Atmospheric Environment 89 (2014) 22–28

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Application of air quality combination forecasting to Bogota

Joakim Westerlund [a,*], Jean-Pierre Urbain [b], Jorge Bonilla [c,d]

[a] Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance, Melbourne Burwood Campus, 221 Burwood Highway, VIC 3125, Australia
[b] Maastricht University, The Netherlands
[c] University of Gothenburg, Sweden
[d] Universidad de los Andes, Colombia

HIGHLIGHTS

- The paper considers a relatively novel forecasting method, combination forecasting.
- Combination forecasting is shown to outperform the standard neural network approach.
- Combination forecasting is applied to a unique data set for Bogota.

ARTICLE INFO

Article history:
Received 16 August 2013
Received in revised form
5 February 2014
Accepted 8 February 2014
Available online 12 February 2014

JEL classification:
C45
C53
Q53

Keywords:
Air quality forecasting
Bogota
Forecast combination
Neural networks

ABSTRACT

The bulk of existing work on the statistical forecasting of air quality is based on either neural networks or linear regressions, which are both subject to important drawbacks. In particular, while neural networks are complicated and prone to in-sample overfitting, linear regressions are highly dependent on the specification of the regression function. The present paper shows how combining linear regression forecasts can be used to circumvent all of these problems. The usefulness of the proposed combination approach is verified using both Monte Carlo simulation and an extensive application to air quality in Bogota, one of the largest and most polluted cities in Latin America.

1. Introduction

Air quality modeling and forecasting have become rapid growth areas in recent years. The main reason for this is the increased awareness of the adverse effects of a wide range of pollutants such as carbon monoxide (CO), particulate matter (PM_{10}), ground level ozone (O_3), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), and sulfur dioxide (SO_2) (see Lippmann, 2003; Brunekreef and Holgate, 2002; Kassomenos et al., 2008). Since air quality is a public good its socially optimal level of provision cannot be ensured through markets. Therefore, the responsibility of this provision is generally given to the environmental authorities, which have to set policies and regulations to mitigate pollution externalities. Air quality forecasting is one of the tools available to these institutions to manage health effects and air pollution events.

In their recent overview of the literature, Zhang et al. (2012) divide the main air quality forecasting approaches in three; (i) physically-based deterministic approaches, (ii) empirical approaches, and (iii) statistical approaches. Deterministic "3-D" air quality forecasting combines models of emissions with those of meteorological and chemical atmospheric processes, and has been shown to lead to accurate forecasts. Unfortunately, this approach involves scarce data, it is computationally costly, difficult to operate, and requires a high level of expertise. Empirical approaches such as climatology and persistence forecasting are, by contrast, reasonably simple and inexpensive to operate. However, they are also quite unreliable.[1] Statistical

* Corresponding author.
E-mail address: j.westerlund@deakin.edu.au (J. Westerlund).

[1] Indeed, persistence forecasting simply says that tomorrow's pollution levels will be the same ones that applied today.

http://dx.doi.org/10.1016/j.atmosenv.2014.02.015
1352-2310/

forecasting requires relatively less detailed data, it is inexpensive and easy to operate, and, in contrast to empirical approaches, is very accurate. In fact, statistical approaches have been shown to outperform even advanced deterministic approaches (see, for example, Diaz-Robles et al., 2008).

But while attractive in many regards, statistical forecasting also has a drawback in that it is not equipped to handle model uncertainty. Forecasters often encounter uncertainty about what variables to include in their models. As a result, they frequently end up estimating a variety of models before deciding on the one to use. This practice leads to a number of pathologies. First, it understates the uncertainty about the forecast. Basing the forecast on a single model implicitly assumes that the probability that the reported model generated the data is one, an assumption that is surely mistaken. Second, some forecasters search the model space until they find a specification with good forecasting performance, a practice that has led to indications of publication bias. As a result, reported results are often fragile to slight variations in model specification.

One way to circumvent these difficulties altogether is to use forecast combination (FC). The idea here is that, with multiple competing models at hand, each having its own strengths and weaknesses, rather than insisting on finding the one best forecast, it should be possible to combine the individual forecasts into a single forecast that is at least as good as any of the individual forecasts. This approach has been shown to work well in econometrics and it will be used also in the present paper. The purpose is to propose FC as an alternative to more common statistical air quality forecasting approaches such as linear regression (LR) and neural network (NN), and in so doing we will consider a unique data set for Bogota.

There are many reasons for looking particularly at Bogota. First, as the capital city of Colombia, Bogota is the fifth most populated city in Latin America with around 7.4 million inhabitants.[2] Second, although industrial emissions have long been considered to be one of the most severe pollution problems, emissions from traffic are nowadays an increasing concern.[3] Third, air pollutant concentrations have at times been well above the national air quality standard (AQS), specially for PM_{10} and O_3.[4] Needless to say, such high levels of pollution are very costly for society. In fact, the price of PM_{10} alone is estimated at a staggering USD 46 million per year (Lozano, 2004). Fourth, at present the environmental and health agencies of Bogota lack an air pollution forecasting system.

2. NN, LR and FC

As alluded in the introduction, users of modern forecasting techniques in environmental sciences are faced with an abundance of predictor variables and a plethora of methods for generating forecasts. An important issue is therefore whether to adopt a forecasting strategy that seeks out a single best forecasting method or, alternatively, attempt to combine forecasts generated by different models.

Most statistical work in the environmental literature is based on using single forecasting methods. In fact, as far as we know, this is the first study to consider FC. Among the many single forecasting methods available, NN have received most attention by far (see, for example, Perez and Reyes, 2006). The main reason for this is its ability to approximate virtually any nonlinear function (Slini et al., 2006). It can therefore be seen as the benchmark statistical approach in the air quality forecasting literature.

In order to describe NN more formally let us denote by y_t the pollutant to be forecasted using the observed sample $t = 1, \ldots, T$. The one-step ahead value of this pollutant is denoted by y_{T+1}. The goal is to construct a point forecast f_{T+1} of y_{T+1} given $\mathbf{x}_{T+1}$, a set of regressors believed to be able to predict y_{T+1}. In case of NN, this means setting

$$f_{T+1} = \mathbf{x}'_{T+1}\beta + \sum_{j=1}^{J} \alpha_j G(\mathbf{x}'_{T+1}\beta_j), \tag{1}$$

where G is the so-called "transfer" function, and β, α_j and β_j are coefficients of the model. In the usual NN terminology, the network is said to comprise three different "layers". At the basis is the "input layer", which is simply the regressors in x_t, which are usually are called "inputs". These inputs are multiplied by so-called "connection strengths" β_j as they enter the "hidden layer", which consists of J "hidden units", which are the logistic functions. Finally, the hidden units are multiplied by coefficients α_j to produce the "output" f_t.

Of course, since G, β, α_j and β_j are all unknown, f_{T+1} is unavailable and we will therefore consider replacing it by an estimate, $\widehat{f}_{T+1}$ say. The idea is to, for a given choice of G, use the observed sample on (y_t, $\mathbf{x}_t$) to obtain estimates $\widehat{\beta}$, $\widehat{\alpha}_j$ and $\widehat{\beta}_j$ of β, α_j and β_j, respectively, which can be accomplished using nonlinear least squares (NLS), and then to make a one-step-ahead forecast from the resulting estimated model. However, for this to be possible there are a number of important choices that have to be made. The most obvious choice is that of G. While a wide range of transfer functions has been considered, the logistic function $G(x) = 1/(1 + e^{-x})$ is perhaps the most popular.[5] The idea is that by allowing the number of logistic components, here denoted J, to increase one can approximate any nonlinear relationship that might exist between y_t and $\mathbf{x}_t$. However, while the network can be made arbitrarily flexible by setting J large enough, this also increases the risk of in-sample overfitting. Another problem is that the NLS objective function is known to possess many local minima (see, for example, Nunnari et al., 2004; Franses and van Dijk, 2000, Chapter 5). Thus, even if the estimation algorithm converges, there are no guarantees that it will be to the global minimum.

Because of these problems it is common to also consider simpler models. A very common choice is LR, which is (1) with $\alpha_1, \ldots, \alpha_J$ put to zero. Thus, in case of LR, one sets

$$f_{T+1} = \mathbf{x}'_{T+1}\beta, \tag{2}$$

which reduces significantly the complexity of the estimation problem. Indeed, since the model is now linear, the estimator $\widehat{\beta}$ of β can be obtained from a simple ordinary least squares (OLS) regression of y_t onto $\mathbf{x}_t$. The resulting estimated forecast is given by $\widehat{f}_{T+1} = \mathbf{x}'_{T+1}\widehat{\beta}$.

Despite its simplicity LR has been found to perform quite well, even in comparison to more general models (see Perez and Reyes, 2006; Slini et al., 2006). One reason for this is parsimony. Indeed, it is a well-known fact that estimating additional parameters can raise the forecast error variance above what might be obtained when using relatively simpler models. Thus, while excluding variables whose parameters are nonzero can adversely affect forecasting accuracy, adding them might lead to an increase in the forecast error variance. This trade-off suggests that combining

[2] Population estimated for 2010 using the 2005 Census (DANE, 2006).

[3] The vehicle fleet contributes 1100 tons of fine particulate matter, 30,000 tons of nitrogen oxides, 450,000 tons of CO and 60,000 tons of hydrocarbons (see Uniandes, Secretaria de Ambiente, 2009).

[4] For example, in 2009 on no less than 190 occasions did PM_{10} surpass the daily norm of 150 $\mu g/m^3$ and the hourly O_3 standard of 61 ppb was exceeded even more often, 210 times (see Secretaria de Ambiente, 2010, for a recent account of air quality). Resolution 601 of 2006 issued by the Ministry of Environment provides a complete description of the AQS for several pollutants.

[5] Thus, with this choice $G(\mathbf{x}'_t\beta_j)$ is nothing but the familiar binary logit probability model.