



OPEN

Comparing quantile regression methods for probabilistic forecasting of NO₂ pollution levels

Sebastien Pérez Vasseur & José L. Aznarte

High concentration episodes for NO₂ are increasingly dealt with by authorities through traffic restrictions which are activated when air quality deteriorates beyond certain thresholds. Foreseeing the probability that pollutant concentrations reach those thresholds becomes thus a necessity. Probabilistic forecasting, as opposed to point-forecasting, is a family of techniques that allow for the prediction of the expected distribution function instead of a single future value. In the case of NO₂, it allows for the calculation of future chances of exceeding thresholds and to detect pollution peaks. However, there is a lack of comparative studies for probabilistic models in the field of air pollution. In this work, we thoroughly compared 10 state of the art quantile regression models, using them to predict the distribution of NO₂ concentrations in a urban location for a set of forecasting horizons (up to 60 hours into the future). Instead of using directly the quantiles, we derived from them the parameters of a predicted distribution, rendering this method semi-parametric. Amongst the models tested, quantile gradient boosted trees show the best performance, yielding the best results for both expected point value and full distribution. However, we found the simpler quantile *k*-nearest neighbors combined with a linear regression provided similar results with much lower training time and complexity.

Pollution has become a worrying issue in cities due to its adverse effects on health and the increase in pollutant concentrations, mainly due to human activity (traffic, heating systems...) ¹. In order to take preventive steps to maintain air quality, forecasting the evolution of pollution levels becomes a useful tool for decision makers: detecting pollution peaks beforehand could give cities enough time to take and communicate effective measures.

Multiple research papers have focused on this issue and have dealt with the prediction of air quality. Bai et al. ² describe the state of the art in this exercise and collect a range of diverse solutions applied to this problem.

However, the prediction of the expected value of pollution concentrations through point-forecasting does not provide enough information about the likelihood of the pollutant levels reaching a certain threshold. Indeed, we have an estimate but we usually do not have a description of the confidence of the model nor the uncertainty in the predictions. Therefore, it is difficult to estimate the probability of the pollutant reaching above a certain threshold.

The reason this probability estimation is so important is because the measures taken by cities to limit pollution (for example, limiting traffic) impact the daily routines of citizens and prove themselves to be quite unpopular. Therefore, local governments need to have an estimation of the confidence in the prediction to safely engage in those preventive measures.

As noted by Hothorn et al. ³, the real objective in a regression analysis is to find the full conditional distribution of the target variable: in our case, the distribution of the concentration of the pollutants. Indeed, this full distribution gives an idea of the uncertainty of our predictions and can be useful to forecast the probability of the signal being above a certain threshold. For example, in the city of Madrid, hourly NO₂ concentrations in the air are considered to be harmful from 180 $\mu\text{g m}^{-3}$.

Previous research on the same dataset has already shown the usefulness of probabilistic forecasting for NO₂ levels ⁴, establishing the advantages of the approach and focusing on 1 h-ahead predictions with a single model (quantile random forests). We hereby extend that work by implementing other six models (quantile linear regression, quantile *k*-nearest neighbours, quantile gradient boosted trees, neural networks, distributional random forests, natural gradient boosting) and by using them for a wide set of forecasting horizons (up to 60 hours). Inspiration is drawn on some of the best approaches from the GEFCom forecasting competition ^{5,6}.

Artificial Intelligence Department, Universidad Nacional de Educación a Distancia — UNED, c/Juan del Rosal, 16, Madrid, Spain. email: jlaznarte@dia.uned.es

Furthermore, improving over these approaches, we also present a novel method to apply statistical inference to the output of the models. This method acknowledges the fact that the results show linear dependences between the predictors and the target, which slightly benefits linear models over nonlinear ones. By combining a linear model with nonlinear probabilistic modelling of its residuals we obtain optimized versions of the standard models.

Probabilistic forecasting with quantile regression

The prediction from most regression models is a point estimate of the conditional mean of a dependent variable, or response, given a set of independent variables or predictors. However, the conditional mean does not provide a complete summary of the distribution, so in order to estimate the associated uncertainty, quantiles are in order. Then the median or 50 percentile can serve as a prediction of the center and for example, the 95 percentile represents the value of the response with 95% of the predicted points below. Recent advances in computing have inducted the development of regression models for predicting given quantiles of the conditional distribution. The technique is called quantile regression (QR) and was first proposed by Koenker in 1978⁷ based on the intuitions of the astronomer and polymath Rudjer Boscovich in the 18th century. Elaborating from the same concept of estimating conditional quantiles from different perspectives, several statistical and CI models that implement this technique have been developed: from the original linear proposal to multiple or additive regression, neural networks, support vector machines, random forests etc.

Quantile regression has gained an increasing attention from very different scientific disciplines⁸, including financial and economic applications⁹, medical applications¹⁰, wind power forecasting¹¹, electric load forecasting¹², environmental modelling¹³ and meteorological modelling¹⁴ (these references are just examples and the list is not exhaustive). To our knowledge, despite its success in other areas, quantile regression has not been applied in the framework of air quality, with the exception of¹⁵.

Thus, as we can estimate an arbitrary quantile and forecast its values, we can also estimate the full conditional distribution and fit it to a known distribution. This will entail us to the results presented in section “[Results and discussion](#)”.

Among the growing array of methods that allow to estimate and forecast data-driven conditional quantiles, in this study we have chosen to compare linear regression, k -nearest neighbors, random forests and gradient boosted trees. We took the point-estimate version of those models and converted them to their quantile or probabilistic counterparts. We can therefore compare each model not only on the accuracy of their point estimation but on the confidence of each model.

We will compare the different algorithms through the Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and bias for the quantile 50 and the continuous ranked probability score (CRPS) metric for the forecast distribution. As described by Gneiting et al.¹⁶, CRPS is a measure of the squared difference between the forecast cumulative distribution function (CDF) and the empirical CDF of the observation.

Data description and experimental design

Nitrogen dioxide. The city of Madrid has an air quality monitoring system composed by 24 stations which capture hourly data for NO₂. For this study, we have selected one of the stations with higher average levels: *Escuelas Aguirre* station (code 28079008).

As we can see in Fig. 1, the shape of the histogram approaches the one from a lognormal distribution and therefore we transformed to the logarithm of the values. This has 2 positive effects: it reduces the tail of the distribution which will enable better quantile estimation and it reduces the skewness of the distribution which helps with linear models like the linear quantile regression.

The time series for this station consists of hourly measured values of the concentrations of NO₂ from 01/01/2013 to 31/12/2019. These values exhibit a clear intraday pattern, in which the higher values are located in two peaks around the morning and evening (with highest average value around 19 h) while the nightly hours (from 00 h to 05 h) have lower average concentrations. Not only are the values higher at those hours, but also the variance is, as we can see in Fig. 1.

In order to analyze the seasonality of the signal, we extract the 5 main factors from the Fourier transform. Those correspond to the main repetitive patterns found on the series, and can be seen clearly from the first 3000 components. The series shows certain seasonality for 12-h, 24-h, 1 week (168-h) and one year. Therefore, we will create, and use as inputs for the models, the output of periodic functions (cosine and sine) whose frequency is equal to the ones stated above. This will enable the machine learning models to learn the seasonality of our time series.

As is common when forecasting with machine learning models, we exploit the inertia of the modelled series by adding lagged variables to the inputs. Of course, in doing so, we are limited by the horizon of the prediction and by the ‘curse of dimensionality’, which implies keeping a limited number of features as input. In our case, the inertia of the series will be modeled by lagged values from the immediate past (hours before) and, based on the seasonal analysis: 1–5 h before and every 11–13 h up to 9 days before.

Ozone. The same station that records Nitrogen dioxide also records the levels of ozone (O₃). It is known that ozone and Nitrogen dioxide are related by chemical reactions occurring in the atmosphere in the presence of sunlight, especially of the UV spectrum. Thus, we will also add lagged values of O₃ as inputs to our models.

ECMWF numerical pollution prediction. The European Centre for Medium-Range Weather Forecasts (ECMWF) implements the Copernicus Atmosphere Monitoring Service. This service delivers a daily production of near-real-time European air quality analyses and forecasts with a multi-model ensemble system. Although

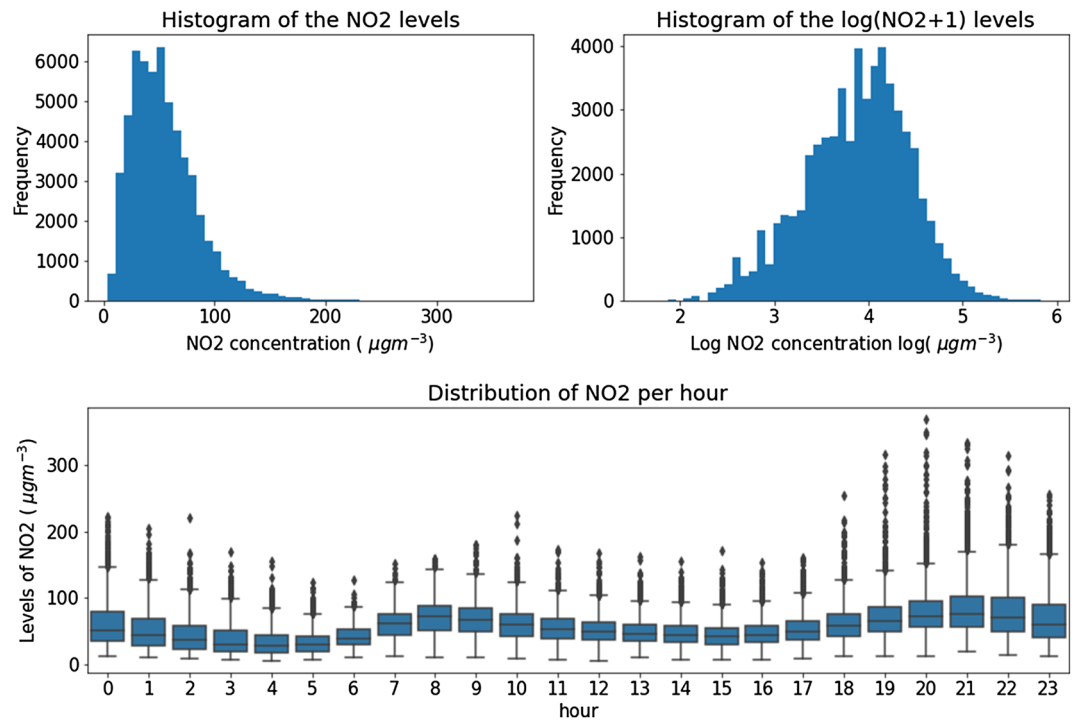


Figure 1. Distribution of logarithmic NO₂ and distribution of NO₂ per hour.

these forecasts are a very good starting point³, the resolution of the model is 10km and hence it is not expected to be capable of modelling the local urban effects of the NO₂ series under study.

Calendar variables. As NO₂ levels are clearly be linked to human activity, we will also flag the hours belonging to a specific type of day. Days could be classified as bank holidays, heavy traffic days (for example, return from holidays), school holidays...We will also use as inputs to the models past values of this variables (1, 2 and 7 days before).

Experimental design. As a summary, we use the following predictors: NO₂ measures lagged 1–5 h and every 11–13 h up to 9 days before, O₃ levels lagged every 24 h up to 4 days before, calendar variables lagged 1, 2 and 7 days before, ECMWF predictions and seasonal features extracted from the Fourier analysis. This amounts to a total of 102 independent variables.

When performing the experiments, first we aligned and gathered all the hourly time series: NO₂, O₃, ECMWF and calendar variables. Then we transformed the signal levels and then we added the lagged values and a seasonal time series with the main periods of the NO₂ time series.

Once all this process is finalized, we train the following probabilistic models: quantile random forests (QRF), *k*-nearest neighbors (QKNN), quantile linear regression (QLR) and quantile gradient boosting (QGB), Multi-layer perceptron (MLP), Distributional random forests (DT) and NGBoost (NGBOOST). Figure 2 shows the data flow in the experimental design. All the hyperparameters of the models have been estimated through grid search on a validation set. In section “Probabilistic models” we provide more information on each of the models.

Concerning cross-validation, there are several accepted methods to separate the train and test set which produce correct estimations of the error¹⁷: We can approximate the error of a model through a cross-validated evaluation as long as we removed from each test set the correlated samples to their respective training dataset. We chose a cross-validation with 5 splits for time execution reasons. We will always test with predictions done at 10:00, as this is the time the forecast is done in the operational setting, as the data is first available at that time.

We want to forecast the full distribution of NO₂ levels for the next 60 h and therefore we will train and evaluate the models for each hour (60 horizons).

After forecasting the quantiles, we will fit them to a normal distribution. Fitting a normal distribution to the predicted quantiles and then generating the percentiles for that fitted distribution has several advantages. It enables the calculation of more percentiles from a small number of them. It also helps estimating the upper tail of the distribution, in spite of the low probability for those values.

We will evaluate the predicted 50 percentile through standard evaluation metrics (RMSE and bias), and the predicted distribution through the CRPS.

We will perform this evaluation for each of the models and each of the horizons.

Probabilistic models. As stated above, we will compare seven different probabilistic models, which are briefly described below for reference. We will provide alongside the models the abbreviation we will use for each

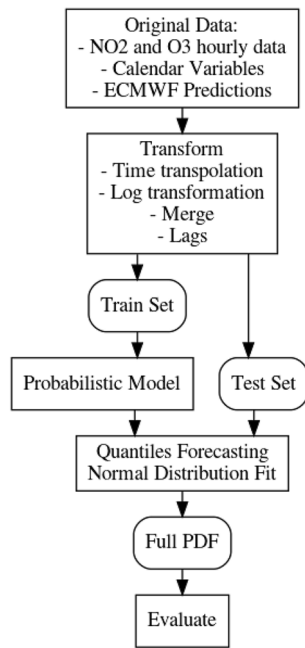


Figure 2. Data flow of the experiments.

of them throughout this article. Also we are adapting point-estimation algorithms to their probabilistic counterparts. This allows us to see the uncertainty these models have. Indeed, metrics like RMSE and Bias are linked to the confidence of the models for point estimation but the predicted CDF is much better at describing it. This also increases the interpretability of those models. Therefore the implementations we described can have applications beyond this forecast exercise.

Quantile linear regression (QLR). As shown in⁷, we can apply linear regression with a modified cost function in order to predict the quantiles of the dependent variable. Given a set of vectors (x_i, y_i) , in the usual point forecasting approach we are usually interested in the prediction $\hat{y}(x) = \alpha_0 + \alpha_1 x$ which minimizes the mean squared error,

$$E = \frac{1}{n} \sum_i \epsilon_i = \frac{1}{n} \sum_i [y_i - (\alpha_0 + \alpha_1 x)]^2. \quad (1)$$

This prediction is the conditional sample mean of y given x , that is, $\hat{y}(x) = \hat{\alpha}_0 + \hat{\alpha}_1 x$, or the location of the conditional distribution. But we could be interested in estimating the conditional median (i.e., the 0.5 quantile) instead of the mean, in which case we should find the prediction $\hat{y}(x)$ which minimizes the mean absolute error,

$$E = \frac{1}{n} \sum_i \epsilon_i = \frac{1}{n} \sum_i |y_i - (\alpha_0 + \alpha_1 x)|. \quad (2)$$

The fact is that, apart from the 0.5 quantile, it is possible to estimate any other given quantile τ . In that case, instead of (2), we could minimize

$$E = \frac{1}{n} \sum_i f(y_i - (\alpha_0 + \alpha_1 x)) \quad (3)$$

where

$$f(y - q) = \begin{cases} \tau(y - q) & \text{if } y \geq q \\ (1 - \tau)(q - y) & \text{if } y < q \end{cases}, \quad (4)$$

with $\tau \in (0, 1)$. Equation (3) represents the median when $\tau = 0.5$ and the τ -th quantile in any other case.

We will train 5 linear regression models to predict 5 percentiles of the signal. Those 5 percentiles will enable us to calculate the mean and standard deviation of a normal distribution. This normal distribution will provide the predicted quantiles.

Quantile k -nearest neighbors (QKNN). We will use the probabilistic k -nearest neighbors algorithm as described in⁶. This algorithm is based on the standard k nearest neighbor, where instead of calculating the mean of the

targets of the k nearest points to the input, it fits a normal distribution to the targets of those neighbors. We build the quantiles from the predicted normal distribution.

Quantile random forests (QRF). Quantile random forests create probabilistic predictions out of the original observations. They work like the usual random forest, except that, in each tree, leaf nodes do not contain a single value as a prediction but the target observations from the training set belonging to that leaf.

Then predictions are calculated by selecting the leaf nodes in each tree corresponding to the input features and combining the weighted histograms in each tree out of the target observations in those leaf nodes. For more information refer to¹⁸. We then fit a normal distribution to the predicted histogram to calculate the predicted quantiles.

Quantile gradient boosted trees (QGB). Tree boosting¹⁹ is a successful machine learning technique that consists of growing trees based on the compromise of a cost function and a regularization function. This technique has already been used to predict air pollution, as documented by Lee et al.²⁰. The cost function is usually used to forecast the mean of the signal. We will modify the cost function (in a similar way as in the quantile linear regression) to predict the quantiles of the target. Also, we will use the `lightgbm` implementation²¹ which provides lower training times and higher accuracy.

We will train 5 gradient boosted trees models to predict 5 percentiles of the NO₂ signal. Then, we will use the percentiles to fit a normal distribution and calculate the predicted quantiles.

Multilayer perceptron. We will also test a multilayer perceptron (MLP)²² with 1 hidden layer. Like for quantile linear regression and quantile gradient boost, we will modify the cost function to predict the quantiles of the target.

We will therefore as well train 5 MLP to predict 5 percentiles of the NO₂ signal and calculate the final quantiles through a normal distribution as for the quantile linear regression and quantile gradient boosted tree.

Distributional random forests. Based on the idea of GAMLSS, Distributional Forests²³ extend the power of this semiparametric method by estimating the LSS (location, scale and shape) through the use of random forests, instead of using GAM's. We will use the R implementation of this method from the package `distree`.

NGBoost. NGBoost²⁴ uses the natural gradient instead of the standard gradient to build a boosted ensemble learner that predicts directly the parameters of the distribution of the target variable. Natural Gradients are presented as offering better stability and robustness than normal gradients.

Probabilistic forecast of linear regression residuals. We suspect there are high linear dependencies between the input predictors and the target. We will then experiment on combining 3 of the models: quantile random forest, quantile k -nearest neighbor and quantile gradient boost with a linear regression.

We decided to train a linear regressor which predicts the NO₂ values and then use the QRF, QKNN and QGB models to predict the full distribution of the residuals of that linear regression. We abbreviate both models respectively with QRFL, QKNNL and QGBL.

Results and discussion

Figure 3 shows the different metrics for each model across all the horizons. First, we see how quantile gradient boosted trees (QGB) outperforms the other models and displays better scores for all metrics. The additive nature of QGB is behind these results.

Quantile random forests and quantile k -nearest neighbors underperform compared to the other models, showing a bias which is clearly higher compared to the others. The main reason for this can be attributed to the highly linear dependence of the data. This also explains the relatively good results of the linear model (QLR). However, the linear model underperforms when compared to other models, as it is unable to learn the non-linear relationships between the predictors and the target.

Regarding the models predicting the residuals of a linear regression, QKNNL and QRFL showcase good performance, which is remarkable for QKNNL due to the simplicity of this approach. The KNNL offers similar results to QGB with lower complexity and lower training times. On the other hand, QGBL brings no real benefit compared to QGB, since QGB can already integrate the linear component.

Distributional forests (DT) and natural gradient boosted trees (NGB) present relatively good results but are still outperformed by QGB. Natural gradients bring no real benefit in this case.

Finally, MLP has good bias results but underperforms against the other methods.

As stated earlier, the NO₂ levels follow a lognormal distribution and thus it seems that it is better modeled with a multiplicative model (since different causes multiply the levels of pollution), therefore the logarithm of the NO₂ levels is better forecast with an additive model. This would explain why quantile gradient boosted trees consistently outperforms the other models: it can naturally add the nonlinear effects.

Table 1 displays the mean and standard deviation of CRPS of the different models. The table shows again the good performance of the quantile gradient boosted trees model for CRPS. The table also shows the training times of the different models. QGB appears as having a good compromise between low training times and metrics. QKNNL offers good performance with really low training times.

As stated above, CRPS is a good summary of the performance of probabilistic models. Notwithstanding, the reliability and sharpness graphs are known to be useful at estimating how the observed values are positioned in the distributions. Figure 4 features the reliability and sharpness of the different models.

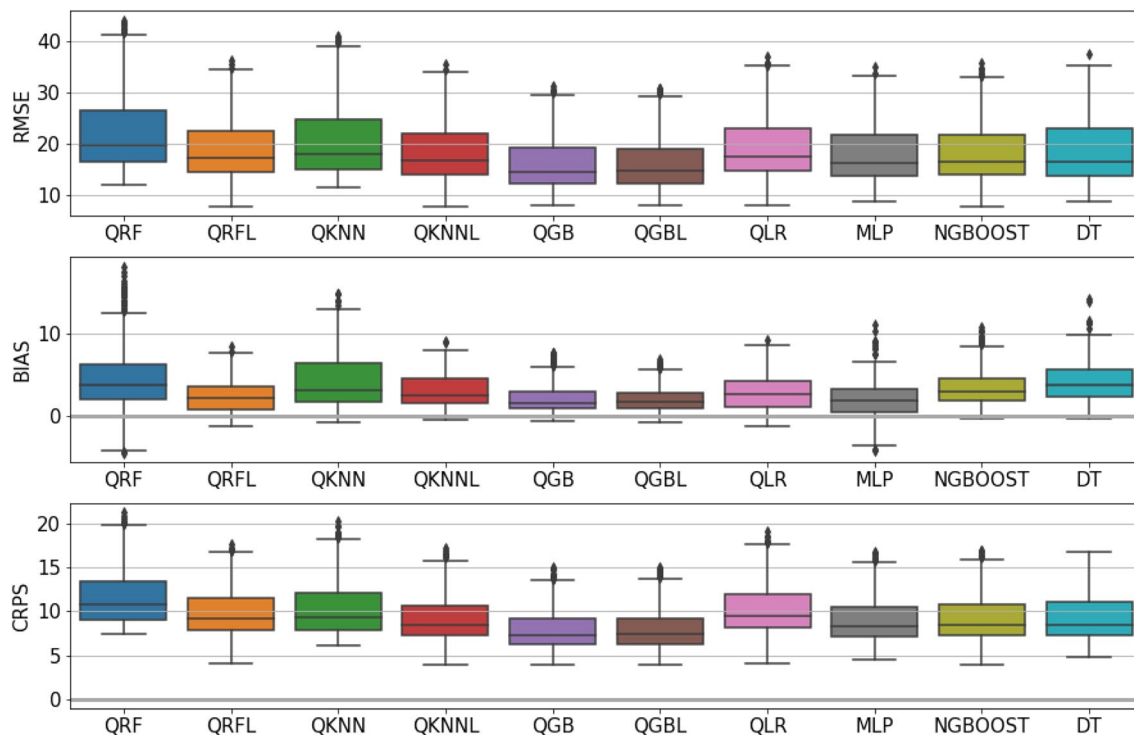


Figure 3. Boxplot of continuous ranked probability score, root mean squared error and bias of the different models for all horizons.

method	RMSE	CRPS	bias	training time (s)
QRF	22.62 (8.37)	11.86 (3.58)	4.68 (4.77)	24.15 (1.01)
QRFL	19.12 (6.52)	9.97 (3.03)	2.32 (2.0)	8.71 (0.36)
QKNN	20.73 (7.91)	10.52 (3.59)	4.26 (3.53)	6.92 (0.31)
QKNNL	18.4 (6.44)	9.34 (2.99)	3.06 (2.01)	7.06 (0.32)
QGB	16.09 (5.59)	8.04 (2.54)	2.08 (1.76)	42.78 (1.19)
QGBL	16.14 (5.55)	8.1 (2.55)	1.95 (1.55)	42.22 (1.04)
QLR	19.44 (6.71)	10.38 (3.31)	2.77 (2.16)	9.68 (0.53)
MLP	18.16 (6.04)	9.27 (2.85)	1.86 (2.49)	73.08 (7.71)
NGBOOST	18.52 (6.43)	9.4 (2.96)	3.42 (2.32)	180.86 (14.2)
DT	18.69 (6.77)	9.45 (2.99)	4.27 (2.84)	289.81 (12.79)

Table 1. Error measures for the proposed models.

The sharpness is given by the distribution of the forecast probabilities (small histogram overimposed in the graphs). NGBOOST, MLP and QKNNL display the best results as their histograms are almost flat.

For QRF, QRFL, QKNN and DT, we see fewer observations on the sides of the predicted distributions. This means the variance of the predicted variance is too high and the predictions have too much uncertainty. It is specially noticeable for QRF and DT.

On the contrary, QGB, QGBL and QLR seem more balanced but display high values at both sides of the sharpness curve. If we consider the distribution to be a Gaussian distribution, this means the forecast distributions tend to have a too small standard deviation and are too narrow. Therefore, the forecast probability of some levels of NO₂ is too small compared to the observed one. QGBL and MLP also display this behaviour. Since, in this framework, it is common that the main interest lies in predicting the higher tail of the distribution (representing peaks of pollution), this is not considered to be a critical problem in this case. Finally, we see again the QKNNL model performing similarly to the QGB model, offering a simpler alternative to QGB.

Reliability is shown in the main charts of Fig. 4. The light blue lines represent each of the reliability diagrams for each horizon, while the dark blue line shows the average reliability diagram for a single model over all the horizons. This let us see how stable is the reliability per horizon.

It is noticeable that QGBL and QKNNL show the best results both in terms of the average diagram and the dispersion of the different diagrams per horizon. MLP and NGBOOST have a good average line but also display a

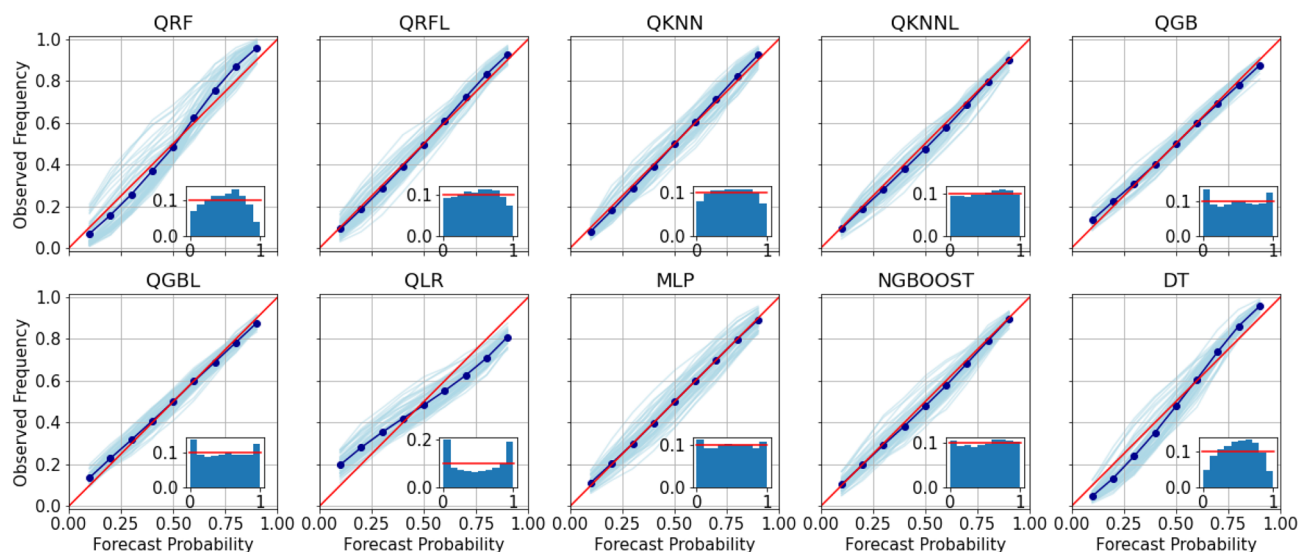


Figure 4. Average reliability and sharpness of the different models across all horizons. The dim blue lines correspond to the different horizons.

Algorithm	Ranking
QRF	7.354393
QRFL	4.888386
QKNN	6.522557
QKNNL	3.379697
QGB	0.967290
QGBL	1.297199
QLR	5.791688
MLP	2.883587
NGBOOST	3.442853
DT	3.921512

Table 2. Average rankings of the algorithms (Quade).

high variability for each horizon. On the other hand, QRF, QRFL, QKNN and DT also show a high variability and again display the high uncertainty of their predictions, which is seen in the extremes of the reliability diagram. On the contrary, QGB and QLR show the opposite behaviour with narrow predicted variances.

We also see that NGBOOST and QKNNL seem to predict better the tail of the distributions (quantile 90). This is relevant as we are interested in the higher values of NO_2 pollution levels.

In order to compare the result of the 10 methods, we will use advanced non parametric tests as described by García et al.²⁵. The Quade test will calculate the rank of all the algorithms and check if they are significantly different from the mean rank. We have 60 different datasets, each one corresponding to a horizon where we have performed a fivefold crossvalidation and kept the mean of the results. The rank is calculated not only based on the results in the different datasets but also on the importance of the dataset (based on the variance of results), this being the main difference with a Friedmann test. The result of the Quade test gives a p-value of 1.11×10^{-16} which rejects the null hypothesis that the methods have similar performance and thus we can use the ranks to classify the different methods. Table 2 shows the ranks of the Quade test.

Conclusions

After extracting and processing the data from one of the pollution stations in Madrid, we have compared 10 different models to build a probabilistic forecast of the levels of NO_2 for up to 60 hours into the future. We have evaluated our models through the forecast quantile 50 and the full forecast distribution. We have also adapted the predicted quantiles to fit a normal distribution.

We have observed a linear dependence between the target and the features. For this reason, the linear quantile regression has performed well compared to random forests and k -nearest neighbors. However, the multiplicative nature of the levels of NO_2 and the nonlinear dependence between target and features have lead to better results for the gradient boosted trees which has outperformed all the other models in all metrics.

However, we have shown how quantile random forest and quantile k -nearest neighbors could be used to improve the results of a linear model when nested to model the full distribution of the residuals of a linear

regression. Those models, specially the k -nearest neighbor are easier to train, so they become worthy and simpler alternatives to the gradient boosted trees.

These results are especially useful for practitioners that need to choose a probabilistic model for air quality time series and other similar problems with a strong anthropogenic component.

Received: 13 July 2020; Accepted: 5 May 2021

Published online: 02 June 2021

References

- Jiang, L. & Bai, L. Spatio-temporal characteristics of urban air pollutions and their causal relationships: Evidence from Beijing and its neighboring cities. *Sci. Rep.* **8**(1), 1279.
- Bai, L., Wang, J., Ma, X. & Haiyan, L. Air pollution forecasts: An overview. *Int. J. Environ. Res. Public Health* **15**(4), (2018).
- Hothorn, T., Kneib, T. & Bühlmann, P. Conditional transformation models. *J. R. Stat. Soc. B* **76**(1), 3–27 (2014).
- Aznarte, J. L. Probabilistic forecasting for extreme no 2 pollution episodes. *Environ. Pollut.* **229**, 321–328 (2017).
- Hong, T. *et al.* Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *Int. J. Forecast.* **32**(3), 896–913 (2016).
- Mangalova, E. & Shesterneva, O. K-nearest neighbors for gecom2014 probabilistic wind power forecasting. *Int. J. Forecast.* **32**(3), 1067–1073 (2016).
- Koenker, R. & Bassett, G. Regression quantiles. *Econometrica* **46**(1), 33–50.
- Yu, K., Lu, Z. & Stander, J. Quantile regression: Applications and current research areas (2003).
- Rejeb, A. B. & Arfaoui, M. Financial market interdependencies: A quantile regression analysis of volatility spillover. *Res. Int. Bus. Financ.* **36**, 140–157 (2016).
- Jang, Y., Kim, J. H., Lee, H., Lee, K. & Ahn, S. A. A quantile regression approach to explain the relationship of fatigue and cortisol, cytokine among Koreans with Hepatitis b. *Sci. Rep.* **8**(1), 16434 (2018).
- Wan, C., Lin, J., Wang, J., Song, Y. & Dong, Z. Y. Direct quantile regression for nonparametric probabilistic forecasting of wind power generation. *IEEE Trans. Power Syst.* **32**(4), 2767–2778 (2017).
- Lebotsa, M. E. *et al.* Short term electricity demand forecasting using partially linear additive quantile regression with an application to the unit commitment problem. *Appl. Energy* **222**, 104–118 (2018).
- Cade, B. S. & Noon, B. R. A gentle introduction to quantile regression for ecologists. *Front. Ecol. Environ.* **1**(8), 412–420 (2003).
- Baur, D., Saisana, M. & Schulze, N. Modelling the effects of meteorological variables on ozone concentration: A quantile regression approach. *Atmos. Environ.* **38**(28), 4689–4699 (2004).
- Martínez-Silva, I., Roca-Pardiñas, J. & Ordóñez, C. Forecasting SO₂ pollution incidents by means of quantile curves based on additive models. *Environmetrics* **27**(3), 147–157 (2016).
- Gneiting, T. & Katzfuss, M. Probabilistic forecasting. *Annu. Rev. Stat. Appl.* **1**(1), 125–151 (2014).
- Bergmeir, C., Hyndman, R. J. & Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Comput. Stat. Data Anal.* **120**, 70–83 (2018).
- Meinshausen, N. Quantile regression forests. *J. Mach. Learn. Res.* **7**, 983–999 (2006).
- Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**(5), 1189–1232 (2001).
- Lee, M. *et al.* Forecasting air quality in taiwan by using machine learning. *Sci. Rep.* **10**(1), 4153.
- Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **30**, 3146–3154 (2017).
- Ramchoun, H. Multilayer perceptron: Architecture optimization and training. *Int. J. Interact. Multimed. Artif. Intell.* **4**, 26–30.
- Schlosser, L., Hothorn, T., Stauffer, R., & Zeileis, A. Distributional regression forests for probabilistic precipitation forecasting in complex terrain. <http://arxiv.org/abs/1804.02921>.
- Duan, T. *et al.* NGBoost: Natural gradient boosting for probabilistic prediction. <http://arxiv.org/abs/1910.03225>.
- García, S., Fernández, A., Luengo, J. & Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.* **180**(10), 2044–2064.

Author contributions

S.P. collected and curated the data, contributed to the design of the experiments, wrote the code, carried out the experiments and wrote the first version of the paper. J.L.A. conceived, planned and designed the analysis, contributed to the interpretation of the results and coordinated the work. Both authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.L.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021