

A Ground-level Ozone Forecasting Model for Santiago, Chile

HÉCTOR JORQUERA,^{1*} WILFREDO PALMA¹ AND JOSÉ TAPIA²

¹ *Pontificia Universidad Católica de Chile, Chile*

² *Universidad de Valparaíso, Chile*

ABSTRACT

A physically based model for ground-level ozone forecasting is evaluated for Santiago, Chile. The model predicts the daily peak ozone concentration, with the daily rise of air temperature as input variable; weekends and rainy days appear as interventions. This model was used to analyse historical data, using the Linear Transfer Function/Finite Impulse Response (LTF/FIR) formalism; the Simultaneous Transfer Function (STF) method was used to analyse several monitoring stations together. Model evaluation showed a good forecasting performance across stations—for low and high ozone impacts—with power of detection (POD) values between 70 and 100%, Heidke's Skill Scores between 40% and 70% and low false alarm rates (FAR). The model consistently outperforms a pure persistence forecast. Model performance was not sensitive to different implementation options. The model performance degrades for two- and three-days ahead forecast, but is still acceptable for the purpose of developing an environmental warning system at Santiago. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS ground-level ozone forecast; forecast evaluation; FIR model; LTF model; STF model

INTRODUCTION

Tropospheric ozone is a distinctive signature of today's large cities. This pollutant is associated with different health effects, such as increases in mortality, respiratory and asthma symptoms, days of restricted activity and eye irritation. The short-term effects are serious enough to prompt public authorities to require the release of information to the public or to take contingency measures whenever ground-level ozone reaches high levels (Elsom, 1995). For instance, the European Economic Community has issued a Directive requiring that information be broadcast if the hourly ozone concentration exceeds 180 ($\mu\text{g}/\text{m}^3$), and that a warning should be sent to the population whenever ozone levels exceed 360 ($\mu\text{g}/\text{m}^3$).

* Correspondence to: Héctor Jorquera, Department de Ingenieria Química, Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago 6904411, Chile. E-mail: jorquera@ing.puc.cl
Contract/grant sponsor: Conicyt-Chile; Contract/grant numbers: Fondecyt 198-0972; Fondecyt 80000004.

The ozone problem is very difficult to manage because of its linking to economic growth: the largest impacts are associated with transportation and large industrial sources; thus it may take several years before public policies effect improvements upon ozone levels. In the meantime, it would be useful for local authorities to have a tool to forecast what the ozone would likely be tomorrow, so that they can take preventive measures as needed. This problem also poses statistical challenges to obtain good forecasting models (see, for example, Beck, 1991; Young *et al.*, 1991).

Ozone generation in the troposphere

Ozone is not emitted by any significant anthropogenic source. It is a secondary pollutant, that is, one produced by chemical reactions upon primary emitted pollutants such as nitrogen oxides ($\text{NO}_x = \text{NO} + \text{NO}_2$) and volatile organic compounds (VOC). These primary pollutants (also called ozone precursors) are emitted by industrial, commercial, residential and mobile sources; the nitrogen oxides are mostly the product of fuel combustion, whereas VOC emissions come from transportation, use of solvents, fuel distribution, natural sources (trees and plants), etc. As the economy grows, emissions tend to be dominated by transportation sources, while the rest of the sources (industrial, commercial and residential) lower their emissions by shifting to cleaner fuels. This pattern of emissions makes the management of urban air quality a challenging task, especially in developing countries (Elsom, 1995; Fenger, 1999).

Ozone precursors are intertwined in a complex network of chemical reactions triggered by solar radiation, so ozone concentration rises during the day, declines at sunset and is low overnight (Seinfeld and Pandis, 1998). There is a complex relationship among daily maximum ozone impacts and the emissions of NO_x and VOC within a given area. Thus it has been found both by simulation and by effected emission reductions that ozone tends to change (in relative terms) less than the NO_x and VOC (relative) emission reductions (Meng *et al.*, 1997). State-of-the-science, photochemical models are capable of simulating the different processes that determine the fate of tropospheric ozone (Zannetti, 1990). However, these models require extensive resources for developing NO_x and VOC emission inventories (with spatial and temporal resolution), monitoring campaigns, meteorological databases, model validation and so on (Hanna *et al.*, 2000). Thus, while these deterministic tools are becoming increasingly available in many cities worldwide, those capable of producing real-time ozone forecasts are available only in a few locations.

Survey of empirical ozone forecasting models

Empirical ozone forecasting models have been developed in many cities worldwide. These models have been usually derived from a correlation analysis between ozone, meteorological parameters and other air pollutants. The forecasting models have used tools such as linear time series (Robeson and Steyn, 1990; Gonzalez-Manteiga *et al.*, 1993), linear and non-linear regression (Ryan, 1995; Hubbard and Cobourn, 1998), discriminant analysis, cluster analysis (Noordijk, 1994), generalized additive models (Niu, 1996; Davis and Speckman, 1999), artificial neural networks (Cannon and Lord, 2000; Comrie, 1997; Gardner and Dorling, 1998; Kao and Huang, 2000; Melas *et al.*, 2000; Yi and Prybutok, 1996), and fuzzy models (Jorquera *et al.*, 1998a). Time series models were used extensively in the 1970s and 1980s; most of them were univariate models, and usually their performance was not good enough to develop environmental warning systems (Myrabo *et al.*, 1976; Aron and Aron, 1978; McCollister and Wilson, 1975; Pryor *et al.*, 1981, Simpson and Leyton, 1983). There was an early attempt to apply the Kalman filter methodology (Desalu *et al.*, 1974), but it was not tested with real data nor was it used for air pollution forecasts.

In the early 1990s, Young *et al.* (1991) introduced a comprehensive framework for analysing time series of environmental data. They advocate an approach in which seasonal and trend components are decomposed into quasi-orthogonal components (the so-called 'unobserved component time series model') by applying forward recursive Kalman filters and the backward recursive fixed interval smoothing (FIS) algorithm to estimate the joint trend and seasonal models. The estimation of model parameters and series decomposition is achieved by the recursive, FIS algorithm based on the state space model for the whole system. Young *et al.* (1997) and Young (1998) extended this unobserved component methodology to include transfer function components and the resulting model has similarities to the model utilized later in the present paper. Schlink *et al.* (1997) and Ng and Yan (1998) have reported applications of this methodology to air quality forecasts.

Summarizing the literature on ozone forecasting, we can conclude that there have been few studies focused on fair comparisons among forecasting methods, especially when real-world data are analysed. Moreover, the lack of publicly available databases hampers efforts oriented towards a systematic comparison among different methodologies.

Case study: Santiago, Chile

Topography, climate and economic growth have turned Santiago, Chile, into one of the most heavily polluted cities in South America. Santiago (current population, 5.4 million) is located in a geographically confined basin between the Andes mountains to the east, with an average altitude of 4500 m, a parallel coastal range to the west with an average altitude near 1500 m, and the basin is further bounded to the north and south by transversal mountain chains. The urban area spreads on a gentle slope from 450 m in the west to 800 m to the east (see Figure 1). At this

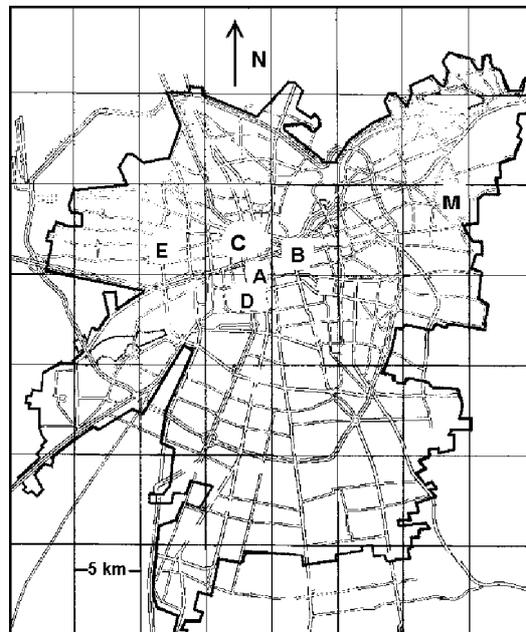


Figure 1. Map of Santiago, Chile, showing the monitoring stations of the MACAM air quality and meteorological network. The thick lines show the urban border, and the thinner lines the major road network

latitude (33.5°S) the radiative transfer controls vertical mixing of the air with frequent low thermal inversions reinforced by the cooling of low-level air in contact with the Pacific Ocean (65 miles to the west). The climate is semi-arid with an average annual rainfall of around 350 mm (Rutllant and Garreaud, 1995), although there is considerable inter-annual variability associated with the El Niño and La Niña phenomena.

Santiago has highly segregated land use; there is a south-west industrial area surrounded by low-income settlements, an old downtown area including many shops and service companies, and a residential eastern zone. The latest urban developments have included expansion of industrial activities on the northern side and rapid housing development in the south-east. Several million trips are made every working day, with many passing by the downtown zone.

The prevailing winds in summer, and springtime conditions (a dry ozone season) are south-westerly. Hence air parcels travelling from the industrial south-west collect ozone precursor emissions and are then subject to strong radiation and high temperatures so ozone rises. By the time an air parcel hits downtown (about noon), some stations record violations of the current hourly standard of 82 ppb, and the highest impacts are measured at the north-eastern side of town between noon and 2 pm (Jorquera *et al.*, 1998b).

THE GROUND-LEVEL OZONE FORECASTING MODEL

The first attempts to construct an ozone forecasting model for Santiago were carried out using artificial neural networks, with data from a downtown station (Acuña *et al.*, 1996). Then in Jorquera *et al.* (1998a) three different modelling techniques were used: linear time series, artificial neural networks and fuzzy models, with data from a suburban, highly impacted area. Finally, the model presented here was derived by means of a systematic simplification of the conservation equation for an air parcel being advected by the wind across an urban area. Hence it should be applicable to other cities as well. More details of the model derivation are given in Jorquera *et al.* (2000); here we quote just the final result:

$$O_3(t) = K + R \cdot \Delta T(t) + \sum_{i=0}^N \omega_i L^i I_1(t) + N_1(t)$$

$$\phi_1(L)N_1(t) = \theta_1(L)\varepsilon_1(t) \quad (1)$$

$$\Delta T(t) = \Delta T_0 + \frac{\omega_p}{(1 - \delta_p L)} I_2(t) + N_2(t)$$

$$\phi_2(L)N_2(t) = \theta_2(L)\varepsilon_2(t) \quad (2)$$

where K is a constant term that represents early-morning contributions to ozone from sources upwind of the city; R is a reduced-ozone production rate and takes into account the different mechanisms for generation, transformation and removal of ozone; ΔT is the daily rise in air temperature ($T_2 - T_1$), with T_1 and T_2 being the lowest and highest air temperatures recorded that day, respectively. In equation (2), an empirical model was developed for the daily rise in air temperature (ΔT), so that the available meteorological information was also included in the forecast of ΔT ; in equation (2) ΔT_0 stands for the average daily temperature rise in the ozone season (see further comments below). In both equations L stands for the lag operator defined as $LX_t = X_{t-1}$ for any discrete signal X_t .

The models in equations (1) and (2) have a linear transfer function (LTF) structure (Pankratz, 1991), sometimes referred to as the finite impulse response (FIR) model, where the following interventions have been added:

- (a) The weekend emission reduction effect upon ozone in equation (1) was modelled as being a binary signal defined by

$$\begin{aligned} I_1(t) &= 0 && \text{if } t \text{ is a working day} \\ &= 1 && \text{if } t \text{ is a weekend day} \end{aligned} \quad (3)$$

- (b) The rainy-day signal effecting changes on ΔT in equation (2) was also modelled as a binary signal defined by

$$\begin{aligned} I_2(t) &= 0 && \text{if no rainfall was recorded on day } t \\ &= 1 && \text{if any rainfall was recorded on day } t \end{aligned} \quad (4)$$

A finite series of powers in L was found to be a good polynomial approximation to the transfer function associated with the input intervention signal $I_1(t)$; the transfer function associated with the rainy day input signal $I_2(t)$ was better modelled as an exponentially damped response; $N_1(t)$ and $N_2(t)$ are underlying disturbances that are modelled as ARMA processes and $\varepsilon_1(t)$ and $\varepsilon_2(t)$ are white noises associated with the LTF models (1) and (2), respectively. Because both series are truncated expansions in powers of L , we will refer to those expressions either as LTF or FIR models of the physical phenomenon analysed.

Rationale for the intervention model

The weekend emissions change is an intervention that has been observed to significantly change ozone concentrations at urban areas (Chaum *et al.*, 1978; Cleveland *et al.*, 1974; Cleveland and McRae, 1978; Karl, 1978; Levitt and Chock, 1976; Pryor and Steyn, 1995). Those studies were aimed at assessing the effect of typical weekend reductions of precursor emissions upon the highest ozone impacts and consequently to evaluate the effectiveness of short-term ozone abatement strategies. Typical conclusions of those analyses were that:

- (1) At downtown sites, ozone on weekends is higher than on workdays, because there are fewer NO emissions to scavenge ozone (by way of the fast reaction $\text{NO} + \text{O}_3 = \text{NO}_2 + \text{O}_2$).
- (2) At suburban sites the weekend ozone impacts may or may not change significantly, and this depends on the spatial patterns of the precursor emissions. Saturday ozone peaks tended to show similar values to those of weekdays, and Sundays tended to be the days on which ozone maxima showed the highest differences compared to workday levels; this means that it takes one or two days for urban photochemistry to respond to weekend reductions.

Therefore, weekends behave differently and so this has to be taken into account within the model. Intervention analysis provides a suitable framework for including these kind of discrete events, within the LTF/FIR formalism (Pankratz, 1991).

In the LTF/FIR formalism, we need to have a forecast of the input variables in the ozone-forecasting equation (1) before we can produce ozone forecasts. In this case, ΔT is forecasted using model (2) and then used as input for the ozone forecast. This procedure is carried out automatically in the software environment by defining the models for ozone maximum and for

ΔT ; in this way, the system knows that it is dealing with a stochastic input variable (ΔT) and not a purely deterministic input (like the binary intervention signal $I_1(t)$). This is the rationale for not combining equations (1) and (2) into a single expression; moreover, since ΔT is an observed quantity it is appropriate to keep it as such in the model structure because it brings more information to the model-identification process.

In the derivation of model (2), we first tried simple ARIMA models for forecasting ΔT , but the results were poor because the univariate series for ΔT cannot anticipate by itself weather changes, so we need a leading meteorological input variable to improve the forecast of ΔT . From an empirical standpoint, whenever a low-pressure front passes by and brings rainfall, the solar radiation and the daily temperature rise (ΔT) are drastically reduced compared to its seasonal average ΔT_0 . A simple way of handling this meteorological effect was to take the daily precipitation records and construct a binary intervention signal as defined by equation (4). A more refined procedure would call for including additional variables such as the cloud cover that indicates a smoother transition between clear skies and rainy days; however, we did not have access to this information, so we cannot improve further on this. Likewise, we would expect higher incoming radiation (and so higher ΔT) to be correlated with mixing height; unfortunately this parameter has not been measured at Santiago so we cannot include it in the ozone model. Clearly, this lack of upper air measurements is a drawback of the proposed model, but the good performance of the forecast (see 'Evaluation of model forecasting' section below) suggests that this limitation does not seem to be a serious problem, at least for Santiago.

Other empirical studies agree with the basic model structure given by equation (1). Ryan (1995) has found that temperature explains most of the variance in ground-level ozone in the Baltimore metropolitan area, and that lower morning temperatures were also negatively correlated with high ozone impacts; the same findings were reported by Hubbard and Cobourn (1998) for the Louisville, Kentucky metropolitan area. Comrie (1997) performed statistical analyses of several cities in the USA and found that daily maximum temperature and daily total sunshine were significant predictands for ozone maxima in all cases; daily total sunshine is a surrogate for UV flux, and it should be strongly correlated with $(T_2 - T_1)$ in equation (1).

Databases used

Table I shows the data sets used for model identification and forecasting. All data were taken from the air-quality monitoring network (MACAM in Spanish) that is operated by the Ministry of Public Health. The data consisted of hourly records of ozone, temperature, etc. at the five monitoring sites (A, B, and C, D, M) shown in Figure 1. The sites cover mostly the downtown area, but the M monitor is located in a north-eastern, suburban area and it has recorded the largest ozone impacts at Santiago; this behaviour is due to that monitor's location in the city: downwind of main sources of ozone precursors (recall that the predominant wind direction is south-west). The daily precipitation data were taken from the meteorological station (E in Figure 1). We chose the 1994–1996 ozone seasons because the earlier records contained too many missing values to be used effectively. From

Table I. Data sets for model identification and forecasting

Set (stations)	Period (dd/mm/yy)	No. identification days	No. evaluation days
1 (A,C,D,M)	1/9/94–28/4/95	190	50
2 (A,C,D,M)	1/9/95–30/4/96	190	53
3 (C,D,M)	1/9/96–31/12/96	90	32

the available monitoring sites, we discarded station B because it is located between two busy streets, so ozone is mostly scavenged by fresh NO emissions from vehicles, resulting in very low ozone readings. In the winter of 1996 station A was discontinued, so it is not included in data set 3.

MODEL IDENTIFICATION

All the model identification and forecasting was carried out using the SCA Statistical System, a proprietary software developed by Scientific Computing Associates Corp., Illinois, USA (Liu *et al.*, 1986). The specific version used was SCAWIN Release VI.2a-Professional, running in a Pentium-II PC. This software comes with an expert system to identify stochastic processes, especially those that have a strong seasonal component (Liu, 1989). The software provides an option for joint parameter estimation and outlier adjustment; this procedure is iterative and has been described by Chen and Liu (1993a). Two types of transfer function models can be handled:

- (a) The linear transfer function (LTF) model in which equation (1) is applied to each monitoring station's data separately.
- (b) The simultaneous transfer function method (STF) where models from several stations are jointly estimated; in this case four models like (1) are identified simultaneously using the STF formalism.

Regarding model (2) for the daily temperature rise, it is identified independently using the LTF formalism. All model-identification methods used in this work are described below.

Methodology to identify LTF and STF models

- (1) The identification of the transfer function (TF) is made following the so-called linear transfer function (LTF) method proposed by Liu and Hanssens (1982) and detailed in Liu *et al.* (1986). In this approach, the ratio of polynomials in the lag operator, $\theta(L)/\phi(L)$, is approximated by a finite polynomial of L with degree k . If k is sufficiently large, then the approximation of the ratio will be reasonably good. After determining the length of the truncation, the approximating polynomial is used to obtain the parameter estimates.
- (2) The simultaneous transfer function (STF) system is identified by means of the stepwise autoregressive fitting. In this method, the possible presence of contemporaneous relationships among series is detected by analysing the residual correlation matrices after fitting an autoregressive vector time series of increasing order; cf. Liu (1991).
- (3) The joint estimation and outlier detection process is carried out by following the methodology proposed by Chen and Liu (1993a). This technique consists of three steps: (a) outlier detection, using multiple regression to take into account possible multivariate effects of outliers in other series; (b) outlier adjustment; and (c) parameter estimation based on the adjusted series. After step (c), the final estimates of the model are performed using maximum likelihood. Then the final model can be used to forecast, following the methodology described by Chen and Liu (1993b).

In addition to the above modelling options, we can also choose whether to update the estimated parameters with all the available information up to current time index, or to keep fixed the parameters in the values estimated with the initial calibration subset (see Table I for the number of data points used in this study); we refer to these modelling options as PR or FP, respectively.

Therefore, we have used the SCAWIN software environment to select different modelling options. In this way we can test the effectiveness of implementation (Adya and Collopy, 1998). The options we have selected are the following:

- (a) Use the LTF or the STF formalism (LTF or STF, respectively), that is, choose between methods (1) and (2) above.
- (b) Use the standard estimation or the joint parameter estimation and outlier adjustment method (F or OF, respectively). Here we use the standard, maximum likelihood estimation (F option) or the more elaborated, iterative outlier adjustment and estimation process detailed in method (3) above (OF option).
- (c) Use fixed parameters or re-estimate parameters after each time step (FP or PR options).

More details on this aspect are given in the 'Evaluation of model forecasting' section.

Results of the identification process

Initial checks were performed upon equation (1) to validate the model structure predicted by the physical derivation. Thus application of the cross-correlation formalism (CCF) indeed showed that the relationship between daily maximum ozone and air temperature rise was a contemporaneous one, and no significant lags were found. The noise process was usually found to be an AR(1) process, with an autoregressive coefficient significantly different from 1.0, a strong indication that the relationship described by equation (1) is a stationary one, and so no data differentiation was performed at all (Jorquera *et al.*, 2000).

Tables II and III show the results of the model identification and diagnostics of equations (1) and (2), respectively, for all data sets and monitoring stations analysed. These results are discussed below.

Results for the ozone-forecasting equation (1)

First, the R parameter stands for a rate of ozone production per unit heating of the airshed, at a given site. The values are all significant and positive, and they are ranked according to the magnitudes of the observed ozone impacts at Santiago. Hence, station M (at which the highest ozone concentrations have been recorded) has the largest values for R , whereas the lowest estimates of R are for station A, at downtown. Recall that station M is located in the north-east side of Santiago and the prevailing winds are south-westerly, so its location is the most downwind of all. Thus ozone impacts there reflect a greater distance travelled by air parcels, longer residence times and consequently higher ozone impacts than those recorded at the Downtown stations A, C and D (closer to transportation sources of NO_x that lower ozone impacts by way of the fast reaction $\text{NO} + \text{O}_3 = \text{O}_2 + \text{NO}_2$). These results show that the R parameter takes into account the relative positions of the monitoring stations with respect to the major emission sources of ozone precursors in the city.

Regarding the weekend effect on reduction of ozone precursor emissions, there is a different behaviour, both across stations and also during the period analysed. For the stations closer to downtown (A, C and D) there is a rise in ozone in Saturdays (linked to decreasing transportation emissions of NO_x), followed by a small net effect in Sundays and then a drop on Mondays. Hence, it seems that it takes about two days for the atmosphere to respond to the weekend reduction of emissions, in terms of photochemical activity. For the case of the suburban station M, the behaviour is different, because the significant changes in ozone levels occur on Sundays and Mondays, with a significant decrease in ozone impacts. These findings are clear for data sets 1 and 2, but not for data set 3. At present we ascribe this behaviour to a shift in spatial and temporal emission

Table II. Model identification for equation (1) LTF method

Parameter	A	C	D	M
Data set 1				
K [ppb]	-9.45 (-3.19)	-16.90 (-4.00)	-7.53 (-2.28)	
R [ppb/°C]	3.587 (19.64)	5.339 (20.58)	3.707 (18.28)	6.187 (60.73)
ω_{10}	6.336 (3.69)	6.339 (2.72)	4.562 (2.50)	
ω_{11}	-4.649 (-2.67)	-4.621 (-1.96)	-4.541 (-2.46)	-7.812 (-3.08)
ϕ_1		0.214 (2.95)	0.222 (3.10)	0.207 (2.85)
θ_1	-0.151 (-2.07)			
Data set 2				
K [ppb]				
R [ppb/°C]	3.860 (41.97)	3.446 (39.41)	3.667 (25.97)	6.205 (33.59)
ω_{10}	6.133 (2.79)			
ω_{11}	-5.898 (-2.70)		-7.000 (-3.73)	-14.469 (-5.58)
ϕ_1	0.646 (6.52)	0.282 (4.04)	0.873 (16.10)	
θ_1	0.539 (4.35)		0.670 (7.57)	0.653 (12.50)
Data set 3				
K [ppb]				
R [ppb/°C]	NA	3.328 (36.13)	2.902 (49.14)	5.074 (37.62)
ω_{10}	NA	7.189 (3.04)		
ω_{11}	NA			
ϕ_1	NA			
θ_1	NA	-0.427 (-4.37)		-0.430 (-4.59)

Notes: Values in parentheses are the t -values for the parameter estimates; blank entries mean that the parameter was not significant and so it was not included in the model; NA means that the station did not have records for that period. All parameters were estimated using the iterative, joint outlier detection and adjustment methodology.

Table III. Model identification for equation (2), LTF method

Parameter	Data set 1	Data set 2	Data set 3
ΔT_0 [°C]	15.76 (60.94)	16.73 (44.06)	16.81 (33.71)
ω_p [°C]	-8.19 (-10.65)	-6.45 (-7.02)	-6.40 (-3.70)
δ_p	0.486 (7.03)	0.428 (3.85)	0.456 (2.53)
ϕ_1		0.345 (4.96)	
θ_1	-0.341 (-4.85)		-0.242 (-2.26)
θ_3			0.251 (2.47)

Notes: Values in parentheses are the t -values for the parameter estimates; blank entries mean that the parameter was not significant and so it was not included in the model.

patterns within the city, promoted by the fast economic growth experienced during the 1990s at Santiago. For instance, many industries (traditionally located in the south and south-west) have been moved to the northern side of town, and there has been an explosive increase of housing units on the eastern side, leading to increased traffic flows on Saturdays associated with shopping and recreational activities.

The noise $N_1(t)$ associated with the LTF/FIR model (1) is modelled as an ARMA process, and the results are usually of the form of a low-order process such as AR(1), MA(1) or ARMA(1,1) structures. Values of the ARMA parameters are listed in Table II, and their numerical estimates

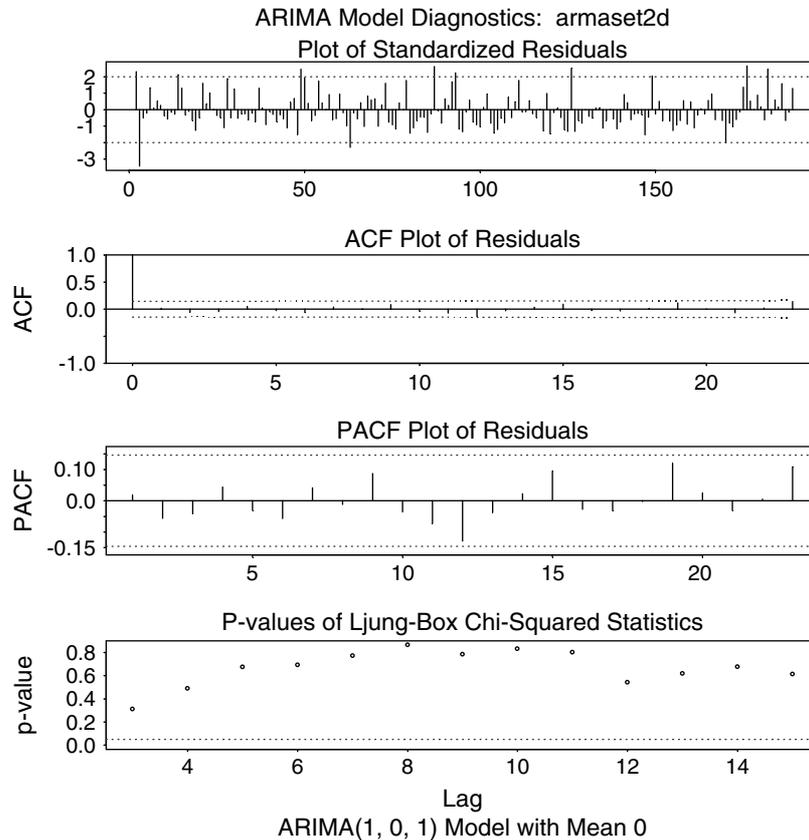


Figure 2. Diagnostic plots for data set 2, equation (1), MACAM station D

were also checked using S-Plus version 4.5. Diagnostics tests for residues show that the models have been successfully fitted, and Figure 2 gives a typical example for data set 2 and station D. Thus the ARMA models fitted to the disturbances $N_1(t)$ display white-noise residuals, showing that all the autocorrelation structure of the linear transfer disturbances has been appropriately described by the respective ARMA models. In particular, observe that the Ljung–Box test for whiteness (first plot from the bottom) is accepted for all lags at the 5% significance level (dotted line).

Results for the temperature rise equation (2)

The results listed in Table III show that the effect of a rainy day in Santiago is a reduction in ΔT of about 6–8°C on the same day, with respect to the seasonal average ΔT_0 . This means that on rainy days the temperature rise is 50–60% of the seasonal average, that is, a significant reduction. This effect decreases in time according to an exponential behaviour, being negligible after two to three days. This period is characteristic of synoptic scale motions in the atmosphere, so the model is consistent with the physics that drives air temperature dynamics.

The noise $N_2(t)$ associated with the LTF model (2) is also modelled as a low-order ARMA process, and the results are usually of the form of an AR(1) or MA(1) structure. The exception is

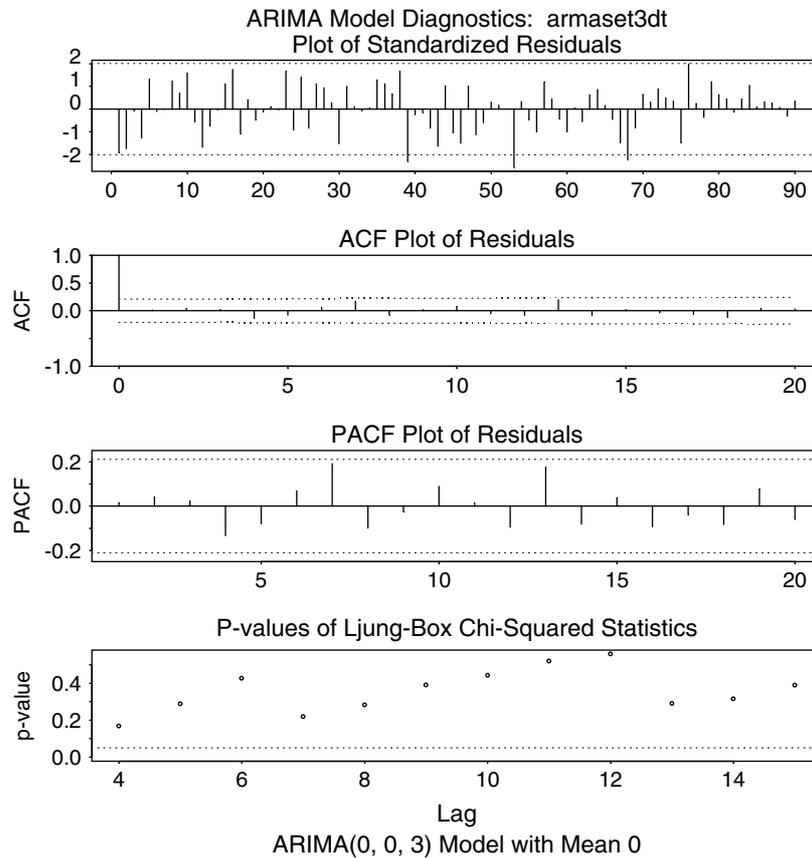


Figure 3. Diagnostic plots for equation (2), data set 3

data set 3, in which the moving average process $\theta_2(L)$ has a third-order term θ_3L^3 that is significant; this might be a consequence of the small size of the database. Figure 3 shows the diagnostics for data set 3, and the other data sets have similar diagnostic plots. All parameter estimations were checked running S-Plus version 4.5, confirming the numerical estimates given by the SCAWIN software.

Results for the STF modelling of equation (1)

The results from the LTF approach suggest that there are only minor differences among the model parameters and forecasting quality from the four stations under study. Moreover, it is possible that by taking into account the joint information from these stations we could be able to improve the performance of the models. Thus, we have chosen the simultaneous transfer function (STF) methodology to build up a multivariate system, which can include the correlation structure of the data. The idea is to test whether a multivariate model will improve upon forecasts by combining the information gathered at all the monitoring stations. These stations differ in their spatial location within the city, but the daily maxima of ground-level ozone are correlated among stations, because

Table IV. Model identification for equation (1), STF method

Parameter	A	C	D	M
Data set 1				
K [ppb]	-9.07 (-3.22)	-16.72 (-4.05)	-8.34 (-2.68)	
R [ppb/°C]	3.344 (18.94)	5.233 (20.43)	3.780 (19.63)	5.795 (50.95)
ω_{10}	5.773 (3.60)	6.147 (20.43)	4.313 (2.56)	
ω_{11}	-4.693 (-2.53)	-5.023 (-2.01)	-4.743 (-2.54)	-8.688 (-3.08)
ϕ_1		0.268 (5.45)	0.261 (5.17)	0.214 (3.46)
θ_1	-0.204 (-4-46)			
Data set 2				
K [ppb]				
R [ppb/°C]	3.654 (34.21)	3.486 (36.72)	3.522 (36.49)	6.220 (37.55)
ω_{10}				
ω_{11}			-2.769 (-2.12)	-9.813 (-4.01)
ϕ_1		0.305 (7.23)	0.450 (4.55)	0.502 (10.96)
θ_1	0.360 (3.95)			
Data set 3				
K [ppb]				
R [ppb/°C]	NA	3.373 (38.29)	3.090 (42.56)	5.071 (39.72)
ω_{10}	NA	5.974 (4.18)		
ω_{11}	NA			
ϕ_1	NA			
θ_1	NA			-0.324 (-4.16)

Notes: Values in parentheses are the t -values for the parameter estimates; blank entries mean that the parameter was not significant and so it was not included in the model; NA means that the station did not have records for that period.

these impacts are produced by the same physical mechanism, photochemistry of VOC and NO_x emissions modulated by identical mesoscale meteorological conditions.

The results of the STF model identification are shown in Table IV. The parameter estimates are quite similar to the values obtained by each model separately (see Table II), meaning that the model dependency among stations is small. The differences may be ascribed to the fact that Table II includes the more refined parameter estimates obtained by using the joint outlier detection and adjustment estimation method, whereas the STF methodology currently does not have such an option in the SCAWIN environment. Despite those small differences, we shall use the multivariate model for evaluating its forecasting performance and for comparing its results with those from other implementations of equation (1) for each station; this is described in the next section.

EVALUATION OF MODEL FORECASTING

We evaluated the proposed model given by equation (1) against a pure persistence forecast defined by

$$O_{3,t+1}^{PP} = O_{3,t} \quad (5)$$

This pure persistence forecast is a common benchmark in the air pollution forecasting literature. Usually ozone episodes are not isolated days, but rather several days of intense solar radiation,

Table V. Forecasting models used

Code	Description
PP	Pure persistence forecast, given by equation (5)
FFP	Forecast using equation (1), LTF model, fixed model parameters
FPR	Forecast using equation (1), LTF model, model parameters re-estimated at each time step
OFFP	Same as FFP, but with the option of joint parameter estimation and outlier adjustment
OFPR	Same as FPR, but with the option of joint parameter estimation and outlier adjustment
STFFP	Forecast using equation (1) for all monitoring stations, STF modeling, fixed model parameters
STFPR	Forecast using equation (1) for all monitoring stations, STF modeling, model parameters re-estimated at each time step

low winds, and no precipitation; these are typical features of summertime high-pressure systems that last several days over a given area (they are synoptic scale motions of the atmosphere). Hence pure persistence tends to perform well in an ongoing ozone episode, and so it is a good alternative forecast to compare against.

As we have mentioned above, there are several possible combinations of implementations of the forecasting model (1), and Table V gives all the modelling options studied in this work, including the pure persistence option. The joint parameter estimation and outlier adjustment method is not available for STF models, so that the combination could not be tested here.

RESULTS OF MODEL EVALUATION

The performance of the different forecasting models was evaluated using two types of measures, whether we considered the forecast to be continuous or categorical, respectively. For the continuous forecast, we chose the root mean squared error (RMSE), the mean absolute error (MAE) and the cumulative sum of errors (CUSUM); all of these have the same units as the independent variable, in this case ozone concentrations in parts per billion (ppb). These indices will give an overall assessment of the similarity among observed and forecasted values, an estimate of a typical forecast error, and a quick comparison between alternative models. Nevertheless, these measures do not necessarily reflect the behaviour of the forecast under extreme circumstances (highest and lowest observed concentrations). To include this, we take the continuous forecast and split it into two categories: 'good' and 'bad' air quality. The definition of what constitutes a good or bad (episodic) day depends on the monitoring site, because ozone impacts tend to vary spatially across an urban area. Here we loosely defined the ozone threshold as the 50th percentile of the ozone daily maxima distribution at each station. Therefore we use indices based upon the classification displayed on Table VI. From this contingency table we can develop several quantitative indices, but here we focus on the following three (Wilks, 1995):

- (a) The probability of detection (POD): this measures the fraction of episodes correctly forecast by the model, and is given by the ratio $POD = n_{11}/(n_{11} + n_{12})$.

Table VI. Contingency table for assessing episodic forecasts

Combinations	Forecast yes	Forecast no
Observed yes	n_{11}	n_{12}
Observed no	n_{21}	n_{22}

- (b) Heidke's skill score (S): this is a measure of the skill of a set of forecasts compared to the skill of a random forecast; it is defined by

$$S = 2(n_{11}n_{22} - n_{12}n_{21}) / \{(n_{12})^2 + (n_{21})^2 + 2n_{11}n_{22} + (n_{12} + n_{21})(n_{11} + n_{22})\}$$

- (c) False alarm rate (FAR): this measures the tendency of the forecasting model to overpredict episodes, and is given by the ratio $FAR = n_{21} / (n_{11} + n_{21})$.

For further details on these categorically derived measures, the reader is referred to Doswell *et al.* (1990) and Wilks (1995).

FORECASTING RESULTS

Tables VII and VIII summarize the evaluation of the ozone-forecasting model based upon equation (1), under different modelling options. In Table VII we can see that the models are significantly better than the pure persistence forecast. There is a significant improvement upon MAE and RMSE when any of the implementations of model (1) is compared against the pure persistence forecast given by equation (5). The exception is CUSUM, and this shows that a model like (1) may be consistently under- or overestimating the ozone and still providing better forecasts (as measured by MAE and RMSE) than a pure persistence forecast (that tends to have low values of CUSUM by definition).

When we analyse the categorically based measures in Table VIII, it is also easy to discriminate among competing models by using measures such as POD, S and FAR. Once again, all implementations of model (1) outperform the pure persistence forecast, in terms of higher values of POD and S and lower values for FAR.

It seems that all variations of implementation of equation (1) yield similar results. There is no modelling option that performs consistently above the rest. All model implementations use essentially the same amount of information and they are all alike (linear), thus a similar behaviour was expected.

Figure 4 shows the forecasting performance for data set number 2, station M. Figure 4(a) displays a comparison of the actually observed daily peak ozone with the different model outputs; Figure 4(b) shows the absolute forecasting errors in each model evaluated. It can be seen in Figure 4 that all methods based upon equation (1) have a similar performance, and that all these methods outperform the pure persistence forecast given by equation (4); the typical standard error of the forecast is around 30 ppb for the methods applied to station M data. Similar results (not shown) are found for the other stations, showing a consistent behaviour across monitoring stations and also the good performance for low and high ozone impacts, that is, the models perform well under extreme impacts. We ascribe this feature to the physical origin of equation (1) and thus to its

Table VII. Evaluation of ozone forecasting models using continuous error measures

Set	Station	PP	FFP	FPR	OFFP	OFPR	STFFP	STFPR
(a) RMSE (ppb)								
1	A	18.16	10.89	10.93	12.30	11.73	10.91	10.96
1	C	21.71	17.48	16.76	17.29	17.03	17.48	16.82
1	D	22.19	14.69	14.45	14.62	14.58	14.39	14.24
1	M	24.89	21.58	20.03	21.32	20.05	21.89	20.28
2	A	27.98	14.88	14.81	15.42	15.13	15.59	15.70
2	C	27.42	14.84	14.83	16.52	15.56	14.97	14.98
2	D	21.09	15.54	15.10	15.62	15.48	16.42	15.38
2	M	41.06	19.48	19.51	19.90	19.84	19.13	19.19
3	C	29.80	16.83	16.80	17.82	18.16	16.76	16.68
3	D	21.57	11.69	11.76	13.00	12.26	11.70	11.78
3	M	34.09	20.21	20.00	20.37	21.86	20.04	19.94
(b) MAE (ppb)								
1	A	14.96	7.58	7.58	9.25	8.98	7.63	7.69
1	C	16.98	12.77	12.21	12.69	12.39	12.74	12.25
1	D	18.28	10.31	10.13	10.64	10.58	10.05	9.89
1	M	20.74	17.20	16.01	16.18	15.25	17.50	16.18
2	A	22.68	12.63	12.61	12.53	12.36	12.54	12.58
2	C	27.42	11.54	11.52	13.05	12.30	11.81	11.70
2	D	16.45	12.90	12.46	12.64	12.28	13.13	12.34
2	M	32.06	16.50	16.50	17.03	16.95	15.89	15.95
3	C	22.94	12.63	12.59	13.51	13.83	12.56	12.52
3	D	16.59	8.79	8.77	9.91	9.28	8.74	8.73
3	M	25.94	16.53	16.24	16.56	16.58	16.21	16.11
(c) CUSUM (ppb)								
1	A	72.0	2.7	6.6	-145.8	-126.7	4.9	6.5
1	C	61.0	316.0	275.2	-157.6	-126.3	305.1	259.7
1	D	64.0	-191.9	-164.9	-87.4	-70.5	-177.7	-153.8
1	M	109.0	651.3	558.7	274.7	272.5	673.9	579.1
2	A	28.0	141.0	130.4	-85.2	-59.7	153.8	137.3
2	C	15.0	-107.7	-100.6	-133.3	-95.0	-155.8	-131.0
2	D	8.0	116.1	88.2	-85.0	-137.0	153.1	106.1
2	M	31.0	49.3	60.2	-3.3	13.3	35.0	46.4
3	C	6.0	-94.7	-88.5	-50.1	-69.8	-102.8	-101.2
3	D	6.0	14.7	7.6	-30.1	-51.3	8.8	1.3
3	M	-14.0	-161.7	-146.8	-138.7	-137.5	-176.2	-165.6

ability to capture the most relevant (measured) variables influencing the peak ozone level; in fact, the input variables used in model (1) are quite similar to those used in other cases reported in the literature.

To explore the effect of increasing the forecast horizon, we compared one-, two-, and three-steps ahead forecasts in terms of the performance indices. Table IX shows such an exercise for data set 2 and four forecasting implementations. It is clear in Table IX that as the forecast horizon increases, the performance deteriorates and the forecast itself becomes more conservative—because it includes less information to make the forecast—so it tends to approach the mean of the observed concentrations. Hence the ratio of the standard deviations of the forecast and observed concentrations decreases as the forecast horizon increases. However, the deterioration in the forecasting

Table VIII. Evaluation of ozone forecasting models using categorically based indices

Set	Station	PP	FFP	FPR	OFFP	OFPR	STFFP	STFPR
(a) POD: Power of detection								
1	A	0.72	0.88	0.88	0.72	0.68	0.88	0.88
1	C	0.73	0.88	0.88	0.58	0.65	0.88	0.88
1	D	0.71	0.63	0.63	0.67	0.67	0.63	0.63
1	M	0.56	1.00	1.00	1.00	0.94	1.00	1.00
2	A	0.50	0.90	0.90	0.65	0.65	0.85	0.85
2	C	0.59	0.70	0.67	0.63	0.67	0.67	0.67
2	D	0.76	0.96	0.96	0.92	0.88	0.92	0.96
2	M	0.63	0.81	0.81	0.81	0.81	0.89	0.89
3	C	0.60	0.93	0.93	0.93	0.93	0.93	0.93
3	D	0.67	1.00	1.00	1.00	1.00	0.80	0.80
3	M	0.59	0.82	0.88	0.82	0.76	0.82	0.88
(b) S: Heidke's skill score								
1	A	0.40	0.72	0.72	0.60	0.64	0.72	0.68
1	C	0.40	0.56	0.56	0.45	0.44	0.56	0.56
1	D	0.40	0.51	0.51	0.52	0.52	0.51	0.51
1	M	0.27	0.39	0.42	0.55	0.53	0.39	0.42
2	A	0.20	0.62	0.62	0.51	0.54	0.51	0.51
2	C	0.17	0.40	0.36	0.40	0.36	0.40	0.36
2	D	0.55	0.70	0.74	0.70	0.66	0.63	0.66
2	M	0.21	0.47	0.47	0.47	0.47	0.55	0.55
3	C	0.25	0.69	0.63	0.69	0.69	0.63	0.63
3	D	0.31	0.51	0.51	0.63	0.63	0.21	0.21
3	M	0.19	0.75	0.81	0.69	0.63	0.75	0.81
(c) FAR: False alarm rate								
1	A	0.31	0.15	0.15	0.14	0.06	0.15	0.19
1	C	0.30	0.26	0.26	0.17	0.23	0.26	0.28
1	D	0.32	0.17	0.17	0.20	0.20	0.17	0.17
1	M	0.47	0.49	0.47	0.40	0.39	0.49	0.47
2	A	0.50	0.31	0.31	0.28	0.23	0.37	0.37
2	C	0.41	0.30	0.31	0.26	0.31	0.28	0.31
2	D	0.24	0.23	0.20	0.21	0.21	0.27	0.25
2	M	0.39	0.29	0.29	0.29	0.29	0.27	0.27
3	C	0.40	0.22	0.26	0.22	0.22	0.26	0.26
3	D	0.38	0.35	0.35	0.29	0.29	0.45	0.45
3	M	0.38	0.07	0.06	0.13	0.13	0.07	0.06

performance is not dramatic, suggesting that the ozone forecasting model could be used to obtain reasonable multi-day forecasts.

Figure 5 shows two examples of the forecast behaviour for one-, two-, and three-steps ahead forecasts, for stations C and M. For the sake of clarity, standard errors of the forecasts are not shown (they are displayed in Table IX for comparison purposes) but typical values are 20 and 30 ppb for stations C and M, respectively, that is, they can be assessed by the tick marks on the y-axis in both plots. The multistep forecasts are alike, except that the one-step-ahead forecasts tend to enclose the other forecasts, that is, they tend to show the highest and lowest forecast values.

More work needs to be done to evaluate our proposed model with data from other cities worldwide. Nevertheless, other studies also agree with the basic structure given by equation (1). For

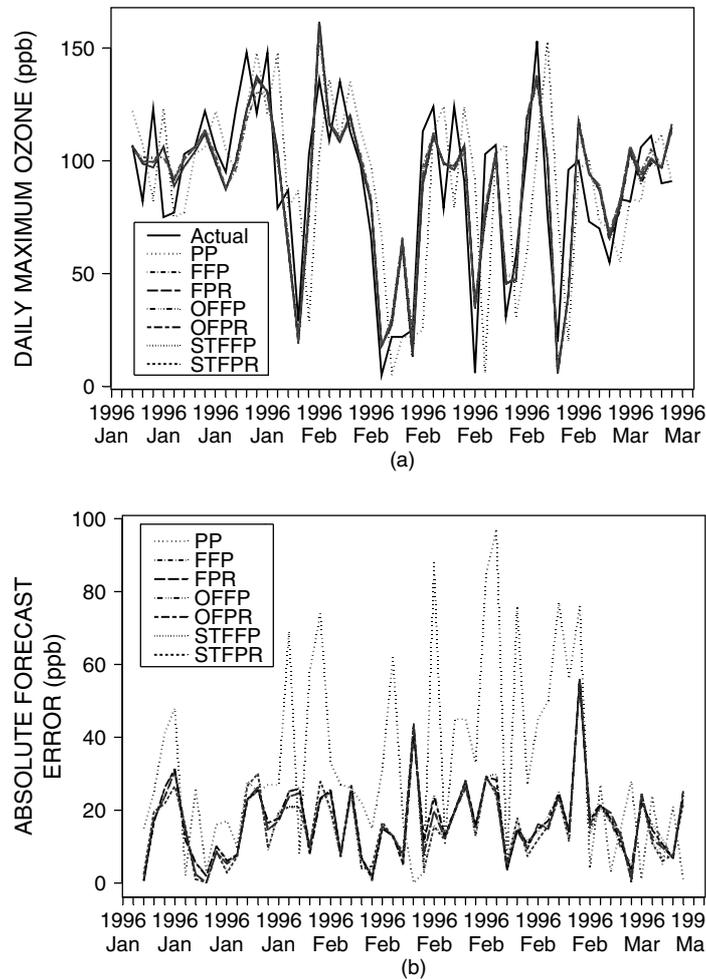


Figure 4. Performance of forecasting models for station M, data set 2; (a): comparison of actual daily peak ozone with the forecasts from the different models evaluated; (b): comparison of absolute forecasting errors for all models evaluated

instance, Ryan (1995) has found that maximum air temperature (T_2) explains most of the variance in peak ground ozone levels in the Baltimore metropolitan area and that the lower morning temperatures (T_1) were also correlated with high ozone impacts; similar findings were reported by Hubbard and Cobourn (1998) for the Louisville, Kentucky, metropolitan area.

CONCLUSIONS

A physically based ground-level ozone-forecasting model has been identified and evaluated for Santiago, Chile. The model is derived from a simplification of the continuity equation followed by an air parcel that moves across an urban region.

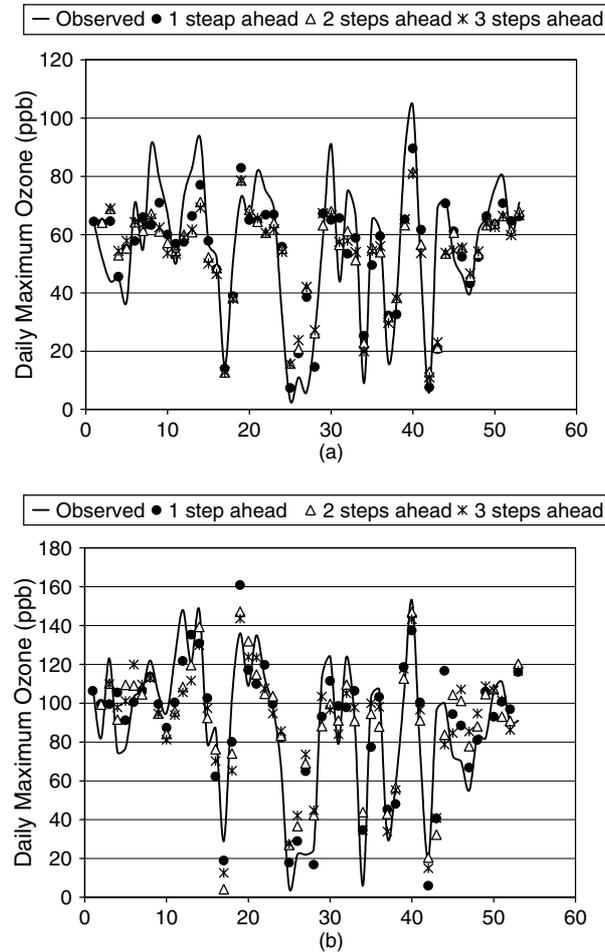


Figure 5. Performance of multistep forecasts for stations C and M, data set 2; (a) comparison of actual daily peak ozone at station C with the 1,2 and 3-steps ahead forecasts from the FFP implementation of the forecasting model; (b) the same, but for station M

The model-identification process produced parameter estimates that were similar across monitoring stations, and that changed slowly with time for each separate site. Parameter estimates were quite significant, especially when the noise was a low-order moving average process. Diagnostics of models (1) and (2) showed essentially white-noise behaviour in the residuals (see Figures 2 and 3); hence the models are adequate for the purposes of forecasting.

In the case of the ozone model (1), weekend effects were similar to those reported in the literature, with Saturdays showing higher ozone impacts than weekdays near downtown Santiago; Sundays and Mondays tended to show lower ozone impacts than on weekdays, especially at the suburban, downwind station.

Rainy days decrease 40–50% the values of the daily rise of air temperature (ΔT), with respect to the seasonal average. The characteristic time for these effects to fade away was 2–3 days, that is, of the same periodicity as synoptic scale motions.

Table IX. Evaluation of ozone forecasting models for one-, two-, and three-steps ahead forecasts

Steps ahead	RMSE (ppb)	MAE (ppb)	CUSUM (ppb)	$\sigma_{\text{Forecast}}/\sigma_{\text{Observed}}$	Standard error of forecast (ppb)
(a) Station A					
1	14.88	12.63	-141.0	0.734	20.31
2	15.58	13.22	-175.4	0.691	20.87
3	15.59	13.12	-183.3	0.683	21.03
(b) Station C					
1	14.84	11.54	107.7	0.764	19.39
2	15.44	12.46	144.9	0.668	20.45
3	15.58	12.58	170.7	0.662	20.56
(c) Station D					
1	15.54	12.90	-116.1	0.795	18.25
2	17.53	14.36	-152.8	0.763	18.83
3	18.69	15.33	-165.3	0.732	19.03
(d) Station M					
1	19.48	16.50	-49.3	0.911	28.36
2	21.19	17.02	-64.4	0.840	31.16
3	21.04	17.15	-53.2	0.829	31.93

Note: Only the FFP forecast results are displayed; results for the FPR, OFFP and OFPR implementations are similar and not shown here.

The model certainly does have limitations, for example the lack of upper air measurements (cloud cover, mixing height) incorporated into the model. Despite this, model performance is reasonable for out-of-sample forecasts in Santiago. We ascribe this to the persistence of wind patterns in the city basin, which are dominated by differential heating in this confined valley (see Figure 1), and to the high correlation between air temperature, mixing height and cloud cover. In other words, the variables not included in the model (wind direction, cloud cover, mixing height) do not seem to influence model performance in a significant manner. This good forecasting performance is in agreement with models with similar input variables that have performed well in other cities abroad (Ryan, 1995; Comrie, 1997; Hubbard and Cobourn, 1998).

One possible explanation for this reasonable model performance is that air temperature is a good surrogate for incoming radiation and mixing height. Hence adding variables such as cloud cover or mixing height do not add further information to the model. On the other hand, wind speed and direction have shown to have significant effects upon ozone impacts, in regions where synoptical motions bring in air masses from different upwind regions (Ryan, 1995; Hubbard and Cobourn, 1998). In these cases of more 'open' topography, the structure of the model increases in complexity and certainly equations (1) and (2) will require improvements in order to obtain suitable forecasting models. Nonetheless, The results of the model evaluation suggest it might be useful in other cities to develop ozone-forecasting tools.

ACKNOWLEDGEMENTS

We are grateful for funding to support this research from Conicyt-Chile through grants Fondecyt 198-0972 and 80000004. We acknowledge support from Mr Ignacio Olaeta (pvca@sesma.cl)

for making the MACAM database available through the Internet. We also thank two anonymous referees for their comments and suggestions that greatly improved the scope and depth of this work.

REFERENCES

- Acuña G, Jorquera H, Pérez R. 1996. Neural network model for maximum ozone concentration prediction. In von der Matsburg C *et al.* (eds). *Lecture Notes in Computer Science No. 1112*, Springer: Berlin, 263–268.
- Adya M, And Collpy F. 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting* **17**: 481–495.
- Aron R, Aron IM. 1978. Statistical forecasting models: II. Oxidant concentration in the Los Angeles basin. *J. Air Pollut. Control Assoc.* **28**: 684–688.
- Beck MB. 1991. Forecasting environmental change. *Journal of Forecasting* **10**: 3–19.
- Cannon A, Lord ER. 2000. Forecasting summertime surface-level ozone concentrations in the Lower Fraser Valley of British Columbia: an ensemble neural network approach. *J. Air & Waste Manage. Assoc.* **50**: 322–339.
- Chaum D, Elkus B, Wilson KR, Rice JR. 1978. A statistically tested short term oxidant control strategy. *Atmospheric Environment* **12**: 1777–1783.
- Chen C, Liu LM. 1993a. Joint estimation of model parameters and outlier effects in time series. *J. American Stat. Assoc.* **88**: 284–297.
- Chen C, Liu LM. 1993b. Forecasting time series with outliers. *Journal of Forecasting* **12**: 13–35.
- Cleveland WS, Graedel TE, Kleiner B, Warner JL. 1974. Sunday and workday variations in photochemical air pollutants in New Jersey. *Science* **186**: 1037–1038.
- Cleveland WS, McRae JE. 1978. Weekday–weekend ozone concentrations in the Northeast United States. *Environmental Science and Technology* **12**: 558–563.
- Comrie AC. 1997. Comparing neural networks and regression models for ozone forecasting. *J. Air & Waste Manage. Assoc.* **47**: 653–663.
- Davis JM, Speckman P. 1999. A model for predicting maximum and 8h average ozone in Houston. *Atmospheric Environment* **33**: 2487–2500.
- Desalu AA, Gould LA, Schweppe FC. 1974. Dynamic estimation of air pollution. *IEEE Trans. Aut. Control* **AC-19**: 904–910.
- Doswell CA III, Jones R, Keller DL. 1990. On summary measures of forecast skill in rare event forecasting based on contingency tables. *Weather Forecasting* **5**: 576–585.
- Elsom D. 1996. *Smog Alert. Managing Urban Air Quality*. Earthscan: London.
- Fenger J. 1999. Urban air quality. *Atmospheric Environment* **33**: 4877–4900.
- Gardner MW, Dorling SR. 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment* **32**: 2627–2636.
- González-Manteiga W, Prada-Sánchez JM, Cao R, García-Jurado I, Febrero-Bande M, Lucas-Domínguez T. 1993. Time-series analysis for ambient concentrations. *Atmospheric Environment* **27A**: 153–158.
- Hanna SR, Lu Z, Frey H, Wheeler N, Vukovich J, Arunachalam S, Fernau M, Hansen AD. 2000. Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmospheric Environment* **35**: 891–903.
- Hubbard M, Cobourn WG. 1998. Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. *Atmospheric Environment* **32**: 2637–2647.
- Jorquera H, Pérez R, Cipriano A, Espejo A, Letelier MV, Acuña G. 1998a. Forecasting ozone daily maximum levels at Santiago, Chile. *Atmospheric Environment* **32**: 3415–3424.
- Jorquera H, Olguín C, Ossa F, Pérez R, Solar I, Encalada O. 1998b. An assessment of photochemical pollution at Santiago, Chile. *Air Pollution VI*, Brebbia CA, Ratto CF, Power H (eds). Computational Mechanics Publications: Boston, MA; 743–753.
- Jorquera H, Palma W, Tapia J. 2000. An intervention analysis of air quality data at Santiago, Chile. *Atmospheric Environment* **34**: 4073–4084.
- Kao J, Huang S. 2000. Forecasts using neural networks versus Box–Jenkins methodology for ambient air quality monitoring data. *J. Air & Waste Manage. Assoc.* **50**: 219–226.

- Karl TR. 1978. Day of the week variations of photochemical pollutants in the St. Louis area. *Atmospheric Environment* **12**: 1657–1667.
- Levitt SB, Chock DP. 1976. Weekday–Weekend pollutant studies of the Los Angeles basin. *J. Air Pollut. Control Assoc.* **26**: 1091–1092.
- Liu LM. 1989. Identification of seasonal ARIMA models using a filtering method. *Communications in Statistics* **18**: 2279–2288.
- Liu LM. 1991. Use of linear transfer function analysis in econometric time series modeling. *Statistica Sinica* **1**: 503–525.
- Liu LM, Hanssens DM. 1982. Identification of multiple-input transfer function models. *Communications in Statistics A* **11**: 297–314.
- Liu LM, Hudak GB, Box GEP, Muller ME, Tiao GC. 1986. *The SCA statistical system: Reference manual for forecasting and time series analysis*. Scientific Computing Associates: DeKalb, IL.
- McCullister GM, Wilson KR. 1975. Linear stochastic models for forecasting daily maximum and hourly concentrations of air pollutants. *Atmospheric Environment* **9**: 417–423.
- Melas D, Kioutsioukis I, And Ziomas IC. 2000. Neural network model for predicting peak photochemical pollutant levels. *J. Air & Waste Manage. Assoc.* **50**: 495–501.
- Meng Z, Dadub B, Seinfeld JH. 1997. Chemical coupling between atmospheric ozone and particulate matter. *Science* **227**: 116–119.
- Merz PH, Painter HJ, Ryason PR. 1972. Aerometric dates analysis—time series analysis and forecast and an atmospheric smog diagram. *Atmospheric Environment* **6**: 319–327.
- Myrabo LE, Wilson KR, Trijonis JC. 1976. Survey of statistical models for oxidant air quality prediction. In *Advances in Environmental Science and Technology, Vol. VII*, Pitts JN, Metcalf RL (eds). Wiley: New York, 391–422.
- Ng CN, Yan TL. 1998. Recursive modelling and adaptive forecasting of air quality data. In *Air Pollution VI*, Brebbia CA, Ratto CF, Power H (eds). Computational Mechanics Publications: Boston, MA; 869–877.
- Niu W. 1996. Nonlinear additive models for environmental time series, with applications to ground-level ozone data analysis. *Journal of the American Statistical Association* **91**: 1310–1321.
- Noordijk H. 1994. The national smog warning system in The Netherlands, a combination of measuring and modelling. In *Air Pollution II: Pollution Control and Monitoring*, Baldesano JM *et al.* (eds). Computational Mechanics Publications: New York, 169–176.
- Pankratz A. 1991. *Forecasting with Dynamic Regression Models*. Wiley: New York.
- Pryor SC, Schiess JR, McDougal DS. 1981. Approach to forecasting daily maximum ozone levels in St. Louis. *Environmental Science and Technology* **15**: 430–436.
- Pryor SC, Steyn DG. 1995. Hebdomadal and diurnal cycles in ozone time series from the Lower Fraser Valley, B.C. *Atmospheric Environment* **29**: 1007–1019.
- Robeson SM, Steyn DG. 1990. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment* **24B**: 303–312.
- Rutllant J, Garreaud R. 1995. Meteorological air pollution potential for Santiago, Chile: Towards an objective episode forecasting. *Environmental Monitoring and Assessment* **34**: 223–244.
- Ryan WF. 1995. Forecasting severe ozone episodes in the Baltimore metropolitan area. *Atmospheric Environment* **29**: 2387–2398.
- Schlink U, Herbarth O, Tetzlaff G. 1997. A component time-series model for SO₂ data: Forecasting, interpretation and modification. *Atmospheric Environment* **31**: 1285–1295.
- Seinfeld JH, Pandis SN. 1998. *Atmospheric Chemistry and Physics*. Wiley: New York.
- Simpson RW, Leyton AP. 1983. Forecasting peak ozone levels. *Atmospheric Environment* **17**: 1644–1654.
- Wilks DS. 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press: San Diego, CA.
- Yi J, Prybutok R. 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution* **92**: 349–357.
- Young PC. 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Env. Modelling and Software* **13**: 105–122.
- Young PC, Ng CN, Lane K, Parker D. 1991. Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data. *Journal of Forecasting* **10**: 57–89.
- Young PC, Jakeman AJ, Post DA. 1997. Recent advances in data-based modelling and analysis of hydrological systems. *Water Science & Tech.* **36**: 99–116.

Zannetti P. 1990. *Air Pollution Modelling—Theories, Computational Methods and Available Software*. Computational Mechanics Publications: New York.

Authors' biographies:

Héctor Jorquera (PhD in Chemical Engineering, University of Minnesota, 1991) is an Associate Professor in the Chemical and Bioprocess Engineering Department, School of Engineering, Pontificia Universidad Católica de Chile. His research interests include air quality modelling and forecasting and environmental regulation.

Wilfredo Palma (PhD in Statistics, Carnegie Mellon University, 1995) is an Assistant Professor in the Department of Statistics, Pontificia Universidad Católica de Chile. His research interests include time series analysis, forecasting financial and environmental data.

José Tapia is an Assistant Professor at the Universidad de Valparaíso and a PhD candidate in the Department of Statistics, Pontificia Universidad Católica de Chile. His research interests are in the areas of long-range time series and environmental dynamics.

Authors' addresses:

Héctor Jorquera, Chemical and Bioprocess Engineering Department, School of Engineering, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago 6904411, Chile.

Wilfredo Palma, Department of Statistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago 6904411, Chile.

José Tapia, Department of Statistics, Pontificia Universidad Católica de Chile, Vicuña Mackenna 4860, Santiago 6904411, Chile.