Transfer-Function Models Predicting Ozone in Urban Air

A contribution to subproject SATURN

U. Schlink, S. John, O. Herbarth

UFZ-Centre for Environmental Research Leipzig-Halle Ltd., POBox 500135, D-04301 Leipzig, Germany

Summary

This contribution addresses the inclusion of atmospheric variables into time-series models for short-term predictions of urban air pollution and their testing in different European urban areas. In particular we worked on the problem of identifying transfer-function models for data series of Ozone from two European regions, one continental site in the urban area of Berlin (see Fig. 1), Germany, and one mediterranean site in Melilli at Sicilly, Italy.

Aim of the research

In the past we mainly worked with univariate time-series models for the prediction of next day's air pollution. It was shown, that the strength of univariate models is to predict pollutants that show more or less regular behaviour (Schlink and Volta, 2000). However, forecasting irregular variation that occurs, for example, at the beginning of a smog episode, appears to be difficult. Apparent irregular dynamics of time series may have three causes: (i) the effect of external covariates, i.e., meteorological parameters and other pollutants (e.g. precursor substances), (ii) the existence of a non-linear dynamics in the observed time-series, or (iii) stochastic noise.

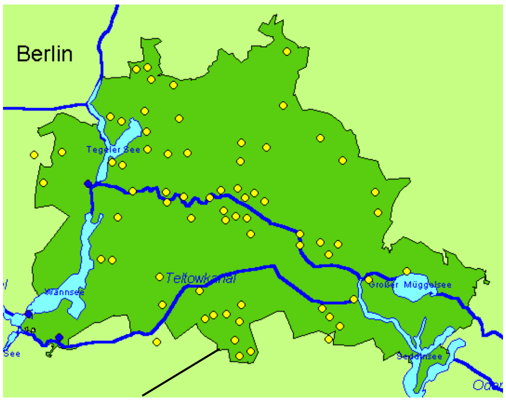

Figure 1. Location of the Berlin site for which Ozone concentration was modelled and predicted.

In previous work we tested Ozone data for dynamical non-linearity. In contrast to recent reports in the literature, we detected only very weak dynamical non-linearity (Haase and Schlink, 2001; Haase, Schlink and Richter, 2001; Schlink and Haase, 2001) and this finding was confirmed by the results of other authors (Palus et al., 2001). As a consequence, to refine statistical techniques for air pollution forecasting, we now studied the effect of external variables.

Activities during the year

The work was based on the identification of transfer-function models, the estimation of their parameters and the calculation of predictions which are assessed by several measures of forecasting quality. Finally, comparison was made to (univariate) predictions based on autoregressive models.

Data analysis and model identification

Modelling Ozone data we used daily maxima of 1hr and 8hr mean concentrations. This decision was guided by the existence of threshold values for 1hr and 8hr means of Ozone in the European Community directives 92/72/EEC and 96/62/EC. These threshold values were used in the final step to assess the quality of exceedence forecasting.

As Ozone concentrations can be considered as approximately log-normally distributed, all statistical methods have been applied on the logarithms of the concentration values. Back-transforming the data after prediction ensures that forecasted Ozone concentrations are always positive.

All modelling work was based on the ARX[r,s] model $d_r(B)Y_t = w_s(B)X_t + h_t$ with the Ozone value Y_t and the external variable X_t (Kendall and Ord, 1990). More complicated models for the random component h_t did not result in an improved prediction performance.

We first studied the external variable 'Temperature'. In a second attempt we modelled Ozone in dependence on Nitrogen Oxides (NO_x). Finally a model $Y_t = d_1 Y_{t-1} + w_{10} X_{1t} + w_{20} X_{2t} + h_t$, comprising both external variables X_{1t} and X_{2t} was identified (multiple input model).

Prewhitening

Transfer-function modelling requires a pre-processing of data, which is called prewhitening. This consists in fitting an ARMA model to the time-series of input data and filtering input and output by means of this model. This ensures a straightforward calculation of the impulse-response function (IRF) from the cross-correlation function (CCF) (see Fig. 2).

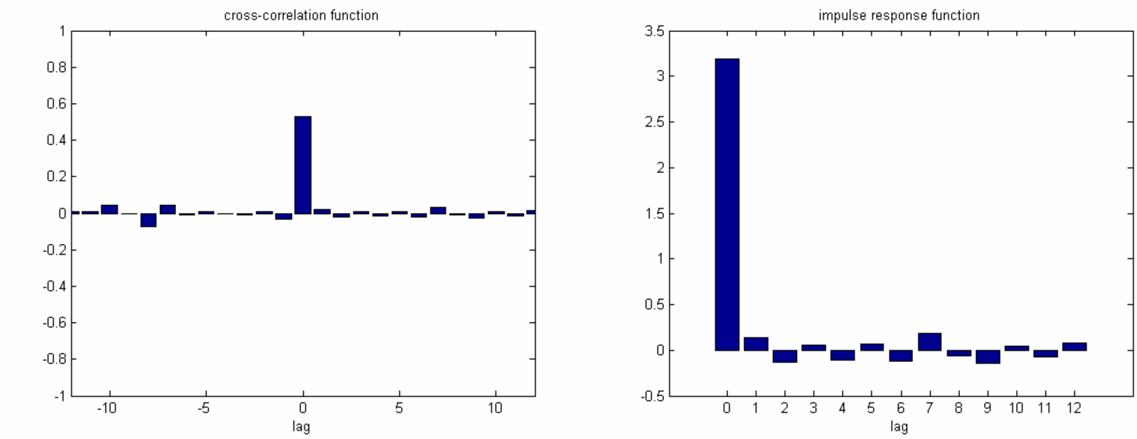

Figure 2. CCF (left) and IRF (right hand side) for Ozone in Berlin (2000) with Temperature as input. In both diagrams the abscissae represent the lag between Ozone and Temperature.

Prediction of the input data

Forecasting tomorrow's Ozone concentration by means of the ARX model requires values of the external covariate (Temperature or Nitrogen oxide concentration) for tomorrow. For that purpose we can utilise either tomorrow observed values of the external covariate, or we can calculate univariate predictions of the covariate and take these predictions as input for the ARX model. Of course, only the latter method is really a prediction.

Principal results

The order of the identified transfer-function models was in general $(r,s) = (1,1)$, with an exception for Melilli and NOx input where the order was $(r,s) = (2,1)$. Predictions have been calculated for a time horizon of one and two days and Figure 3 illustrates the multiple input forecasts for Berlin.

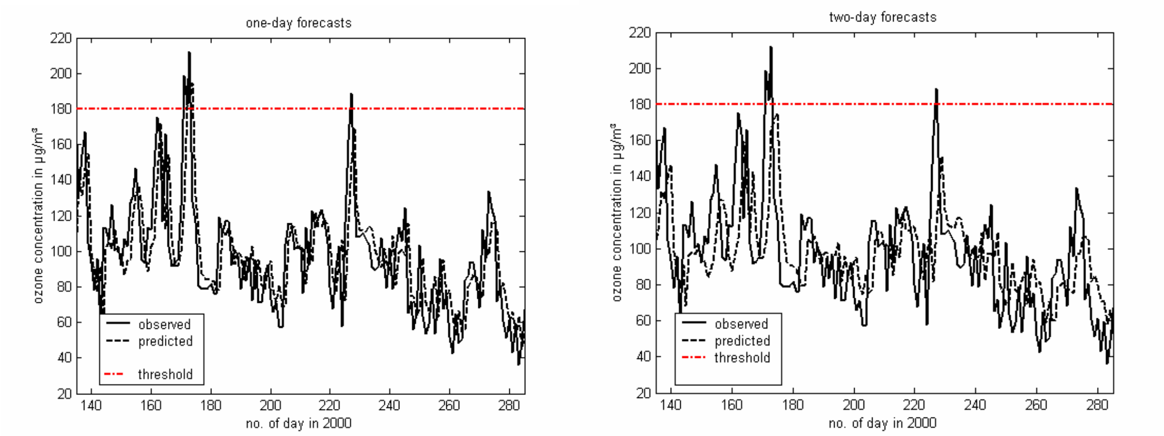

Figure 3. Daily maxima of 1hr Ozone observations in Berlin; day No. 1 is the 1st January 2000. The dashed line represents the predictions of the transfer-function model based on Temperature and NOx concentration as input. The horizontal line is the threshold that defines the exceedences.

All predictions P_i were compared with the observations O_i, and the quality of forecasts was assessed using following indices: Mean Bias Error: $MBE = \frac{1}{n}\sum_{i=1}^{n} P_i - O_i$, Mean Absolute Error: $MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right|$, Root Mean Square Error: $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$, Index of Agreement: $d = 1 - \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n}\left(\left|P_i - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2}$, and the Squared Correlation Coefficient R^2. The comparative Table 1 presents the results for one-day forecasting and Table 2 for two-day forecasting. Considering the different models in Tables 1 and 2, one should bear in mind that only the relative measures d and R^2 are useful for comparison purposes. Additionally, the prediction performance of two autoregressive models is presented in Table 3.

The application of the transfer-function models to forecast Ozone demonstrates, that the Temperature is a much more important input variable than NOx concentration. There is only minor improvement in forecasting performance using both external variables in a multiple input model.

Based on the Berlin data we found it difficult to assess the quality of exceedence forecasting as there are just 4 exceedences of the threshold in 2000. For Melilli, which is characterised by higher Ozone pollution, one-day forecasts predicted 20 of 61 observed exceedences, and 19 predictions were false alarms.

Main conclusions

We find that the multivariate transfer-function approach results in an improved quality of the Ozone predictions, compared to AR(1) forecasts. Interestingly, the multivariate approach is not much superior to the higher-order autoregressive approach (AR(10))! As the fluctuations of the data series for Melilli are much higher than for Berlin, Ozone forecasts are of poor quality for the former.

While the multivariate approach per se does not guarantee a higher-quality of predictions we observed that using tomorrow observed meteorological variables, like Temperature, results in considerable improvement in prediction accuracy (For this case the performance measures are not presented here). Of course, these variables are not available at the time when the prediction is calculated. Therefore, we just can recommend to use the Temperature value that is predicted for tomorrow by the meteorological forecast. Since the meteo forecast can not be expected to be absolutely correct, this approach will be limited as well.

Table 1: Prediction performance for one-day forecasts of the transfer-function model.

Site	Berlin			Melilli		
Input	Temp.	NOx	Temp + NOx	Temp.	NOx	Temp + NOx
MBE	-2.43	-3.45	-2.04	-16.97	-15.32	-16.08
MAE	15.89	15.73	15.76	39.67	40.98	39.65
RMSE	20.18	20.72	19.97	89.06	91.02	88.77
d	0.91	0.88	0.91	0.553	0.548	0.56
R^2	0.69	0.67	0.69	0.26	0.21	0.26

Table 2: Prediction performance for two-day forecasts of the transfer-function model.

Site	Berlin			Melilli		
Input	Temp.	NOx	Temp + NOx	Temp.	NOx	Temp + NOx
MBE	-3.67	-5.75	-3.23	-21.09	-19.01	-18.89
MAE	19.46	19.04	19.24	42.62	44.50	42.30
RMSE	24.56	25.27	24.24	91.88	95.95	91.48
d	0.84	0.79	0.85	0.47	0.46	0.49
R^2	0.54	0.54	0.55	0.24	0.14	0.22

Table 3: Prediction performance for one- and two-day forecasts of autoregressive models AR(1) and AR(10).

Site	Berlin				Melilli			
Model	AR(1)		AR(10)		AR(1)		AR(10)	
Horizon	1-d F	2-d F	1-d F	2-d F	1-d F	2-d F	1-d F	2-d F
MBE	-3.79	-6.23	-2.93	-4.14	-17.02	-24.67	-13.09	-14.49
MAE	15.85	19.15	15.13	17.72	41.80	46.24	38.83	42.18
RMSE	20.82	25.43	19.57	22.63	92.60	97.22	87.74	90.45
d	0.88	0.79	0.90	0.85	0.55	0.40	0.58	0.52
R^2	0.67	0.53	0.70	0.61	0.19	0.14	0.27	0.22

Aims for the coming year

Future research will address novel approaches to quantifying the effects of Ozone and other urban air pollutants as well as meteorological parameters on human health.

References

Haase P, Schlink U (2001) Non-parametric short-term prediction of ozone concentration in Berlin, paper presented at the Air Pollution Modelling and Simulation Conference, Paris 2001.

Haase P, Schlink U, Richter M (2001) Critical reconsideration of phase space embedding and non-parametric prediction of ozone time-series, paper presented at the Urban Air Quality Conference 2001, Loutraki, Greece.

Kendall M, Ord JK (1990) Time Series; Edward Arnold.

Palus M, Pelikan E, Eben K, Krejcir P, Jurus P (2001) Nonlinearity and prediction of air pollution, Artificial Neural Nets and Genetic Algorithms, Proceedings of the ICANNGA 2001 Conference, 473-476.

Schlink U, Haase P (2001) Application of the surrogate test to detect dynamic non-linearity in ground-level ozone time-series from Berlin, Artificial Neural Nets and Genetic Algorithms, Proceedings of the ICANNGA 2001 Conference, 469-472.

Schlink U, Volta M (2000) Grey box and component models to forecast ozone episodes: a comparison study, Environmental Monitoring and Assessment 65(1/2), 313-321.