Fluvial water turbidity: a neuronal network approach to turbidity forecast in small population centres.

Raúl Del Valle García

Universidad Internacional de la Rioja, Logroño (España)

15th July, 2021

ABSTRACT

The presence of turbidity in water is a problem that impacts human health and water treatment infrastructures. In the proposed approach, a local turbidity model is developed to predict its behavior. In small towns, due to the lack of means, turbidity prediction is the only possibility to guarantee water quality at an affordable economic cost. Hence, water turbidity is taken as a univariate non-seasonal signal. Setting a time horizon of 7 days, the proposed neural network model can forecast water turbidity performing a SMAPE=11.6% error. This improves an ARIMA-based model outcome by 1%.

KEYWORDS

Water Treatment, Turbidity, Time Series Forecasting, Neuronal Networks

I. INTRODUCTION

THE approach to water turbidity is a common topic that concerns hydrology and water treatment engineering. Numerous efforts are continuously ongoing for its study and modeling. Turbidity is a critical indicator that demonstrates water quality in terms of its salubrity and acceptance as an indispensable commodity. The prediction of turbidity in river water has numerous fields of application. One of the most common use concerns to plan operations and costs related to drinking water. Multiple factors affect turbidity, the most obvious being those related to hydrology, climatology, and anthropocentric. The result is that turbidity can only be modeled locally. Cities of a certain size can perfectly justify the expenses derived from the physical and chemical processes for fluvial water turbidity remedy when such continental water is available for the catchment. Turbidity increase always means skyrocketing costs related to infrastructure and water treatment chemical consumption. For small population centres, turbidity prediction is a vital strategy for planning the best water-taken conditions. Hence, facilities being able to capture an adequate volume of river water, at the precise moment of low turbidity, means a very important advantage. The purpose of this work is allowing a fair approach to the problem of water turbidity modeling. Drawbacks derived from the fluvial water turbidity model as a non-stationary univariate signal are shown by employing a local case study. The objective is to explain all the necessary concepts (and provide proper tools) to model a local turbidity predictor, especially in small communities. Such a tool allows an estimation of the economic feasibility of the sizing and maintenance of the water treatment infrastructure. The predictor becomes an important issue for estimating associated costs related to equipment and chemical consumption.

At the early stage, it is essential to provide adequate data preprocessing. Turbidity can undergo pronounced changes in its measurement. Hence, mean-based statistics operators should be carefully used. As result, they may eliminate important information by considering a succession of high turbidity values as unlikely. Therefore, mentioned operators may not attach sufficient importance to signal geometry. It implies that convolutional methods could lose crucial features in the geometry of the signal hampering the model accuracy. On the contrary, median-based operators have proven to be more appropriate for preprocessing tasks while maintaining the coherence of the statistical moments and preserving the geometrical shape of the signal.

The next step is to perform a qualitative approach to turbidity, carrying out a detailed study. For this purpose, turbidity is assimilated as a stochastic process defined by a non-stationary signal. Determining the time horizon at this stage is also very important. Local modeling always requires a trade-off between accuracy and the practicalities of water treatment. To further study the signal, an approximation is modeled through autoregressive and exponential smoothing models. This step will determine whether the computational cost of a neural network is worthwhile compared to a more conventional approach and determines the success of neural network-based modeling. In our case, given a prediction horizon of seven days, the autoregressive ARIMA(1,1,2) model was beaten by a convolutional predictor based on one of the architectures proposed in this paper for prediction. ARIMA showed an error in the SMAPE prediction =12.56% versus the SMAPE=11.66% error of the neural network.

Currently, many predictive models rely on a hybrid approach based on both statistics and neural networks. The insight behind this concept is to take full advantage of statistical modeling and assign to neural networks the more difficult task of the search for patterns where statistical techniques are difficult to apply. However, the mentioned hybrid models are limited when the signal decomposition (in its additive or multiplicative form) does not produce good results. Therefore, it is pointed out in the conclusions that in the case of turbidity, other types of signal decompositions may be more suitable. Such is the case of the Empirical Decomposition Model (EMD) based on the Hilbert Huang decomposition.

The approach taken in this paper makes contributions to enable personnel with some background in computer science to have the capability to develop local turbidity predictors. It broadens the possibilities of applying water treatment engineering techniques to specific locations in a more site-specific manner.

II. State of the Art

Turbidity is a physical property in which certain types of elements present in the water give it opacity. Therefore, opacity reveals the presence of impurities in the water. The more turbidity the water has, the less purity or quality can be attributed to it. Not all turbidity is visible to the naked eye. In particular, certain dissolved solids having organic or inorganic origin can alter the turbidity of water by acting as contaminants without the human eye being able to perceive them.

Opacity is measured in nephelometric units or NTU. The presence of turbidity hinders the effectiveness of additives and disinfection products used in water treatment. Thus [1] states that turbidity greater than 1-2 NTU reduces significantly the effectiveness of disinfection. Water should be kept below 1 NTU to benefit from the savings in disinfectant chemical consumption. Many water quality standards seek to maintain water turbidity under 0.2 NTU. If reduction of turbidity below mentioned thresholds is not possible, at least a limitation below 5 NTU should be guaranteed [2]

One of the most important features of turbidity is that the process is local and extremely variable. Turbidity variations in river water have many different causes. In well-preserved river water, the main reason is the fluvial sediments streamed by the water flow. After contrasting fifteen turbidity models in Indonesia, [3] cites that there is no single general formulation applicable to all transported sediment data sets. Therefore, not all sediment transport models include a version applicable for the case of a particular river. Other hydrological, meteorological [4] or anthropogenic [5] events also take part into turbidity levels. They add up complexity to sediment transport modeling. **The result is that turbidity must be modeled on a case-by-case basis to be accurate in the estimates**.

The methods used to predict turbidity levels in water are very diverse. The paper published by [6] cites several groups of them. Some are autoregressive or similar models (such as ARIMA). The algorithms classified as machine learning is also quite numerous with Support Vector Machine, K-nearest neighbor, and Naive Bayes standing out. In the field of neural networks, architectures based on Long Short-Term Memory (LSTM) are widely used.

In the M4 competition [7] sixty-four predicting models were compared. The categories included pure mathematical methods, methods based solely on neural networks, and hybrid methods combining the two previous categories. The competition evaluated the predictive capability of over 100,000 datasets of different types. Among **the most important findings, it is stated that the hybrid models that combined statistical methods with neural networks were superior to the rest. In contrast, the group based purely on neural networks obtained the poorest results.** Another interesting finding is that the prediction of one of the variables improved if multiple series were used in a multivariate approach. Slawek Smyl, contest winner and data scientist at Uber, used a hybrid model [8] based on an exponential smoothing model combined with an LSTM-based architecture. According to [9] although many other models (Depth Gated LSTM, Clockwork RNN, Stochastic Recurrent Networks, Temporal Convolutional Networks, Bidirectional RNN) have been introduced in the academic world, there is hardly any literature in the field of time series prediction. Recursive Neural Networks (RNN)-based architectures are currently the most widely used for time series forecasting, especially those based on LSTM and Gated Recurrent Unit (GRU). The same author states that predictive methods **using neural networks are not a silver bullet for every forecasting puzzle. It is difficult to replace other more classical and popular methods (such as ARIMA). However, neural networks can be a very competitive alternative in some scenarios. The work done for the present paper can undoubtedly confirm the same result.**

III. Objectives and Methodology

The present work proposes a set of tools and techniques that enable the local study of flow water turbidity at a specific point. The final intention is to show the implications and problems of predicting turbidity through a predictor developed as a use case. Eventually, the turbidity predictor should be usable for decision-making related to water treatment planning. During the use case, with orientation to its use in small communities, the pursued purpose is:

- To perform a temporal characterization of turbidity, with the capacity to understand and predict its occurrence cycles to prevent its entry into the system.
- To provide tools and analysis techniques for data processing that allow the correct representation and interpretation of turbidity, in the context of river water consumed by small populations.
- To model the behavior of local turbidity through a predictor. Predictions should be made over a practical and consistent time horizon. The relatively low turbidity level should be determined in the time interval to improve water quality.
- The work will be developed with Open Source software tools and public resources to facilitate access to the widest possible audience.

Tools

The following tools were employed: Anaconda Individual Edition 2021.05, Python 3.5, Prophet 1.01 (available at

https://pypi.org/project/prophet/), Pmdarima 1.8.2 (available at https://pypi.org/project/pmdarima/), statsmodels v0.12.2 (available at https://www.statsmodels.org/stable/index.html), Keras on TensorFlow 2.5 running on Anaconda, and Google Colab.

There are many data sources for the study of turbidity, although unfortunately, they are not always accessible to the general public. The best quality data sources correspond to automatic measurement systems maintained by public institutions or private companies involved in environmental studies or water treatment. In many cases, data are not available data for a specific location. Sometimes satellite technology can help to obtain data when automated field measurements are not available.

Datasets

Turbidity datasets for our paper were provided by the United States Geological Survey (USGS). The data are available at: https://waterdata.usgs.gov/or/nwis/uv?site_no=11501000. Frequency sampling is 15 minutes. No downsamplings were performed. This interval is the most interesting for us. It allows us to estimate events on an hourly basis. The place corresponds to a small community called Chiloquin, located in the State of Oregon, USA (42°34′30″N 121°51′51″W). This small community (2.12 km2) has an estimated total of 749 people in 2019 (United States Census Bureau, n.d.). The Köppen-Geiger climate classification [10] classifies this area along with Northwest Spain as "Csb" class: temperate or mesothermal, mild summer oceanic Mediterranean climate, with temperatures between 22ºC to -10ºC at least four or more months per year.

Preprocessing

The source data in RDB format is transformed into a CSV format. The transformation is simple because both formats are tabulated ASCII text. The next part of the preprocessing is the search and imputation of missing values. The occurrence of missing values can be due to multiple aspects, some of them can be [11]: failure of the field instrumentation/loss of equipment communications, poorly calibrated measurement equipment, failure of the system to interpret special format values or failure to interpret numerical format errors. These events usually cause the appearance of missing values in the datasets. Correction involves imputing new values through interpolation using degree two polynomials. Inconsistent values (zero or negative values) are also treated as missing values because they are meaningless in the domain. For the following treatment of outliers, outliers math z-score and the Hampel filter are compared. The Hampel filter results to be more suitable for this purpose. In the first approach, descriptive statistical techniques are used. Finally, Facebook's Prophet library is employed to decompose turbidity signal into the trend, hourly, weekly and monthly series. This decomposition is easy to perform and provides a powerful and adequate descriptive tool for turbidity.

Modeling Turbidity as a Time Series

Turbidity is defined as a non-stationary univariate stochastic process for our shake (otherwise, it would be easy to model it in an autoregressive way e.g.). Its modeling is done through additive and multiplicative decomposition of the signal. The statsmodels.tsa.seasonal (additive and multiplicative seasonal modeling), statsmodels.tsa.holtwinters (exponential smoothing) libraries of the statsmodels Python package are used for this purpose. Finally, predictive models based on Simple Exponential Smooth (SES), Exponential Smooth (a.k.a Holt Winters, the full model), ARIMA, and Prophet are built.

All models were fitted setting a 0.95 confidence interval. The error metrics are defined by: MAE, MAPE, MDAPE, MEDAE, MSE, RMSE, and SMAPE. Of these, SMAPE is the most important metric that will define the best model. To determine the prediction time horizon, the above models passed two tests. The first is the prediction capability of fewer than twelve hours by performing the calculation in one, three, six, and twelve hours. The second is the prediction capacity in days, taking one, two, five, seven, and fifteen days. By comparing the results and observing where the prediction error starts to increase significantly (based on the SMAPE metric) the maximum prediction horizon can be estimated. Finally, other practical aspects determine the appropriate prediction horizon, always below the maximum prediction horizon. The best autoregressive model will be used to evaluate the forecasting models based on neural networks.

Modeling Turbidity as a Neuronal Network

The objective here is to model a neural network. When it comes to turbidity, complex signal behaviors are expected due to the multiple causes. It is possible that some components of the decomposed signal can be explained by autoregressive and exponential models. However, depending on the case, neural networks may have to be used as a form of modeling if the previous methods are not entirely satisfactory. The approach may well be a hybrid model between neural networks and other statistical methods. In our case, the following neural network architectures are compared as turbidity predictors: Convolutional MPL, CNN + LSTM, Wavenet, and Bidirectional LSTM. A neural network may not obtain better results than other statistical methods. Thus, the accuracy of the already trained neural networks must be compared in the same terms as the aforementioned models.

IV. CONTRIBUITON

The main contribution of this paper is to bring together the fields of water treatment engineering and neural network modeling for the study of river water turbidity. Through this methodology and by using a use case, the feasibility of adopting neural networks for turbidity prediction has been shown. The work, made available to the general audience, offers a set of tools and approximation criteria for the study and modeling of turbidity. The use of selected open-source tools and the simplicity of the approach is particularly significant for small population centers. Hence, non-specialized personnel can build turbidity models for planning infrastructures and processes related to water treatment.

V. Evaluation and Results

Preprocessing

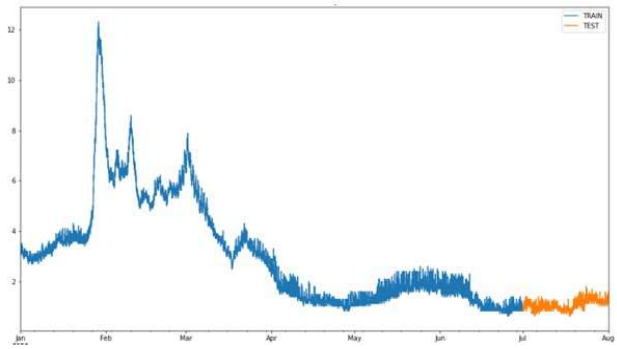

Fig.1: Training and Test data

The dataset used for the use case illustration comprises the turbidity signal during the period January - August 2020. The period is selected because it contains mostly the rainy season (February to March). The next period corresponding to the summer months has relatively stable values. This configuration gives some advantage to autoregressive methods such as ARIMA, which have a more stable trend several for many months before the prediction period. In Fig.1 in blue the training data (85.45%), the orange line corresponds to the prediction (14.55%).

Two statistical tests were used to prove that the signal is not stationary: Augmented Dick-Fuller [12] and [13]. In both cases, it was concluded that it is not stationary. By using the KPSS test it was shown that it was also not deterministically trending (a condition assimilated to stationarity or trend-stationary, without becoming formally stationary). If the signal is stationary, it'll be modeled autoregressively or exponentially with high accuracy. Phophet library was used for trending modeling (Fig.2) having great success as turbidity trend descriptor. The Prophet library was successfully used for turbidity trend description. Fig.2, clearly shows that the best time to drink water is between Saturday and Tuesday of the following week (preferably between 10:15 and 17:15). The practical limitations imposed by the infrastructure of small communities make it impractical to apply monthly or annual trends. The infrastructure of a small town cannot take advantage of the monthly minimal turbidity levels between May and December. It simply cannot store enough low turbidity water to supply the population in the rainy season, when turbidity is much higher. This consideration gives the prediction horizon a practical limit of seven days (one week) in this case since it will not be of much use to predict turbidity later turbidity values.

Modeling Turbidity as a Time Series

An additive and multiplicative decomposition of the signal into trend, seasonality, and residual was performed. The multiplicative decomposition yielded no other results. The signal y(t) is decomposed in the additive model as y(t) = t(t)+s(t)+r(t), where t(t) is the trend, s(t) is the seasonality and r(t) the residual. The decomposition (Fig.3) shows that the signal is examined in a wider interval than the one considered for training and testing. The signal is explained through the trend and residual part. The stationarity s(t) has hardly any influence on the decomposition.

Fig. 4 shows the additive decomposition before, during, and after the period considered for training and testing. The trend explains and the residual explain practically the totality of the signal.

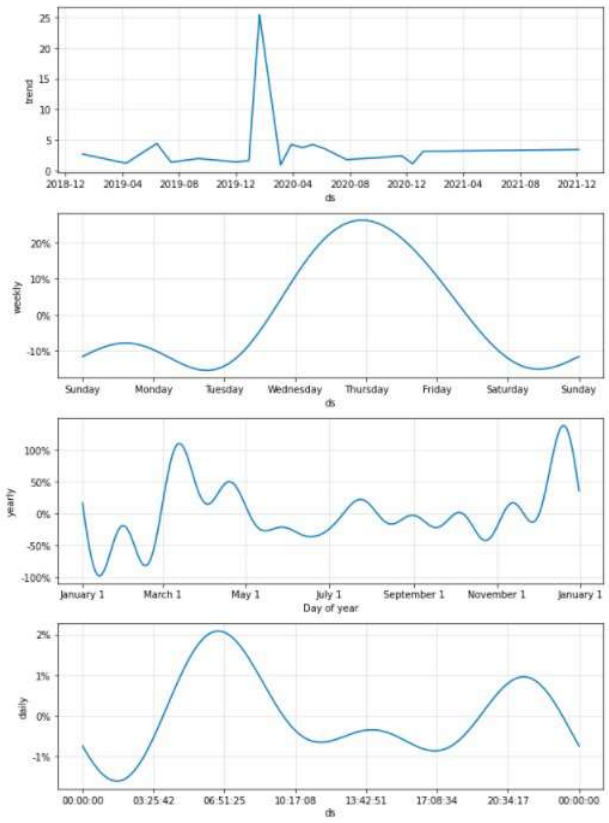

Fig.2: Prophet's Trend Decomposition

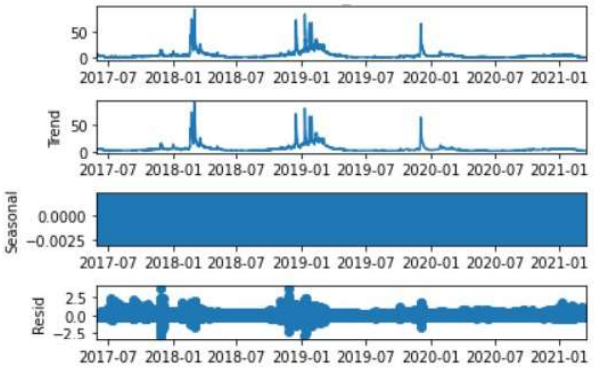

Fig3: Turbidity Decomposition July 2017- January 2021

Four non-neural network-based predictors were tested. The two of them belong to the family of predictors based on

exponential smoothing. The simple exponential smoothing (SES) only takes into account the trend. The Holt-Winters exponential smoothing (ETS) a.k.a. exponential smoothing or full exponential smoothing is a more complete version of SES. Two more models were considered: ARIMA and Prophet's model. The orders of the ARIMA model are automatically calculated by the Pmdarima package without user intervention.

The result was an ARIMA(1,1,2) model. A second autoregressive ARIMA model with seasonality was manually optimized. The result was a SARIMA(0,1,1)(2,1,2,6) model. The error metrics of both models were very close, so ARIMA(1,1,2) was chosen because it's a simpler model.

	Estimator	Horizon	RMSE	SMAPE
(1)	Holt-Winters : Confidence Interval 0.95	1D	0.110255	8.787175
(2)	Holt-Winters : Confidence Interval 0.95	1D	0.109995	8.933436
	AUTOARIMA Confidence Interval 0.95	1D	0.105591	8.586032
	Prophet : Confidence Interval 0.95	1D	0.268158	28.280358
(1)	Holt-Winters : Confidence Interval 0.95	2D	0.143603	10.561279
(2)	Holt-Winters : Confidence Interval 0.95	2D	0.147547	10.963818
	AUTOARIMA Confidence Interval 0.95	2D	0.137931	10.171148
	Prophet : Confidence Interval 0.95	2D	0.367557	39.932455
	Estimator		RMSE	SMAPE
(1)	Holt-Winters : Confidence Interval 0.95	5D	0.151895	13.358853
(2)	Holt-Winters : Confidence Interval 0.95	5D	0.152307	13.009673
	AUTOARIMA Confidence Interval 0.95	5D	0.149630	13.248050
	Prophet : Confidence Interval 0.95	5D	0.425838	53.053558
			RMSE	SMAPE
(1)	Holt-Winters : Confidence Interval 0.95	7D	0.143730	12.417276
(2)	Holt-Winters : Confidence Interval 0.95	7D	0.144349	11.995640
	AUTOARIMA Confidence Interval 0.95	7D	0.142849	12.559702
	Prophet : Confidence Interval 0.95	7D	0.427191	55.659389
			RMSE	SMAPE
(1)	Holt-Winters : Confidence Interval 0.95	15D	0.122030	9.989042
(2)	Holt-Winters : Confidence Interval 0.95	15D	0.162186	14.804596
	AUTOARIMA Confidence Interval 0.95	15D	0.120703	10.148950
	Prophet : Confidence Interval 0.95	15D	0.575664	86.223182

Fig 4. Evaluating day-based forecast (1) SES (2) ETS

	Estimator	Horizon	RMSE	SMAPE
(1)	Holt-Winters : Confidence Interval 0.95	1H	0.065106	5.264319
(2)	Holt-Winters : Confidence Interval 0.95	1H	0.059378	5.265382
	AUTOARIMA Confidence Interval 0.95	1H	0.059449	5.165546
	Prophet : Confidence Interval 0.95	1H	0.215913	25.013151
(1)	Holt-Winters : Confidence Interval 0.95	3H	0.059530	4.539951
(2)	Holt-Winters : Confidence Interval 0.95	3H	0.054837	4.697469
	AUTOARIMA Confidence Interval 0.95	3H	0.049630	4.129375
	Prophet : Confidence Interval 0.95	3H	0.216061	25.112757
(1)	Holt-Winters : Confidence Interval 0.95	6H	0.058932	4.433506
(2)	Holt-Winters : Confidence Interval 0.95	6H	0.057419	4.791510
	AUTOARIMA Confidence Interval 0.95	6H	0.052536	4.360589
	Prophet : Confidence Interval 0.95	6H	0.198060	22.754783
(1)	Holt-Winters : Confidence Interval 0.95	12H	0.070218	5.903666
(2)	Holt-Winters : Confidence Interval 0.95	12H	0.069883	6.185771
	AUTOARIMA Confidence Interval 0.95	12H	0.068831	6.083571
	Prophet : Confidence Interval 0.95	12H	0.184289	20.172248

Fig 5. Evaluating day-based forecast (1) SES (2) ETS

Neural Network Turbidity Predictor

The networks were trained with a maximum of 20 epochs with early stopping for a seven-day prediction. No hyperparameter optimization was performed. The results obtained are compared with the ARIMA model. The CNN + LSTM model has performed better. The rest of the models have not been as accurate as ARIMA.

	RMSE	SMAPE
Autorima ARIMA(1,1,2) (1)	0.14	12.56%
Convolutional MLP	0.16	14.22%
CNN + LSTM (2)	**0.14**	**11.61%**
Bidirectional LSTM	0.16	14.05%
WaveNet	0.15	13.73%

Fig 6. Seven days forecast. ARIMA vs NN Models

The internal architecture of the best trained neural network model is shown in Fig. 7.

```
Model: "CNN_LSTM"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_7 (Conv1D)            (None, 671, 512)          1536
_________________________________________________________________
lstm (LSTM)                  (None, 671, 100)          245200
_________________________________________________________________
lstm_1 (LSTM)                (None, 671, 25)           12600
_________________________________________________________________
time_distributed (TimeDistri (None, 671, 1)            26
=================================================================
Total params: 259,362
Trainable params: 259,362
Non-trainable params: 0

Mean Absolute Error (MAE): 0.11
Median Absolute Error (MedAE): 0.1
Mean Squared Error (MSE): 0.02
Root Mean Squared Error (RMSE): 0.14
Mean Absolute Percentage Error (MAPE): 11.72 %
Median Absolute Percentage Error (MDAPE): 11.09 %
Symmetric Median Absolute Percentage Error (SMAPE): 11.61 %
```

Fig 7. CNN + LSTM Model

The parameters self-estimated by the Pdmarima package for the ARIMA(1,1,2) model are shown in Fig. 8.

SARIMAX Results

Dep. Variable:	y	**No. Observations:**	17473
Model:	SARIMAX(1, 1, 2)	**Log Likelihood**	12907.895
Date:	Sun, 20 Jun 2021	**AIC**	-25807.791
Time:	14:28:41	**BIC**	-25776.717
Sample:	0 - 17473	**HQIC**	-25797.557

Covariance Type: opg

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.4323	0.060	7.183	0.000	0.314	0.550
ma.L1	-0.9715	0.061	-15.848	0.000	-1.092	-0.851
ma.L2	0.1799	0.037	4.812	0.000	0.107	0.253
sigma2	0.0134	6.92e-05	193.008	0.000	0.013	0.013

Ljung-Box (L1) (Q):	0.00	**Jarque-Bera (JB):**	37520.49
Prob(Q):	0.96	**Prob(JB):**	0.00
Heteroskedasticity (H):	1.80	**Skew:**	1.45
Prob(H) (two-sided):	0.00	**Kurtosis:**	9.56

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).

Fig 8. ARIMA (1,1,2) Model

VI. Discussion

Throughout the paper, a set of techniques that have proven to be efficient have been employed to illustrate how to approach the turbidity modeling problem. Several models have been proposed and of them, only CNN+LSTM has shown to be slightly better in the seven-day forecast. During preprocessing two conditions have been illustrated that can affect the representativeness of the data. Regarding outlier removal, the Z-score technique was shown to be rather inefficient. This technique even pointed out as outliers large ranges of the signal that corresponded to legitimate events of rapid turbidity growth.

The Hample filter, on the other hand, was much more respectful of outlier removal, while respecting the signal geometry. This is very important when using neural network-based models since CNN models require mentioned geometry to get features from a signal. Generally speaking, due to the large variability of the turbidity signal, it is recommended that statistical operators based on medians be used in priority to

those based on means. Median is a more robust operator to outliers as they have proven to be more reliable: this includes error metrics and outlier removal. Z-score is based on the mean operator while the Hample filter is based on medians. All error metrics yielded fairly similar estimates, SMAPE however, clearly showed a real difference between the error of different models. This confirms in this work the quality of SMAPE as a comparator for time-series predictions as [7] pointed out.

The exponential smoothing and ARIMA methods have obtained very similar values. However, in general, ARIMA has been slightly more accurate. In the end, a good error estimate at 7 days is a value similar to that obtained by ARIMA or better. A variation of 1.5% in the estimate over the error made by ARIMA is quite acceptable. Virtually all models could serve as an acceptable predictor in the case at hand except the one modeled by Prophet. We should note, looking at the results (Fig. 4 and Fig 5) that although the quantitative value of the RSME predictions is not especially high, SMAPE especially penalizes Prophet. Thus, in the analyzed use case, Prophet's prediction is noticeably worse than the rest. It is important to note that the signal is not stationary, which by definition makes autoregressive and exponential smoothing methods much less accurate. Seasonality has hardly any weight in the model. This is not good news for autoregressive models, as they are not particularly accurate under these circumstances. That is why some hybrid models extract the trend from the signal. By doing so, the trendless signal becomes automatically stationary and therefore easily modeled by exponential or regressive methods. The analysis of the rest part, composed of noise and stationarity, is usually the part assigned to neural networks for empirical modeling.

In the scenario studied, the proposed neural models do not perform any signal decomposition. The reason is that the low importance of the seasonality leaves the prediction in the hands of the noise and the tendency components, so nothing would be gained by performing an additive decomposition prior to neural network. This is precisely one of the reasons why it is suggested to use other types of decompositions such as STL [14], SEATS [15] or X11 [15]. Any decomposition is fair as far as it allows to model the largest possible number of signal components, while the neural network takes care of the complex. EMD decomposition (T. Wang et al., 2012) is now very popular in the implementation of predictors following the hybrid model of component modeling by using neural networks.

VII. Conclusion

This is a too short route to draw many global conclusions. Especially when it comes to turbidity, which is featured as it cannot be generalized. However, this work fulfills its objective of providing researchers and technicians with a set of tools and criteria that allow an approach to the modeling of fluvial water turbidity. The tools and the criteria are strongly aligned with the objectives: they are easy to use and require minimal training in data science. We hope that this work not only provides a fair first contact with turbidity modeling, but also helps to ensure that modeling is not an unaffordable challenge for countless small community water treatment projects. To this end, a repository in Github, where the source code of the developed use case can be found in notebooks, is included in the appendix.

It's also hoped that this paper could be a starting point for many other applications related to water turbidity prediction. An example of this might be to develop an autonomous pumping system with the ability to catch highest possible volume of water with the lowest possible turbidity. The local turbidity predictor can also be applied to predict catastrophic events such as flooding or ecosystem contamination. In ecosystems where excess turbidity may represent an opportunity for the development of certain organisms or just the opposite, the predictor could be used for local assessment of environmental quality. In conjunction with cloud computing and the internet of things, local turbidity predictors can also be an interesting tool for water infrastructure management, predicting maintenance costs and chemical consumption. Overall, the paper's contribution broadens the possibility of facilitating new projects of interest to small communities that require a technical basis on local turbidity forecast.

Appendix

The Python source code (Jupiter Notebook format) and the datasets used in this work can be accessed through the following link: https://github.com/rulrulesforever/WaterTurbidityForecast

References

[1] LeChevallier, M. W., Evans, T. M., & Seidler, R. J. (1981). Effect of turbidity on chlorination efficiency and bacterial persistence in drinking water. Applied and Environmental Microbiology, 42(1), 159–167.

[2] Organization, W. H. (2017). Water quality and health-review of turbidity: information for regulators and water suppliers. World Health Organization.

[3] Gunawan, T. A., Daud, A., Haki, H., & Sarino. (2019). The Estimation of Total Sediments Load in River Tributary for Sustainable Resources Management. IOP Conference Series: Earth and Environmental Science, 248(1).

[4] Lawler, D. M., Petts, G. E., Foster, I. D. L., & Harper, S. (2006). Turbidity dynamics during spring storm events in an urban headwater river system: the Upper Tame, West Midlands, UK. Science of the Total Environment, 360(1), 109–126.

[5] Mendoza, L. M., Mladenov, N., Kinoshita, A. M., Pinongcos, F., Verbyla, M. E., & Gersberg, R. (2020). Fluorescence-based monitoring of anthropogenic pollutant inputs to an urban stream in Southern California, USA. Science of the Total Environment, 718.

[6] Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., & Maashi, M. (2020). Water Quality Prediction Using Artificial Intelligence Algorithms. Applied Bionics and Biomechanics, 2020.

[7] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54–74.

[8] Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. International Journal of Forecasting, 36(1), 75–85.

[9] Hewamalage, H., Bergmeir, C., & Bandara, K. (2021a). Recurrent Neural Networks for Time Series Forecasting: Current status and future directions. International Journal of Forecasting, 37(1), 388–427.

[10] Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15(3), 259–263.

[11] Grobbelaar, J. U. (2009). Turbidity. Encyclopedia of Inland Waters, 1, 699–704.

[12] Mushtaq, R. (2011). Augmented Dickey Fuller Test.

[13] Kokoszka, P., & Young, G. (2016). KPSS test for functional time series. Statistics, 50(5), 957–973.

[14] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. J. (1990). STL: A seasonal-trend

decomposition procedure based on loess. *Journal of Official Statistics*, *6*(1), 3–33.

[15] Dagum, E. B., & Bianconcini, S. (2016). *Seasonal adjustment methods and real time trend-cycle estimation*. Springer.

[16] Wang, T., Zhang, M., Yu, Q., & Zhang, H. (2012). Comparing the applications of EMD and EEMD on time-frequency analysis of seismic signal. Journal of Applied Geophysics, 83, 29–34.