Water Turbidity Prediction in Desalination Plants Using Machine Learning-Based Models

S. Lozano Iglesias[1], E. Lavín Pallero[2] & M. Ruiz García[1,3]

[1]Faculty of Physical Sciences, Complutense University of Madrid, 28040, Madrid, Spain
[2]School of Industrial Engineering, Polytechnic University of Madrid, 28006, Madrid, Spain
[3]Complex Systems Interdisciplinary Group, Complutense University of Madrid, 28040, Madrid, Spain

Introduction

Turbidity is a measure of the transparency of a liquid, quantified by the amount of light scattered by the material and expressed in NTU (Nephelometric Turbidity Units). This physical variable is challenging to describe analytically due to the unclear relationships it has with factors such as wind, ocean currents, dust, water temperature, and salinity.

However, **forecasting water turbidity is essential** in numerous contexts, particularly in the production of potable water.

Fig 1. High turbidity events pose significant challenges for desalination plants, as membrane-based processes and reverse osmosis systems cannot operate effectively under elevated turbidity levels [2], often resulting in operational shutdowns.

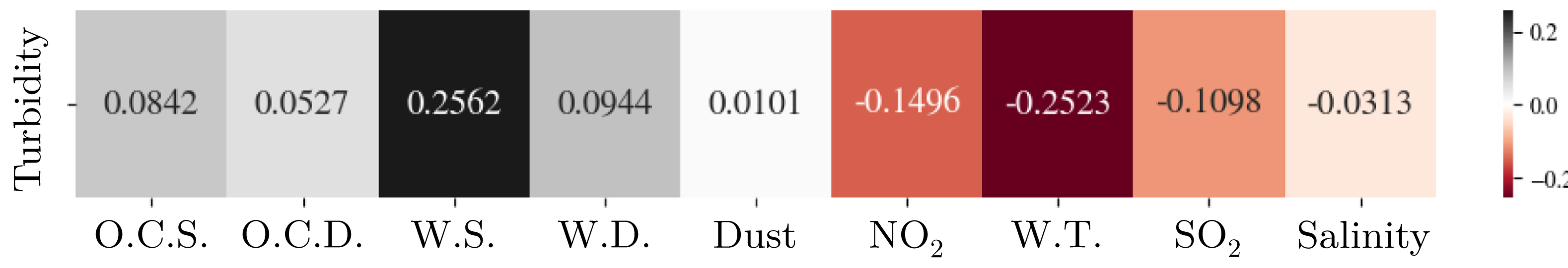


Fig 2. Pearson correlation coefficients between turbidity and various related variables indicate the presence of a relationship, though its exact nature remains unclear.

Regression Model vs. Classification Model

The initial intuitive approach is to predict the *exact turbidity value* in NTU that will be obtained at a specific moment in the future. This corresponds to a **regression model**.

Conversely, an alternative approach is to predict the *probability* that turbidity will reach a *certain level* at a specific moment in the future. This corresponds to a **classification model**.

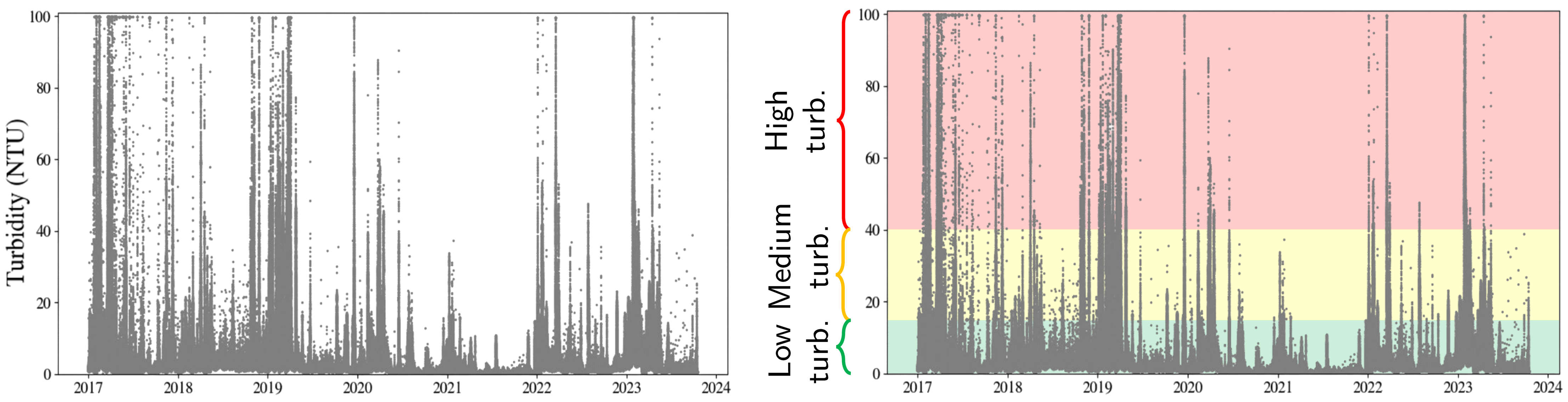


Fig 3. Turbidity data used for the regression model (left) and transformed into classes for the classification model (right).

Model Characteristics

LSTM (*long short-term memory*) neural network with six hidden layers and a total of 1.259.235 parameters.

Glorot normal initialization

$$w_{n,j} \sim U\left(-\sqrt{\frac{6}{M+S}}, \sqrt{\frac{6}{M+S}}\right)$$

ReLU activation function

$$\mathrm{ReLU}(a) = \begin{cases} 0 & \text{si} \quad a < 0 \\ a & \text{si} \quad a \geq 0 \end{cases}$$

Regression Model: mean squared error as loss function.

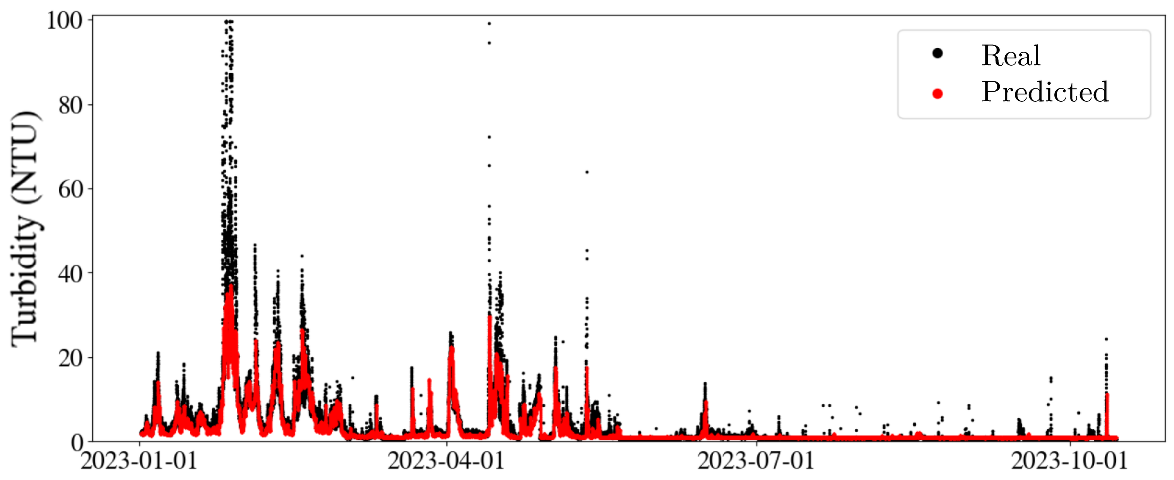
$$\mathcal{E}_{X,Y}(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N}\left(y^i - \hat{f}(\boldsymbol{x}^i)\right)^2$$

Classification Model: cross entropy as loss function.

$$\mathcal{E}_{X,Y}(\hat{f}) = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^i \cdot \log\left(\hat{f}_k(\boldsymbol{x}^i)\right)$$

Model Results

Regression Model

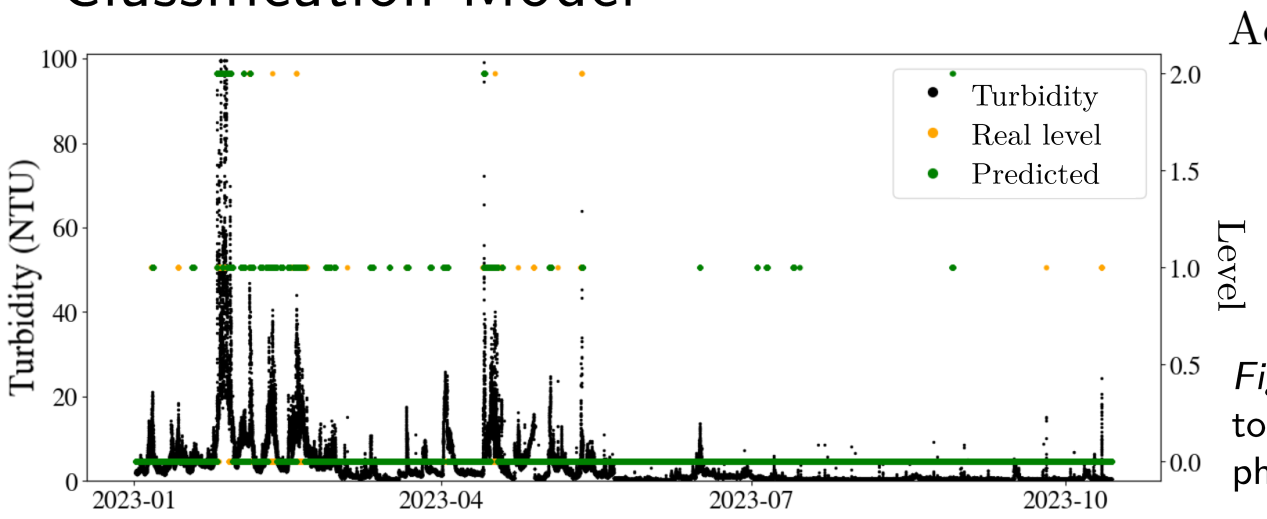


$$\text{Accuracy} = 1 - \frac{1}{N}\sum_{i=1}^{N}\frac{|y^i - \hat{f}(\boldsymbol{x}^i)|^2}{y^i}$$

	Training	Test
Accuracy	0,553	0,496

Fig 4. Predicted turbidity (red) compared to actual turbidity (black) during the test phase of the regression model.

Classification Model



$$\text{Accuracy} = 1 - \frac{1}{N}\sum_{i=1}^{N}|y^i - \hat{f}(\boldsymbol{x}^i)|$$

	Training	Test
Accuracy	0,951	0,938

Fig 5. Predicted levels (green) compared to the actual levels (yellow) in the test phase of the classification model.

- 87.9% of actual turbidity events are predicted to be at least a medium-class turbidity event.
- In 4.4% of cases where the model predicts a high turbidity event, a low turbidity event actually occurs.
- In 0.5% of instances where the model predicts a low turbidity event, a medium or high turbidity event is observed.

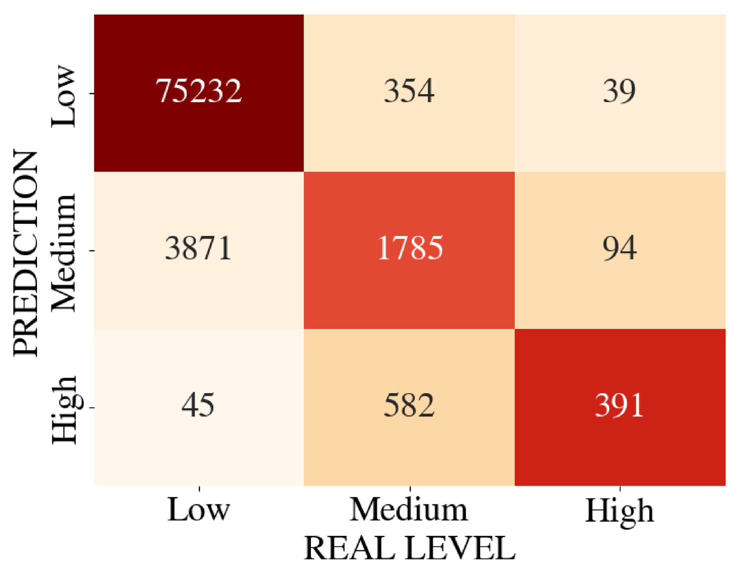


Fig 6. Confusion matrix illustrating predicted versus actual turbidity levels obtained during the test phase.

Classification in a Real Environment

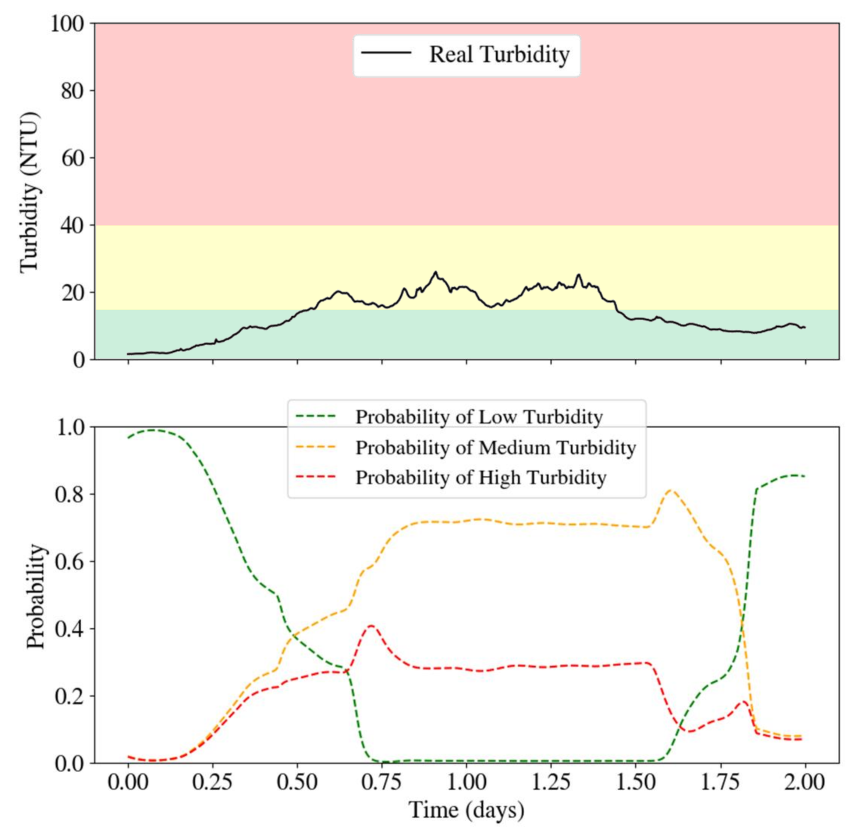


Fig 7. Comparison of the actual turbidity levels with the probabilities of the different classes predicted by the classification model, for an unknown future time during model training.

In this example, the model **accurately** captures the transition of turbidity from low to medium level and then back again to low level.

Further steps: Shap Results

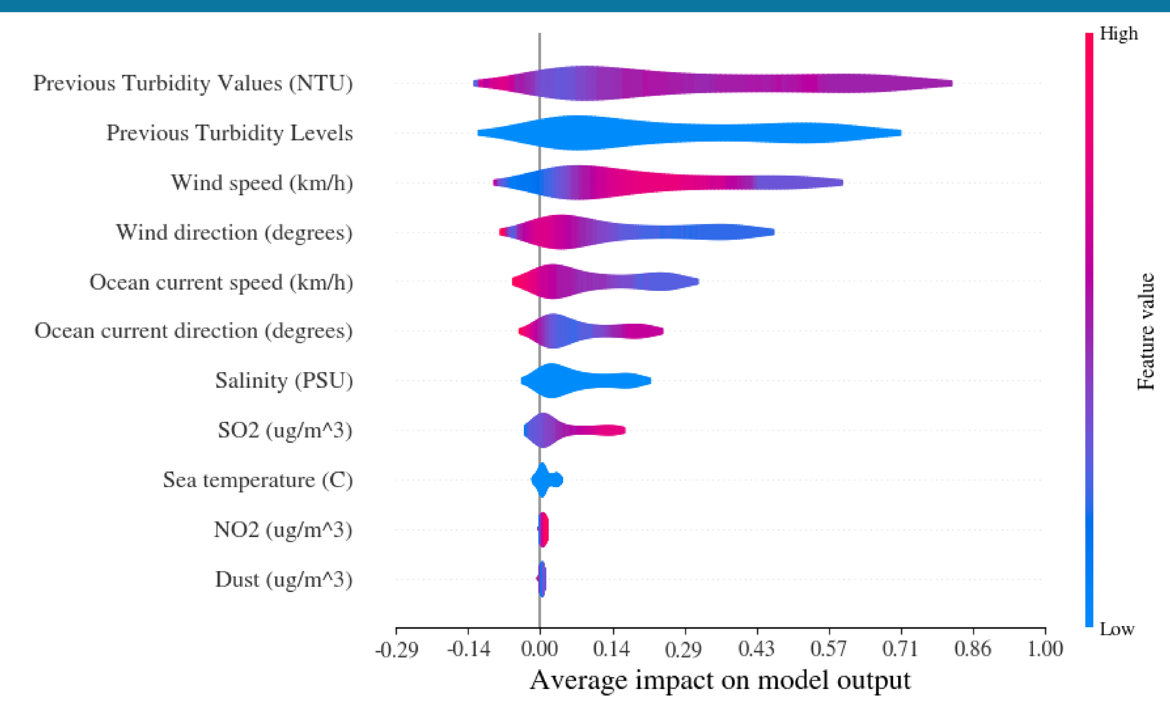


Fig 8. Analysis of variable impact on the classification model indicates that prior turbidity (value and level) is the most significant feature, followed by wind (speed and direction) and ocean currents (speed and direction).

References

[1] C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, Germany, August 2006.

[2] A. Altaee et al. Impact of High Turbidity on Reverse Osmosis: Evaluation of Pretreatment Processes. *Desalination and Water Treatment*, 208:96–103, 08 2020.

[3] S. Haykin. *Neural Networks and Learning Machines*. Pearson, Upper Saddle River, NJ, 3rd edition, November 2008.

[4] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

Contact

Samuel Lozano Iglesias
samuloza@ucm.es
+34 661 061 011