Environmetrics

WILEY

DISCUSSION OPEN ACCESS

Discussion on "Assessing Predictability of Environmental Time Series With Statistical and Machine Learning Models"

Paolo Maranzano[1] | Paul A. Parker[2]

[1]Department of Economics, Management and Statistics, University of Milano-Bicocca, Milano, Italy | [2]Department of Statistics, University of California, Santa Cruz, California, USA

Correspondence: Paolo Maranzano (paolo.maranzano@unimib.it)

Received: 6 November 2024 | **Revised:** 28 January 2025 | **Accepted:** 29 January 2025

Funding:

Keywords: air quality | citizen science data | heteroscedasticity | skewness | statistical and machine learning algorithms | time series forecasting

ABSTRACT

We contribute to the discussion of the insightful article "Assessing predictability of environmental time series with statistical and machine learning models" by Bonas et al. (2024), in which the authors commend their effort in comparing a wide range of methodologies for the challenging task of predicting environmental time series data. We focus our discussion on two topics of interest to us. First, we consider extensions of the explored methodologies that allow for heteroscedastic error terms. Second, we consider non-Gaussianity and fitting models on transformed data. For both of these points, we will make use of the authors' supplied code and data in order to extend their examples. Ultimately, we find that modeling of heteroscedasticity error terms has the potential to improve both point and interval estimates for these environmental time series. We also find that the use of transformations to handle non-Gaussianity can improve interval estimates.

1 | Introduction

It is a pleasure to contribute to the discussion of the insightful article "Assessing predictability of environmental time series with statistical and machine learning models" by Bonas et al. (2025). The authors are commended for their effort in comparing a wide range of methodologies for the challenging task of predicting environmental time series data. The article clearly highlights the role of short- and long-term autocorrelation in determining the quality of forecasting.

One aspect of the article in particular that stands out is the commitment to transparency and reproducibility. By making both the data and code available to the public, the authors have provided the necessary tools to replicate their findings. This also allows for the extension of their work, encouraging others to consider additional models or alternative datasets, a goal that we will explore here.

There are many interesting avenues of discussion here, however, we will narrow our discussion to two topics of interest to us. First, we will consider extensions of the explored methodologies that allow for heteroscedastic error terms. Second, we will consider non-Gaussianity and fitting models on transformed data. For both of these points, we will make use of the authors' supplied code and data in order to extend their examples.

2 | Heteroscedastic Error Terms

Environmental time series can exhibit non-constant variance for a variety of reasons. For example, particulate matter may exhibit different degrees of variation by season or due to regulatory changes. Bonas et al. (2025) show clear examples of environmental time series with this behavior, most notably their wind speed example. Understanding and modeling these heteroscedastic error terms may allow for improved predictive performance

https://doi.org/10.1002/env.70001

and uncertainty quantification, as well as new opportunities for inference.

Heteroscedastic error terms can be considered for both the classical and machine learning models explored by Bonas et al. (2025). For example The Autoregressive conditional heteroscedasticity (ARCH) model (Engle 1982) posits that the variance at time t (σ_t^2) given observations up to time $t-1$ is a linear combination of previous squared error terms(ϵ_{t-i}^2):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

Here, $\alpha_i,\ \ i = 0, 1, \ldots, q$ are coefficients to be estimated. The generalized ARCH (GARCH) (Bollerslev 1986) extends this variance model to be analogous to an ARMA model:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \gamma_i \sigma_{t-i}^2$$

where γ_i are additional coefficients. Either of these two models for the variance could be combined with the traditional time series models explored by Bonas et al. (2025), in order to jointly model the mean and variance of an environmental time series.

The recent advancements in machine learning models offer even further opportunities to jointly model the mean and variance. One notable example is the recent development of Heteroscedastic Bayesian Additive Regression Trees (HBART) (Pratola et al. 2020). The traditional Bayesian Additive Regression Trees (BART) model of Chipman et al. (2010) is a homoscedastic error model, assuming that

$$Y_i = f(\boldsymbol{x}_i) + \sigma Z_i$$

where $Z_i \overset{iid}{\sim} \mathrm{N}(0, 1)$, and $f(\boldsymbol{x})$ is modeled as a sum of trees,

$$f(\boldsymbol{x}) = \sum_{j=1}^{m} g\left(\boldsymbol{x}; \boldsymbol{T}_j, \boldsymbol{M}_j\right)$$

Here, $g(\cdot; \boldsymbol{T}, \boldsymbol{M})$ denotes a single tree with node parameters $\boldsymbol{M}$ and tree structure parameters $\boldsymbol{T}$, that operates on a p-dimensional vector of covariates $\boldsymbol{x} \in \mathbb{R}^p$. The tree structure is determined by which nodes will be split, and what variable/value they will be split on. Meanwhile, the node parameters consist of mean values for every terminal node that serve as a prediction for data points within the node. For more details see Chipman et al. (2010). Alternatively, HBART assumes a heteroscedastic error structure,

$$Y_i = f(\boldsymbol{x}_i) + \sigma(\boldsymbol{x}_i) Z_i$$

where the variance is modeled as a product of trees,

$$\sigma(\boldsymbol{x}) = \left(\prod_{l=1}^{m'} g\left(\boldsymbol{x}; \boldsymbol{T}_l, \boldsymbol{M}_l\right) \right)^{1/2}$$

Thus, HBART models the log-variance as a sum of trees. This allows for both the mean and the variance to be flexibly modeled as a non-linear function using the sum of trees structure.

In order to gauge the impact of heteroscedastic error term modeling for environmental time series, we consider the same test datasets used by Bonas et al. (2025). Specifically, we fit the HBART model to the pollution, temperature, and wind speed time series examples. In all three examples, we kept the mean structure the same and used the same inputs ($\boldsymbol{x}$) for the variance function, $\sigma(\boldsymbol{x})$ as for the mean $f(\boldsymbol{x})$. HBART was fit using the `rbart` package in R (McCulloch et al. 2019) with default settings of 40 trees for the variance model and 200 trees for the mean model. Note that the standard BART model also uses 200 trees, so any improvements are attributable to the improved variance model. Table 1 summarizes these results using the same metrics: Mean squared error (MSE), continuous ranked probability score (CRPS), 95% prediction interval coverage rate, and interval score. For the pollution data, the results are more or less the same as what was obtained using standard homoscedastic BART. However, for the other two datasets, HBART offers considerable improvement in terms of both superior predictions and uncertainty quantification. For example, with the temperature data, HBART results in about a 12% reduction in MSE compared to standard BART and roughly a 21% reduction in interval score. Similarly, for the wind speed dataset, HBART results in a roughly 11% decrease in MSE and a 95% interval coverage rate that is 4% points closer to the nominal value compared to the standard BART. This indicates that the inclusion of heteroscedastic modeling can improve both the MSE of point estimates, but also the quality of interval estimates.

Finally, Figure 1 presents a visual comparison of BART and HBART applied to the wind speed data from February 20 to 28, 2021. The black line in each plot represents the observed data, the red line indicates the posterior mean of each model's fitted values, and the green shaded region denotes the point-wise 95% credible interval (i.e., 2.5% and 97.5% percentiles of the posterior distribution) around the predictions. Visually, both BART and HBART capture the general trends and fluctuations in the data, with the credible intervals providing further insight into the models' uncertainty. The credible intervals for BART have constant width throughout the temporal domain. In contrast to this, HBART adapts the uncertainty to changes in the data. For example, HBART is able to reduce the uncertainty in periods of low variance, such as February 25 and 27. In periods of high variance, such as February 23 and 24, HBART allows for greater uncertainty, resulting in better coverage of the observed data points. Finally, although only a slight difference, HBART seems to allocate more model complexity in periods of low variance and less complexity in periods of high variance. This results in slightly smoother point estimates for HBART compared to BART when there is considerable noise present. In other words,

TABLE 1 | Comparison of forecasting and uncertainty quantification performance using HBART in terms of the MSE, CRPS, 95% Coverage, and interval score for three environmental time series.

Dataset	MSE	CRPS	95% Coverage	Interval scores
Pollution	20.44	2.46	0.95	0.63
Temperature	63.32	4.64	0.80	1.27
Wind speed	0.50	0.39	0.88	0.11

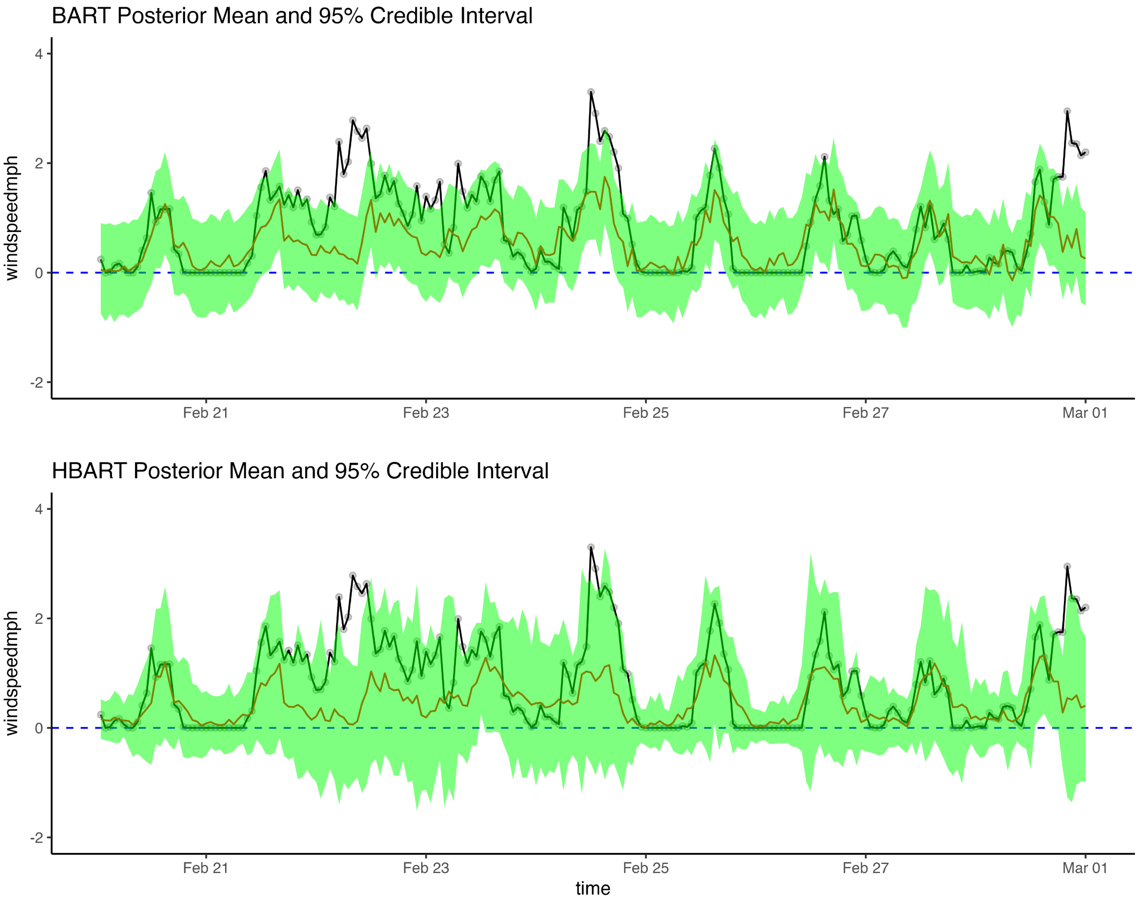

FIGURE 1 | BART and HBART fitted values for the wind speed data from February 20 to 28, 2021. The shaded regions represent point-wise 95% credible intervals.

the homoscedastic BART model has a tendency to overfit the data in periods of higher variance. This is most likely the reason that HBART was able to improve prediction MSE along with the improved uncertainty quantification.

In addition to HBART, there is potential to model heteroscedastic error terms within other machine learning frameworks. For example, Parker et al. (2021) construct an echo-state network volatility model by linking an echo-state network to a model for the variance (on the log scale) rather than the mean. Similar to HBART, it would be straightforward to extend this and use an echo-state network to model both the mean and the variance simultaneously. Machine learning approaches such as this that model both the mean and the variance in a flexible manner may offer superior uncertainty quantification by capturing complex non-linear relationships between the model inputs and the response variability. This is indicated by the improved interval coverage rate for HBART. At the same time, they may result in improved predictions by allowing the model the appropriately allocate complexity within the mean function. This could be particularly valuable in the context of environmental time series, where volatility may vary with changing environmental conditions and uncertainty quantification is critical for risk assessment and decision-making.

3 | Non-Gaussianity and Transformations

A further crucial aspect to consider when examining environmental time series is the shape of the data distribution. Measurements of atmospheric concentrations are, by definition, non-negative values that typically exhibit a positively skewed distribution characterized by a considerable number of high values. This property obviously contrasts with the classical assumption of symmetry and Gaussianity of the data, which are fundamental building blocks of statistical models. As shown in the left panel of Figure 2, the air quality data for Toronto are no exception to this situation of non-normality. In fact, the distribution of atmospheric concentrations is clearly skewed towards low values, with only a few instances of elevated concentrations. Moreover, for all monitoring stations, we can observe a considerable range of variation, reaching up to 40 $\mu g/m^3$ in some cases.

In Bonas et al. (2025) the authors compare models which treat the distribution of the response variable in different ways. For instance, ARIMA models assume the error term (conditionally to the covariates and past values) to be a sequence of independent and normally distributed random variables (see Section 7 of Wei 2006), reflecting in a Gaussian distribution for the response time series. However, in the case of skewed data, such as the airborne pollutant concentrations commented on above, this may not be ideal.

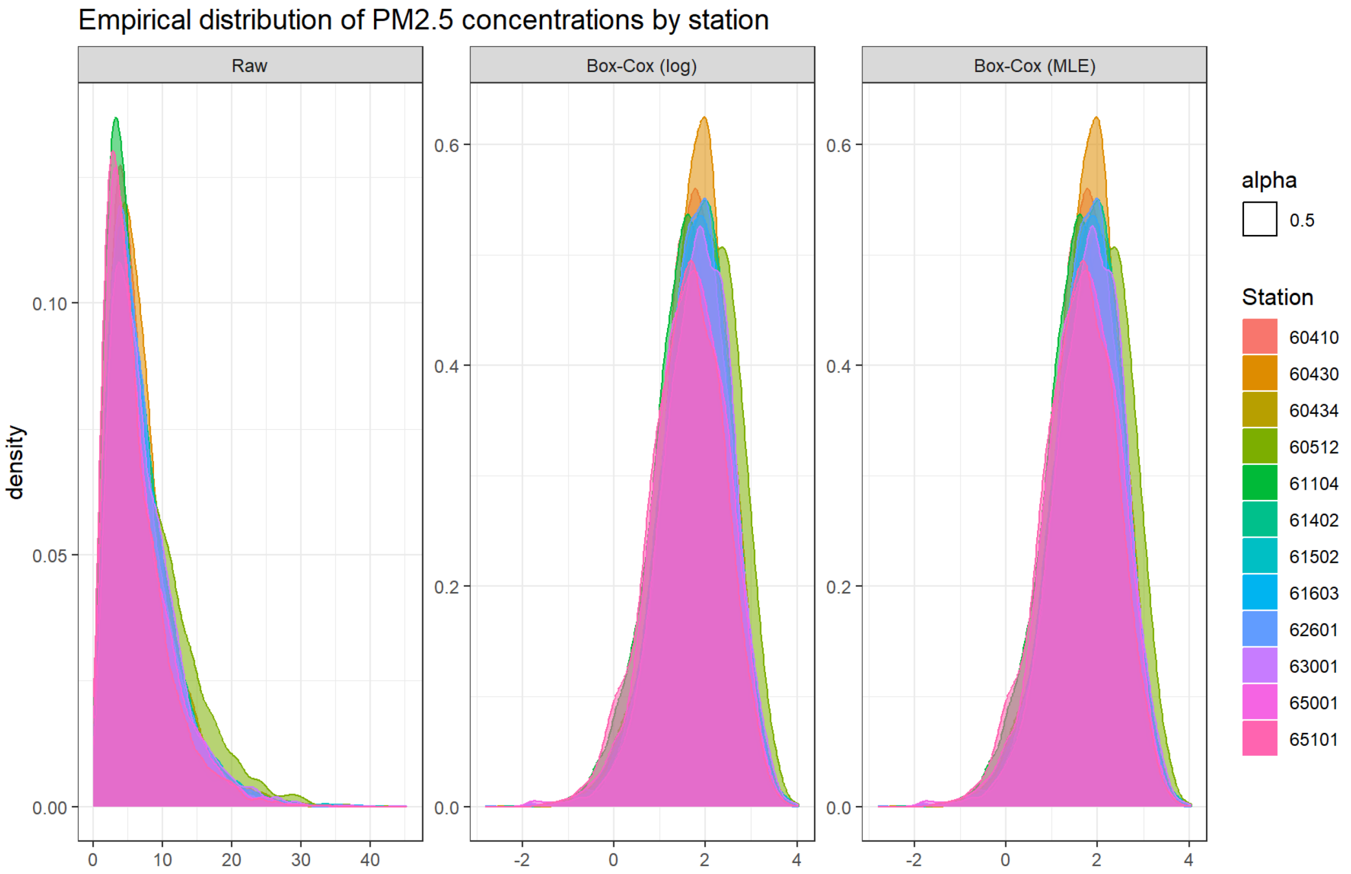

FIGURE 2 | Empirical density distributions of site-specific daily concentrations of $PM_{2.5}$ in Toronto (Canada) from 2009 to 2019. The left panel represents data in the original scale ($\mu g/m^3$), the central panel uses log-transformed data ($log[\mu g/m^3]$), the right panel uses Box–Cox transformed data with λ parameter estimated using maximum likelihood.

In order to deal with the apparent skewness, the authors propose the use of transformations such as the Box–Cox transformation (Box and Cox 1964; Sakia 1992) or logarithmic transformations (see, for example, the case of TBATS models or mixtures with log-normal components). The family of Box–Cox transformation could be used as a pre-processing step for any model, where the estimation of the parameters (i.e., the λ value) is performed in a separate stage w.r.t. the model's parameter estimation. Also, The Box–Cox transform is known to be helpful in addressing specific forms of heteroscedasticity (i.e., the variance stabilizing property of the Box–Cox transform). Figure 2 illustrates the empirical densities of the data following the application of a logarithmic transformation (middle panel), which corresponds to the Box–Cox transform with parameter $\lambda = 0$, and after the Box–Cox transformation in which the parameter is estimated by maximum likelihood (right panel). It is evident that the transformations result in distributions that are highly similar and more symmetric than the original data[1].

To test the influence of transformations on the predictive ability of models, we implement the rolling window validation forecasting strategies presented in the article for a subset of model classes. In particular, we consider ARIMA models without covariates (i.e., ARIMA) and with deterministic twenty-harmonic Fourier term to capture seasonal patterns (i.e., Fourier + ARIMA), TBATS models, and homoscedastic BART. Model parameters are estimated using observations up to 31 December 2018 as the training set[2], while forecasts[3] and corresponding error metrics are calculated out-of-sample using 2019 data.

In Figure 3, we show the site-specific daily $PM_{2.5}$ concentrations observed in 2019 (black time series) and the corresponding one-step-ahead predictions obtained with the models applied to the original and transformed data. The plot shows that the methods are mutually consistent as yield very similar point forecasts; in fact, the overlapping among the series is very high, and the range in which the values move is almost identical[4]. Thus, from an initial visual inspection, no obvious advantages in point prediction can be identified from the data transformations. This belief is confirmed by the results shown in Table 2, in which we report (for each class) the MSE of the models on the transformed data to the MSE obtained by the model on the original data. Note that for models fit on transformed data, the MSE is computed on the original data scale (i.e., after back-transforming). Although both transformations are almost always more effective than using the raw data, the out-of-sample gain rarely exceeds 5% and in some cases, even leads to worse results (e.g., BART).

Although the transformations seem to have little effect on the point predictions, we find the uncertainty of the predictions can be considerably improved. In Table 3, we report two of the uncertainty quantification metrics discussed by the authors, namely the coverage of the generated 95% prediction intervals and the interval score, both for the models on raw data and for those with

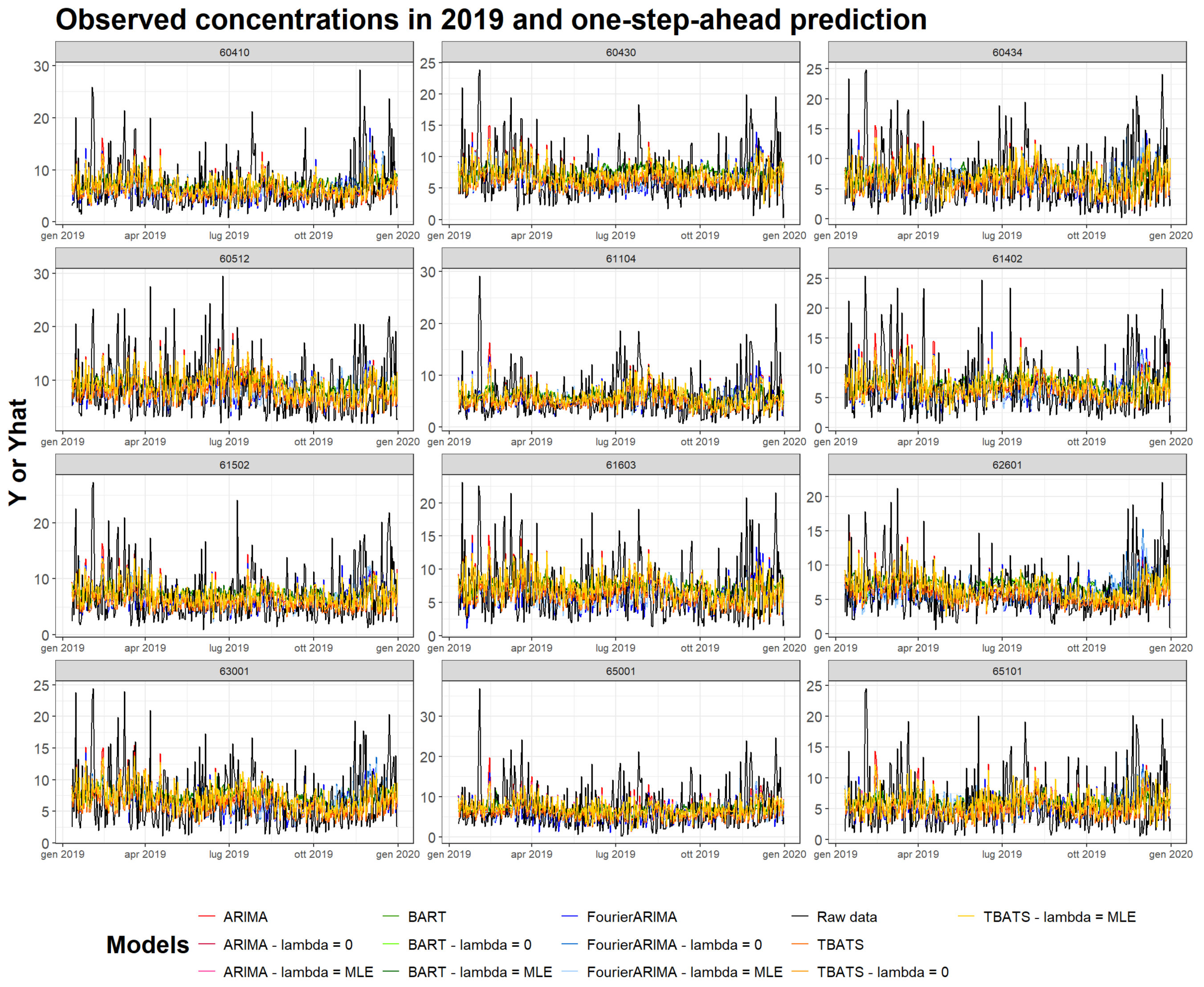

FIGURE 3 | Station-specific pollution concentrations in 2019 (black lines) and one-step-ahead out-of-sample predictions obtained by several models (colored lines).

TABLE 2 | Station-specific relative MSE of models with transformed data (and with bias adjustment) compared to the corresponding benchmark model with untransformed data. Metrics are computed using only out-of-sample observations from January 1st, 2019 to December 31st, 2019.

Model station	60,410	60,430	60,434	60,512	61,104	61,402	61,502	61,603	62,601	63,001	65,001	65,101
ARIMA ($\lambda = 0$)	0.97	0.98	0.98	1.01	0.99	0.98	0.98	1.00	1.00	0.99	0.97	0.98
ARIMA (λ_{MLE})	0.97	0.98	0.98	1.00	0.99	0.98	0.98	1.00	1.00	0.99	0.97	0.98
Fourier + ARIMA ($\lambda = 0$)	0.93	0.95	0.96	1.00	0.99	0.96	0.98	0.93	0.97	0.96	0.93	0.99
Fourier + ARIMA (λ_{MLE})	0.93	0.95	0.97	1.00	0.99	0.95	0.98	0.93	0.98	0.96	0.93	0.98
TBATS ($\lambda = 0$)	0.96	0.98	0.98	1.00	0.98	0.98	0.99	0.99	1.00	0.98	0.96	0.99
TBATS (λ_{MLE})	0.96	0.98	0.96	1.00	0.98	0.97	0.99	0.99	1.00	0.98	0.96	1.02
BART ($\lambda = 0$)	1.01	1.02	1.02	1.00	1.04	1.03	1.04	1.01	1.05	1.01	1.01	1.02
BART (λ_{MLE})	1.01	1.03	1.02	0.99	1.04	1.03	1.03	1.01	1.06	1.01	1.01	1.03

transformations. The estimates show that the prediction intervals have very similar coverage regardless of whether the logarithmic or Box–Cox transformation is used. However, the interval scores for the transformed data are significantly lower than for the untransformed case, resulting from a significantly more precise prediction interval. The reductions are always greater than 10%, with an observed a mass of around 17% for the ARIMA models with deterministic seasonal components.

TABLE 3 | Uncertainty quantification performance in terms of the 95% coverage and IS across all spatial locations, time horizon and windows for the transformed and untransformed Canada air pollution data.

	Coverage probability	Interval score	Variation w.r.t. benchmark model
ARIMA	0.94	0.62	/
ARIMA ($\lambda = 0$)	0.95	0.53	−14.72%
ARIMA (λ_{MLE})	0.95	0.53	−15.84%
Fourier + ARIMA	0.92	0.66	/
Fourier + ARIMA ($\lambda = 0$)	0.93	0.55	−17.27%
Fourier + ARIMA (λ_{MLE})	0.93	0.54	−17.88%
TBATS	0.95	0.62	/
TBATS ($\lambda = 0$)	0.95	0.53	−14.81%
TBATS (λ_{MLE})	0.95	0.53	−15.14%
BART	0.95	0.60	/
BART ($\lambda = 0$)	0.96	0.53	−10.91%
BART (λ_{MLE})	0.96	0.52	−13.09%

Other techniques for handling skewed data could also be considered. In the context of environmental data, techniques such as warping of the input space (Colombo et al. 2024; Snelson et al. 2003; Agou et al. 2022) or models with non-Gaussian errors (Alodat and Shakhatreh 2020; Jafari Khaledi et al. 2023) or responses (Otto, Fusta Moro et al. 2024) are becoming popular. However, while these tools are typically implemented in a spatio-temporal smoothing context, their performance in forecasting tasks is still poorly studied.

4 | Final Remarks and Conclusions

We have discussed the excellent article by Bonas et al. (2025) and tried to highlight its strengths and potential future extensions and comparisons to be developed. The article presented two applications to environmental data, where the goal was to compare the performance of a variety of forecasting models for time series derived from the statistical and machine learning worlds. The comparison showed that, depending on the application, the two philosophies can work antithetically, but also complement each other, and that no clear superiority can be highlighted. Based on an exploratory analysis of the data provided by the authors, we have emphasized in our commentary that at least two statistical properties can further be considered in order to obtain more accurate and meaningful predictions. The first concerns the heteroscedasticity of environmental time series, while the second concerns the skewed distribution of many environmental data, especially air quality. In both cases, the most important contribution is made to the quantification of the uncertainty of the predictions. On one side, models for heteroscedastic data such as HBART fit the dynamics of variability over time remarkably well by assigning more or less complexity depending on the periods. On the other hand, preprocessing transformations such as logarithmic and Box–Cox transformations strongly reduce the overall uncertainty over the entire forecast window.

In addition to the two specific issues addressed, others could be considered and developed in the future. For instance, the following may all be interesting points to consider:

- The forecasting horizon can affect model performance and the selection of the best predictor. When considering a single-horizon framework, in order to compare forecasts is common practice to implement the Diebold & Mariano test (Diebold and Mariano 1995; Diebold 2015), which tests the null hypothesis of no difference in the prediction accuracy of two competing forecasts using a broad class of accuracy measures and mild assumptions. Several extensions of the test have been developed to allow for more complex situations (e.g., (Giacomini and White 2006; White 2000; Hansen 2005)). When considering a multi-horizon forecasting setup (as the one implemented here), models can exhibit either uniform or average superior predictive ability (Quaedvlieg 2021). The former is defined as a model with lower loss at each individual horizon, while the latter allows poor performance at some horizons to be compensated by superior performance at other horizons. To test for these properties one can consider several test statistics, including those proposed in Quaedvlieg (2021), which extend the previously cited statistics at both the multiple-model and multi-horizon framework.
- The article compares models that leverage lagged values of the time series under consideration. However, some models allow one to use not only the information embedded in the time series (e.g., ARIMA without exogenous variables), but also external information provided by covariates (e.g., regARIMA or BART). In this particular case, the authors only considered deterministic components (e.g., Fourier bases for seasonality) that can be predicted into the future without error. In general, the presence of covariates in forecasting tasks can offer great advantages (Masini et al. 2023; Medeiros et al. 2021) by exploiting the correlation between variables to predict short-term patterns. However, one complication is the need to produce reliable predictions for exogenous information to be used as input. A potential solution is to forecast both covariates and the target variable in a multivariate framework as in (Salinas et al. 2019) and Salinas et al. (2020).
- A crucial issue in forecasting is to implement an adequate evaluation setup, typically involving a sample split strategy. In Bonas et al. (2025), the authors propose a rolling window approach in which the model's parameters are estimated once and taken as fixed for the whole evaluation test set while the h-step-ahead predictions are obtained sequentially. As discussed in Hewamalage et al. (2023), other structures, such as the expanding window or temporal cross-validation schemes could be taken into consideration. Two issues in choosing the desired setup are the computational burden and the presence of missing data. For instance, ARIMA models allow for missing data only when estimated using a state space form or through the expectation-maximization algorithm. Another important consideration is the stationarity assumption, which can be misleading, as the unknown future may differ from

the training sample, the test sample, or both (Otto, Fassò et al. 2024).

Acknowledgments

Open access publishing facilitated by Universita degli Studi di Milano-Bicocca, as part of the Wiley - CRUI-CARE agreement.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

Endnotes

[1] To further argue on the benefits provided by the log-transform, for all the monitoring sites, we computed the Jarque–Bera (JB) test, the Shapiro–Wilk (SW) test, and the Kolmogorov–Smirnov (KS) test (Das and Imon 2016) on both natural scale and logarithmic scale values. According to the BJ and SW tests for most of the stations, the logarithmic transformation mitigates the positive skewness, leading to normality in half of the cases. Conversely, the KS test, being more conservative, does not support the hypothesis of normality even after transforming.

[2] To avoid bias from outliers, all time series were pre-processed by implementing a robust STL decomposition to identify outliers and substitution via the boxplot rule on the residual component as suggested by (Hyndman 2021).

[3] In the case of Box–Cox transformed data models, forecasts are back-transformed and the bias-adjustment formula (see Section 5.6 of Hyndman and Athanasopoulos 2021) is applied.

[4] In order to assess the coherence among the point forecasts from different models we computed the station-specific pairwise Pearson's linear correlation. Indeed, as two series predict similar values toward the same direction, they should exhibit a positive and close-to-one linear correlation. For every station, we estimate that the proposed models are mutually consistent. Indeed, the linear correlation is very high for ARIMA and TBATS models (often above 0.80) while decreases for BART models (around 0.30 linear correlation with respect to other models but over 0.90 for other BARTs).

References

Agou, V. D., A. Pavlides, and D. T. Hristopulos. 2022. "Spatial Modeling of Precipitation Based on Data-Driven Warping of Gaussian Processes." *Entropy* 24: 321.

Alodat, M., and M. K. Shakhatreh. 2020. "Gaussian Process Regression With Skewed Errors." *Journal of Computational and Applied Mathematics* 370: 112665. https://doi.org/10.1016/j.cam.2019.112665.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics* 31: 307–327.

Bonas, M., A. Datta, C. K. Wikle, et al. 2025. "Assessing Predictability of Environmental Time Series With Statistical and Machine Learning Models." *Environmetrics* 36, no. 1: e2864. https://doi.org/10.1002/env.2864.

Box, G. E. P., and D. R. Cox. 1964. "An Analysis of Transformations." *Journal of the Royal Statistical Society: Series B* 26: 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Chipman, H. A., E. I. George, and R. E. McCulloch. 2010. "BART: Bayesian Additive Regression Trees." *Annals of Applied Statistics* 4: 266–298.

Colombo, P., C. Miller, X. Yang, R. O'Donnell, and P. Maranzano. 2024. "Warped Multifidelity Gaussian Processes for Data Fusion of Skewed Environmental Data." *arXiv*, 2407.20295.

Das, K. R., and A. Imon. 2016. "A Brief Review of Tests for Normality." *American Journal of Theoretical and Applied Statistics* 5: 5–12.

Diebold, F. X. 2015. "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests." *Journal of Business & Economic Statistics* 33: 1. https://doi.org/10.1080/07350015.2014.983236.

Diebold, F. X., and R. S. Mariano. 1995. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 13: 134–144. https://doi.org/10.2307/1392185.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of United Kingdom Inflation." *Econometrica* 50, no. 4: 987–1007. https://doi.org/10.2307/1912773.

Giacomini, R., and H. White. 2006. "Tests of Conditional Predictive Ability." *Econometrica* 74: 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x.

Hansen, P. R. 2005. "A Test for Superior Predictive Ability." *Journal of Business & Economic Statistics* 23: 365–380. https://doi.org/10.1198/073500105000000063.

Hewamalage, H., K. Ackermann, and C. Bergmeir. 2023. "Forecast Evaluation for Data Scientists: Common Pitfalls and Best Practices." *Data Mining and Knowledge Discovery* 37: 788–832. https://doi.org/10.1007/s10618-022-00894-5.

Hyndman, R. J., and G. Athanasopoulos. 2021. *Forecasting: principles and practice*, 3rd ed. OTexts. Accessed on December 02, 2025.

Hyndman, R. J. 2021. *Detecting Time Series Outliers*. https://robjhyndman.com/hyndsight/tsoutliers/.

Jafari Khaledi, M., H. Zareifard, and H. Boojari. 2023. "A Spatial Skew-Gaussian Process With a Specified Covariance Function." *Statistics & Probability Letters* 192: 109681. https://doi.org/10.1016/j.spl.2022.109681.

Masini, R. P., M. C. Medeiros, and E. F. Mendes. 2023. "Machine Learning Advances for Time Series Forecasting." *Journal of Economic Surveys* 37: 76–111. https://doi.org/10.1111/joes.12429.

McCulloch, R., M. Pratola, and H. Chipman. 2019. "Rbart: Bayesian Trees for Conditional Mean and Variance." R package version 1.0.

Medeiros, M. C., G. F. R. Vasconcelos, Á. Veiga, and E. Zilberman. 2021. "Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods." *Journal of Business & Economic Statistics* 39: 98–119. https://doi.org/10.1080/07350015.2019.1637745.

Otto, P., A. Fassò, and P. Maranzano. 2024. "A Review of Regularised Estimation Methods and Cross-Validation in Spatiotemporal Statistics." *Statistics Surveys* 18: 299–340. https://doi.org/10.1214/24-SS150.

Otto, P., A. Fusta Moro, J. Rodeschini, et al. 2024. "Spatiotemporal Modelling of Pm 2.5 Concentrations in Lombardy (Italy): A Comparative Study." *Environmental and Ecological Statistics* 31: 245–272. https://doi.org/10.1007/s10651-023-00589-0.

Parker, P. A., S. H. Holan, and S. A. Wills. 2021. "A General Bayesian Model for Heteroskedastic Data With Fully Conjugate Full-Conditional Distributions." *Journal of Statistical Computation and Simulation* 91: 3207–3227.

Pratola, M. T., H. A. Chipman, E. I. George, and R. E. McCulloch. 2020. "Heteroscedastic Bart via Multiplicative Regression Trees." *Journal of Computational and Graphical Statistics* 29: 405–417.

Quaedvlieg, R. 2021. "Multi-Horizon Forecast Comparison." *Journal of Business & Economic Statistics* 39: 40–53. https://doi.org/10.1080/07350015.2019.1620074.

Sakia, R. M. 1992. "The Box-Cox Transformation Technique: A Review." *Journal of the Royal Statistical Society. Series D* 41: 169–178. https://doi.org/10.2307/2348250.

Salinas, D., M. Bohlke-Schneider, L. Callot, R. Medico, and J. Gasthaus. 2019. "High-Dimensional Multivariate Forecasting With Low-Rank Gaussian Copula Processes." *Advances in Neural Information Processing Systems* 32.

Salinas, D., V. Flunkert, J. Gasthaus, and T. Januschowski. 2020. "Deepar: Probabilistic Forecasting With Autoregressive Recurrent Networks." *International Journal of Forecasting* 36: 1181–1191. https://doi.org/10.1016/j.ijforecast.2019.07.001.

Snelson, E., Z. Ghahramani, and C. Rasmussen. 2003. "Warped Gaussian Processes." *Advances in Neural Information Processing Systems* 16.

Wei, W. W. S. 2006. *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley Publishing Company, Incorporation.

White, H. 2000. "A Reality Check for Data Snooping." *Econometrica* 68: 1097–1126. https://doi.org/10.1111/1468-0262.00152.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.