Hybrid Machine Learning Models Applied to Daily Urban Water Consumption Prediction

Ramon G. da Silva, Luis F. R. Agottani, Anderson Schamne, Andre B. da Silva, Gustavo R. C. Possetti, Leandro S. Coelho, and Viviana C. Mariani

Pontifical Catholic University of Parana, Federal University of Parana, Sanitation Company of Parana

Introduction

- Efficient planning of resources are based on accurate predictions;
- Data-driven decision-making approaches can reduce the impacts of drought periods;
- Several studies have been developed for nonlinear hydrological systems;
- Decomposition techniques alleviate the adverse impacts of the series characteristics;
- This paper applied diverse decomposition modes to preprocess time series for daily water level forecasting;
- A case study associated to a water station system in Curitiba, Brazil.

Objectives

- Evaluating the exogenous variables by correlation analysis;
- Using heterogeneous decomposition techniques for data-preprocessing;
- Employing diverse machine learning (ML) models with different learning processes.

Decomposition techniques

- Complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN);
- Empirical wavelet transformation (EWT);
- Seasonal-trend decomposition (STL);
- Variational mode decomposition (VMD);

Forecasting Models

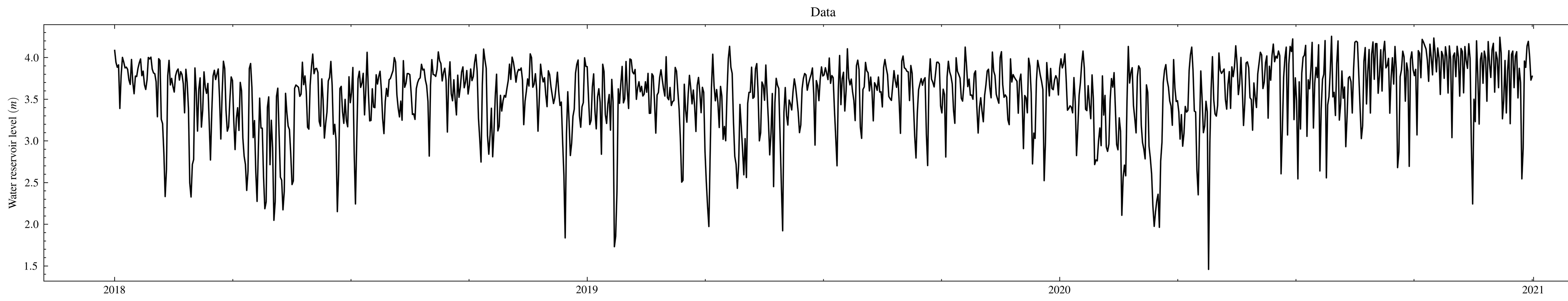
- Bayesian Regularized Neural Network (BRNN);
- Extreme Learning Machine (ELM);
- K-Nearest Neighbors (KNN);
- Multilayer Perceptron (MLP);
- Support Vector Regression with linear kernel (SVRL) and RBF kernel (SVRR).

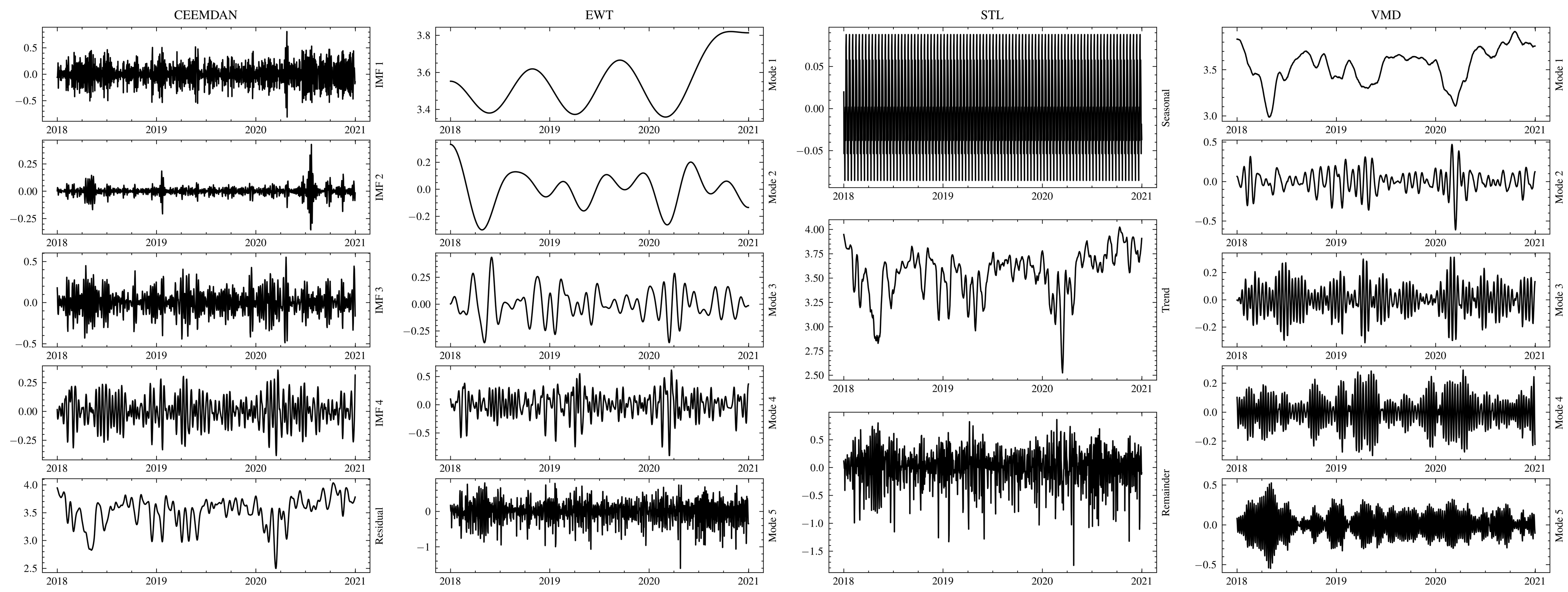
Conclusion

- SVRL, SVRR, and KNN consistently outperformed other models;
- VMD-based models achieved superior performance compared to those employing CEEMDAN, EWT, or STL;
- Even when forecasting extreme patterns over extended horizons, the VMD–SVRL model forecasts achieved its objective;
- Signal decomposition techniques combined with ML models offer a promising solution by accurately learning and predicting these nonlinear relationships.

Case Study and Methodology

- Modeling a daily series (2018 - 2020) of a water station system located in Curitiba, Brazil;
- Decomposing the water reservoir level using CEEMDAN, EWT, STL, and VMD into modes;
- Forecasting using ML models within the windows of 1 to 60 days ahead;
- Evaluating the models' performance by improvement performance index (IP), mean absolute error (MAE), and root mean squared error (RMSE) criteria;
- The experimental setup generated 150 forecasting models, representing 30 unique combinations across five forecasting horizons.





Results

Forecasting Model	1 day MAE	RMSE	15 days MAE	RMSE	30 days MAE	RMSE	45 days MAE	RMSE	60 days MAE	RMSE
(A) BRNN	0.4206	0.5232	0.6526	0.7734	0.6354	0.7496	0.6326	0.7477	0.6467	0.7581
(B) ELM	2.3137	2.6213	2.6423	2.9253	2.7785	3.0338	2.8458	3.0691	2.9152	3.1425
(C) KNN	0.5195	0.6576	0.5241	0.6642	0.5269	0.6633	0.5220	0.6621	0.5328	0.6765
(D) MLP	0.6642	1.1418	0.7341	1.2471	0.7194	1.2161	0.8587	1.4320	0.7049	1.1984
(E) SVRL	0.6833	0.8767	1.1683	1.3551	1.1848	1.3725	1.2096	1.3965	1.2539	1.4459
(F) SVRR	0.4618	0.5491	0.5177	0.6049	0.5224	0.6090	0.5295	0.6165	0.5385	0.6250
(G) CEEMDAN–BRNN	**0.2334**	**0.2920**	0.4451	0.5594	0.5305	0.6452	0.5569	0.6679	0.5787	0.6930
(H) CEEMDAN–ELM	2.6781	2.9372	3.0169	3.2002	3.0398	3.1988	3.1676	3.3177	3.2209	3.3348
(I) CEEMDAN–KNN	0.4155	0.5154	0.4255	0.5361	0.4303	0.5383	0.4446	0.5528	0.4473	0.5536
(J) CEEMDAN–MLP	0.3523	0.4928	0.5184	0.7096	0.6102	0.8753	0.6459	0.9057	0.5506	0.7757
(K) CEEMDAN–SVRL	0.2966	0.3745	0.3929	0.4997	0.4319	0.5484	0.4444	0.5601	0.4511	0.5685
(L) CEEMDAN–SVRR	0.4004	0.4801	0.4528	0.5353	0.4591	0.5436	0.4684	0.5548	0.4680	0.5546
(M) EWT–BRNN	0.3332	0.4198	0.3661	0.4574	0.4455	0.5546	0.4839	0.6061	0.6215	0.8391
(N) EWT–ELM	1.7367	2.1784	1.82E+09	1.22E+10	1.19E+20	1.13E+21	2.21E+31	2.21E+32	1.17E+42	1.17E+43
(O) EWT–KNN	0.4341	0.5452	0.4449	0.5564	0.4514	0.5671	0.4533	0.5699	0.4538	0.5718
(P) EWT–MLP	0.4276	0.5333	0.4853	0.5938	0.5111	0.6276	0.4906	0.6064	0.5410	0.6582
(Q) EWT–SVRL	0.2903	0.3717	0.4200	0.5293	0.4637	0.5782	0.4685	0.5984	0.5017	0.6276
(R) EWT–SVRR	0.4181	0.5057	0.4420	0.5278	0.4462	0.5356	0.4419	0.5345	0.4442	0.5372
(S) STL–BRNN	0.4140	0.5082	0.5913	0.7145	0.6244	0.7463	0.6405	0.7651	0.6116	0.7314
(T) STL–ELM	2.5930	2.8451	2.9406	3.1246	3.1337	3.2826	3.1654	3.3077	3.2750	3.3944
(U) STL–KNN	0.4160	0.5200	0.4223	0.5302	0.4208	0.5295	0.4206	0.5270	0.4225	0.5331
(V) STL–MLP	0.6150	0.7560	0.7954	1.0027	0.8138	1.0503	0.8746	1.0966	0.9053	1.1089
(W) STL–SVRL	0.4866	0.6048	0.5037	0.6387	0.5213	0.6696	0.5411	0.6815	0.5365	0.6848
(X) STL–SVRR	0.4565	0.5484	0.4772	0.5696	0.4800	0.5712	0.4798	0.5737	0.4789	0.5737
(Y) VMD–BRNN	0.2431	0.2987	0.3612	0.4408	0.4547	0.5772	0.4848	0.6244	0.5278	0.6843
(Z) VMD–ELM	2.7060	2.9400	3.1603	3.3144	3.1991	3.3481	3.1827	3.3355	3.2290	3.3580
(AA) VMD–KNN	0.3876	0.4707	0.3995	0.4831	0.4106	0.4947	0.4130	0.4999	0.4204	0.5076
(AB) VMD–MLP	0.3631	0.4691	0.3987	0.5085	0.4509	0.5751	0.4595	0.5927	0.4631	0.5838
(AC) VMD–SVRL	0.2403	0.2973	**0.2624**	**0.3248**	**0.3265**	**0.4076**	**0.3684**	**0.4654**	**0.3799**	**0.4722**
(AD) VMD–SVRR	0.3801	0.4568	0.4018	0.4781	0.4109	0.4878	0.4198	0.4970	0.4200	0.4968

- For 1 day forecasting, the (G) CEEMDAN–BRNN model presented the best performance, and for the remaining horizons, the (AC) VMD–SVRL model was the best.
- In overall, (AC) VMD–SVRL model demonstrated the highest accuracy, while the (N) EWT–ELM model exhibited the lowest performance.

References

[1] L. Ni, D. Wang, J. Wu, Y. Wang, Y. Tao, J. Zhang, and J. Liu (2020). Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. *Journal of Hydrology*, 586, 124901.

[2] K. Dragomiretskiy and D. Zosso (2013). Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62(3), 531–544.

[3] M. Awad and R. Khanna (2015). Support vector regression. In *Efficient Learning Machines* (pp. 67–80). Springer.

[4] R. G. da Silva, S. R. Moreno, M. H. D. M. Ribeiro, J. H. K. Larcher, V. C. Mariani, and L. S. Coelho (2022). Multi-step short-term wind speed forecasting based on multi-stage decomposition coupled with stacking- ensemble learning approach. *International Journal of Electrical Power & Energy Systems*, 143, 108504.