.69%	4.81%	4.28%
	Three-months-ahead	RMSE	45702.71	22195.03	362693.16	**15374.36**	22314.41	23507.29	19519.84
		MAE	43937.13	19184.13	362065.51	**12825.63**	19632.81	20657.13	17117.20
		MAPE	12.03%	5.22%	100.00%	**3.51%**	5.32%	5.59%	4.65%

Results – Diebold-Mariano test

STATISTIC OF DM TEST FOR STATISTICAL COMPARISON OF PROPOSED APPROACH VERSUS OTHER MODELS

Model	Palmas			Pato Branco		
	One-month -ahead	Two-months -ahead	Three-months -ahead	One-month -ahead	Two-months -ahead	Three-months -ahead
STL-GP	-19.82***	-4.60***	-3.76***	-3.29***	-1.29	-12.15***
STL-SVR	-1.63*	-1.66*	-30.82***	-22.69***	-1.43*	-6.20***
STL-RR	-0.96	-2.99***	-38.15***	-20.44***	-3.70***	-29.13***
GP	-4.45***	-2.96***	-3.20***	-5.18***	-3.65***	-2.55***
SVR	-4.33***	-2.84***	-3.16***	-5.14***	-3.68***	-2.78***
RIDGE	-4.47***	-2.99***	-3.20***	-5.21***	-3.70***	-2.00**

Note: ***1% significance level; **5% significance level; * 10% significance level.

Results – Predicted versus Observed water consumption

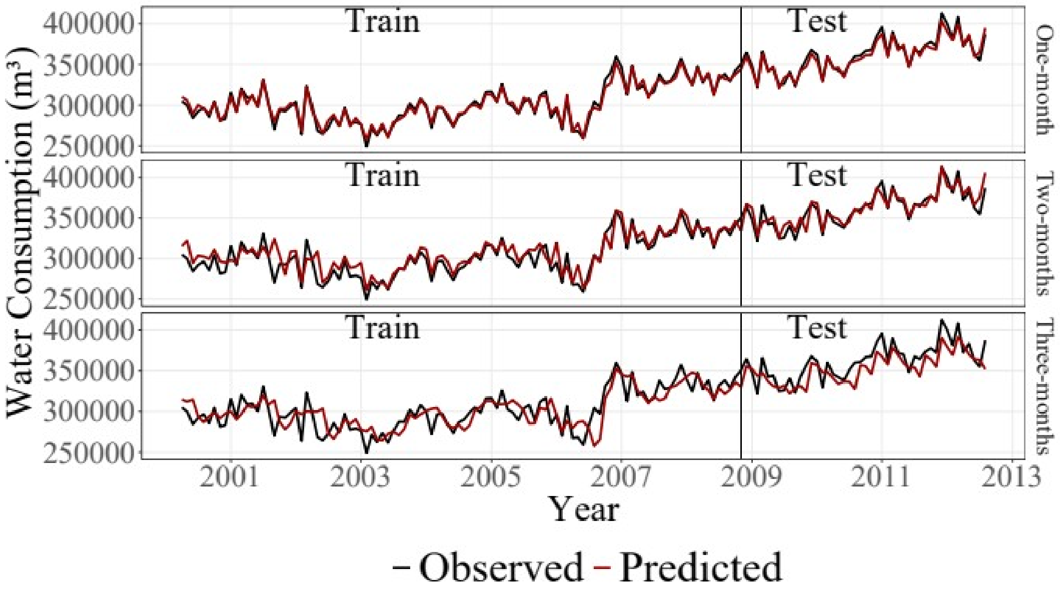


Figure 6: Pato Branco city

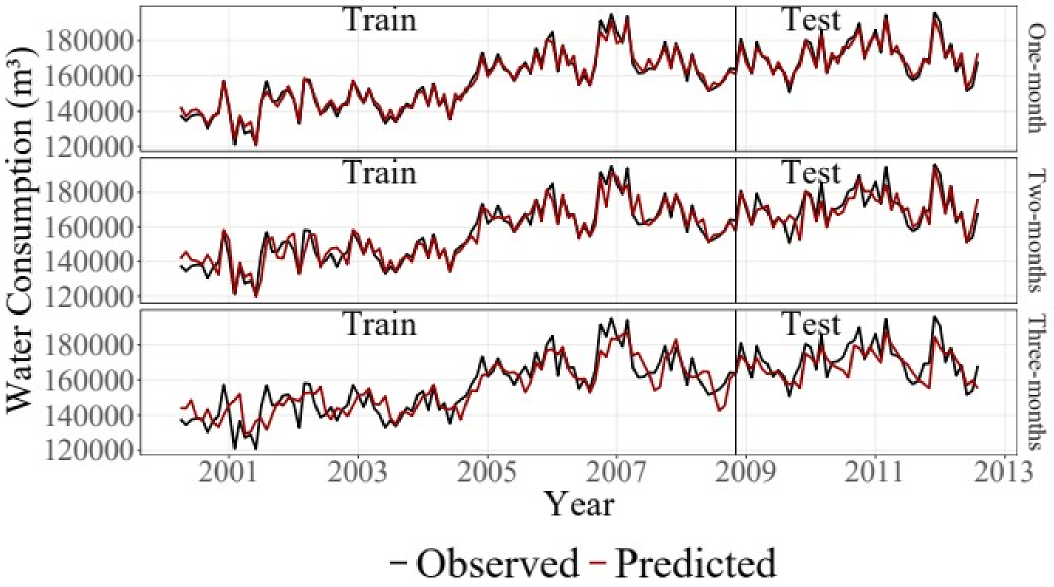


Figure 7: Palmas city

Conclusion

- STL approach allowed to deal explicit with the seasonal behavior of the data and improves the final results regarding not applying decomposition;
- Using heterogeneous ensemble learning model allow to improve the final accuracy concerning the use of a homogeneous ensemble model of components;
- Employing different weights through MOO allow different components contributing in a better way to construct an efficient forecasting model;
- Limitations of the model including: Complexity in relation to single models, and absence of exogenous factors such as climatic and demographic factors;
- Future works intend to using of exogenous factors associated with feature engineering and selection, hybridization of decomposition techniques, and to adopt RF, artificial neural networks, and persistence model to perform comparisons with the proposed framework;

References

- H. Allende and C. Valle, Ensemble Methods for Time Series Forecasting. Berlin, Germany: Springer International Publishing, 2017, vol. 349, ch. 13, pp. 217–232.
- A. F. Atiya, “Why does forecast combination work so well?” International Journal of Forecasting, vol. 36, no. 1, pp. 197 – 200, 2020.
- R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, “STL: a seasonal-trend decomposition,” Journal of Official Statistics, vol. 6, no. 1, pp. 3–73, 1990.
- M. H. D. M. Ribeiro, V. C. Mariani, and L. S. Coelho, “Multi-step ahead meningitis case forecasting based on decomposition and multi-objective optimization methods,” Journal of Biomedical Informatics, vol. 111, no. 103575, 2020.
- J. Neville and D. Jensen, “Bias/variance analysis for relational domains,” in Inductive Logic Programming, H. Blockeel, J. Ramon, J. Shavlik, and P. Tadepalli, Eds. Heidelberg, Germany: Springer, 2008, pp. 27–28.
- M. H. D. M. Ribeiro, and L. S. Coelho, “Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series”, Applied Soft Computing, vol. 86, no. 105837, 2020.
- F. X. Diebold and R. S. Mariano, “Comparing predictive accuracy”, Journal of Business and Economic Statistics, vol. 13, no. 3, pp. 253–263, 1995.

Acknowledgments

CAPES

CNPq
Conselho Nacional de Desenvolvimento Científico e Tecnológico

FUNDAÇÃO ARAUCÁRIA
Apoio ao Desenvolvimento Científico e Tecnológico do Paraná

PUCPR

Research Team

M. H . D. M. Ribeiro
Phd Candidate

Pontifical Catholic
University of Parana
(PUC-PR).

R.G. da Silva
Phd Candidate

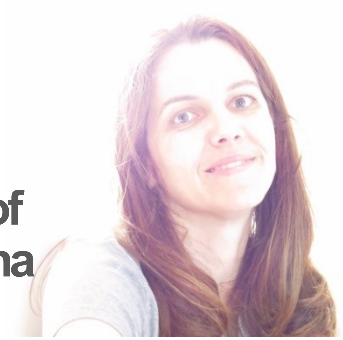
Pontifical Catholic
University of Parana
(PUC-PR).

J.H.K. Larcher
Master Student

Pontifical Catholic
University of Parana
(PUC-PR).

Dr. J. D. de Lima
Full Professor

Federal University of
Technology – Parana
(UTFPR)



Drª. V. C. Mariani
Full Professor

Pontifical Catholic
University of Parana
(PUC-PR).

Dr. L.S. Coelho
Full Professor

Pontifical Catholic
University of Parana
(PUC-PR).

PUCPR
GRUPO MARISTA