



**HAL**  
open science

## Application of nonlinear forecasting techniques for meteorological modeling

V. Pérez-Muñuzuri, I. R. Gelpi

► **To cite this version:**

V. Pérez-Muñuzuri, I. R. Gelpi. Application of nonlinear forecasting techniques for meteorological modeling. *Annales Geophysicae*, 2000, 18 (10), pp.1349-1359. hal-00316807

**HAL Id: hal-00316807**

**<https://hal.science/hal-00316807>**

Submitted on 18 Jun 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Application of nonlinear forecasting techniques for meteorological modeling

V. Pérez-Muñuzuri, I. R. Gelpi

Group of Nonlinear Physics, Faculty of Physics, University of Santiago de Compostela, E-15706 Santiago de Compostela, Spain

Received: 15 November 1999 / Revised: 13 June 2000 / Accepted: 14 June 2000

**Abstract.** A nonlinear forecasting method was used to predict the behavior of a cloud coverage time series several hours in advance. The method is based on the reconstruction of a chaotic strange attractor using four years of cloud absorption data obtained from half-hourly Meteosat infrared images from Northwestern Spain. An exhaustive nonlinear analysis of the time series was carried out to reconstruct the phase space of the underlying chaotic attractor. The forecast values are used by a non-hydrostatic meteorological model ARPS for daily weather prediction and their results compared with surface temperature measurements from a meteorological station and a vertical sounding. The effect of noise in the time series is analyzed in terms of the prediction results.

**Key words:** Meteorology and atmospheric dynamics (mesoscale meteorology; general) – General (new fields)

---

## 1 Introduction

The accuracy of Earth radiation budget estimates is highly dependent on how well cloud variability is taken into account. Because of its dynamic nature and pronounced optical characteristics, cloud cover is one of the most important variables affecting the radiation balance, which, for example is determinant for dispersion modeling of pollutants and, ultimately, the global climate, among other effects.

Standard meteorological models are based on finite difference solutions of the hydrothermodynamical equations. Among the different parameterizations that are taken into account in these models, subgrid-scale cloud dynamics must be implemented in order to describe its

influence in the atmospheric flow and pollutants transport (Donner, 1993; Gimson, 1997; Mölders *et al.*, 1994). The degree of sophistication of cumulus parameterizations for weather forecasting varies widely, according to the computer resources needed for the host model, and the complexity of physical and chemical processes that the model is required to describe. That is, the parameterized clouds may be diagnosed under some assumptions regarding their areas, mass fluxes, and entrainment rates, or more closely linked to the host model dynamics. They may be in a steady-state balance or evolve in time, depending on the time step of the host model. Then, to describe the effect of cumulus convection, model resolution must be sacrificed to limit computational costs, and cumulus effects must be parameterized.

Here, we have tried to combine two different points of view of analyzing meteorological data, namely: the standard atmospheric circulation models and a new approach based on the analysis of time series data of physical observations, whose dynamics exhibit irregular or chaotic behavior. The latter case, which may be called *nonlinear time series analysis*, explores the possibility and the extent to which the dynamics generating this time series is deterministic, i.e., occurs on a low-dimensional chaotic attractor. The main idea is that the aperiodicity in the data is not due to stochasticity but due to nonlinearity. The study of irregular time series data by means of nonlinear methods has become a popular task within the last years. Nonlinear time series analysis is described in two recent monographs; one by Abarbanel (1996) and one by Kantz and Schreiber (1997) as well as in several reviews, including Grassberger *et al.* (1991), Abarbanel *et al.* (1993), Schreiber (1999) and Hegger *et al.* (1999)<sup>1</sup>. Moreover, under appropriate conditions the phase space corresponding to the low-dimensional attractor can be reconstructed such that it predicts the future using a local forecasting function computed from the attractor. The main idea is

---

Correspondence to: V. Pérez-Muñuzuri  
e-mail: vicente@fmneteo.usc.es; http://fmneteo.usc.es

---

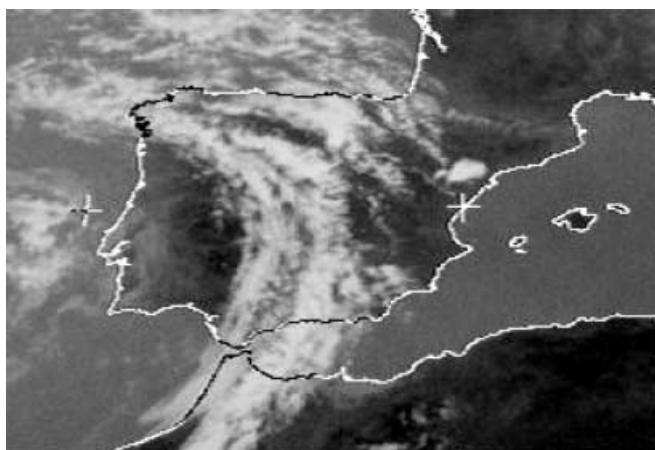
<sup>1</sup> See also <http://www.mpi-pks-dresden.mpg.de/~tisean>.

that for deterministic systems, points which are close in the reconstructed phase space stay close under forward iteration.

In this paper, techniques of nonlinear analysis are applied to a time series of semi-hourly cloud absorption values obtained from infrared Meteosat images during the last four years for cloud coverage forecasting. The predicted cloud absorption values are then used by a meteorological model for weather forecasting. Previous results with a shorter time series of data were presented in two former papers by Souto *et al.* (1998) and Pérez-Muñuzuri (1998), where the effect of cloud coverage in plume dispersion modeling in Northwestern Spain was studied in detail. Here, in Sects. 2 and 3 we will focus on the nonlinear analysis of the four year time series of cloud absorption values. To avoid verbosity, the derivation of the methods will be kept to a minimum. The use of the predicted cloud absorption values for weather forecasting by using a non-hydrostatic meteorological model is presented in Sects. 4 and 5.

## 2 Nonlinear prediction

We have collected a time series of cloud absorption values  $\xi$  during four years from semi-hourly Meteosat infrared images from Northwestern Spain (Fig. 1) in order to predict  $\xi$ . We obtain a single average value of the cloud cover over the area of interest (Galicia) from these images. Values range from 0 to 1 where  $\xi = 1$  means that no radiation is emitted from the surface. For clear skies, the stored value is close to zero; intermediate values depend on cloud thickness and water phase and content. Values of  $\xi(t)$  could also be considered to depend on the grid position  $(x, y)$ , but this would require higher precision satellite images than those available from Meteosat. Furthermore, the noise in the time series increases with a finer image resolution, which decreases the precision of the nonlinear forecasting method described below.



**Fig. 1.** Infrared Meteosat image of the Iberian Peninsula corresponding to July 26, 1999 (12:00 UTC)

In the following, the cloud dynamics is supposed to belong to the category of dissipative system dynamics, which can reveal dynamics with strange attractor structure. The Lyapunov exponents determine how far into the future a successful forecast can be made and the number of variables needed to make the prediction is governed by the fractal dimension (Grassberger and Procaccia, 1983; Wolf *et al.*, 1985). The general nonlinear prediction method is to reconstruct the strange attractor from the set of data in a minimum embedding space dictated by the correlation fractal dimension and then predict the future using a local forecasting function computed from the attractor (Farmer and Siderowich, 1987; Casdagli, 1989; Abarbanel *et al.*, 1993; Abarbanel, 1996). Takens (1981) has proven that, after embedding, and if the embedding dimension  $m$  and time delay  $\tau$  are chosen appropriately, there exists a smooth map  $\mathcal{F}$  such that

$$\xi_{n+T} = \mathcal{F}(\Xi_n) \quad (1)$$

where  $\Xi_n$  is a vector of data points defined by

$$\Xi_n = (\xi_n, \xi_{n+\tau}, \dots, \xi_{n+(m-1)\tau}) \quad (2)$$

and  $n$  is the number of data points in the time series and  $T$  is the forecasting time.

The meaning of the theorem is that, under some general conditions, the orbit followed by  $\Xi_n$  in this  $m$ -dimensional embedding space will differ from the actual solution of  $\mathcal{F}(\Xi_n)$  in Eq. (1) only by a smooth change of coordinates. The most common solution of Eq. (1) is to compute the Jacobian of the strange attractor in the vicinity of a target point and then use it to predict the future of that point. Typically, one finds two to five times the minimum number of nearest neighbors  $N_B$  needed to compute the Jacobian and makes a least squares fit using the nearest neighbors and their future iterates to compute the prediction function  $\mathcal{F}$ . This must be repeated for every target point. The procedure consists of the following steps<sup>2</sup>:

1. Choose an embedding time delay  $\tau$  using the mutual information (Fraser and Swinney, 1986) defined as,

$$I(\tau) = - \sum_{i,j} p_{ij} \ln \frac{p_{ij}(\tau)}{p_i p_j} \quad (3)$$

where  $p_i$  is the probability to find a time series value in the  $i$ -th interval, and  $p_{ij}(\tau)$  is the joint probability that an observation falls into the  $i$ -th interval and the observation  $\tau$  later falls into the  $j$ -th. Unlike the autocorrelation function, the mutual information also takes into account nonlinear correlations. It can be shown, that if the time delayed mutual information exhibits a marked minimum at a certain value of  $\tau$ , then this is a good candidate for a reasonable time delay to reconstruct the phase space of the chaotic attractor.

<sup>2</sup> See Hegger *et al.* (1999) for a practical implementation of these nonlinear time series methods through the use of the TISEAN package and their proper usage.

2. Choose an embedding dimension  $m$  using the false nearest neighbor method (Kennel *et al.*, 1992; Hegger and Kantz, 1999). The idea of this method is the following: For each point  $\Xi_i$  in the time series look for its nearest neighbor  $\Xi_j$  in an  $m$ -dimensional space. Calculate the distance  $\|\Xi_i - \Xi_j\|$ . Iterate both points and compute,

$$R_i = \frac{\|\Xi_{i+1} - \Xi_{j+1}\|}{\|\Xi_i - \Xi_j\|} \quad (4)$$

If  $R_i$  exceeds a given threshold  $R_t$ , this point is marked as a false nearest neighbor. The criterion to consider the embedding dimension high enough is that the number of points verifying  $R_i > R_t$  is zero or, at least, small enough. In other words, the false nearest neighbors are the nearest neighbors on an attractor that arrive near each other by projection because the attractor is being viewed in a space dimension too low to completely unfold it.

3. Construct a vector  $\Xi_i$  by using Eq. (2).

4. Find the  $N_B$  nearest neighbors  $\Xi_j$  of  $\Xi_i$  ( $j < i$ ) that minimize  $\|\Xi_i - \Xi_j\|$ .

5. Select a forecasting time  $T$ .

6. Order the neighbors from closest to furthest, and find an affine model in the following form (Farmer and Siderowich, 1987),

$$\xi_{j(l)}^N = \alpha_0 + \sum_{s=1}^m \alpha_s \xi_{j(l)-(s-1)\tau}^{N-1} \quad l = 1, \dots, N_B \quad (5)$$

where  $\xi_{j(l)}^{N-1}$  is the  $j$ th coordinate of the  $l$ th neighbor of  $\Xi_{N-1}$  for a time series consisting of  $N - 1$  data, and  $\xi_{j(l)}^N$  is the  $j$ th coordinate of the point of the attractor to which  $\xi_{j(l)}^{N-1}$  goes in one time step. The coefficients  $\alpha_0, \dots, \alpha_m$  are the parameters of the model, computed by ordinary least squares. Then, to forecast a new point of the attractor  $\xi_j^{N+1}$  we use  $\xi_j^{N+1} = \alpha_0 + \sum_{s=1}^m \alpha_s \xi_{j-(s-1)\tau}^N$ , and the process is repeated  $T$  steps ahead.

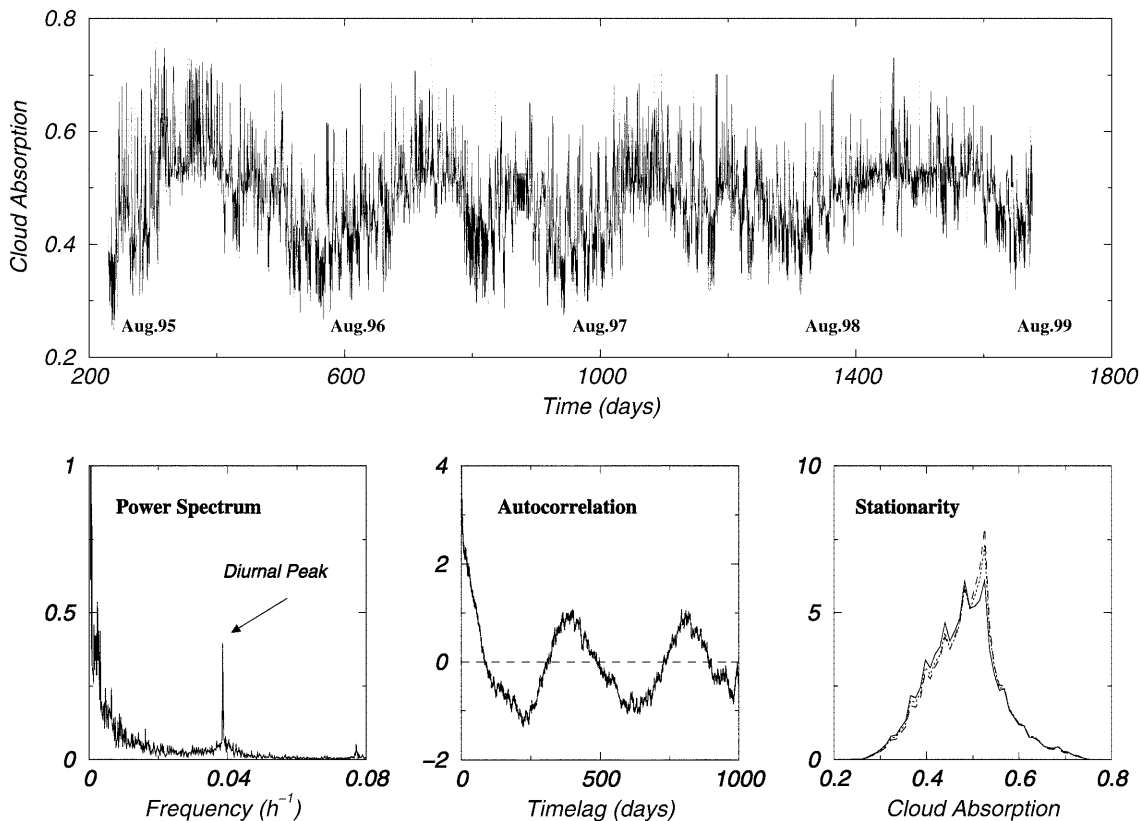
To evaluate the accuracy of the nonlinear predictions, the normalized root-mean-square error (RMSE) (Farmer and Siderowich, 1987) is used,

$$E = \left[ \frac{\sum_i (\xi_{\text{pred}}(i) - \xi(i))^2}{\sum_i (\xi(i) - \bar{\xi})^2} \right]^{1/2} \quad (6)$$

where  $\xi_{\text{pred}}(i)$  is the predicted value of the measured  $\xi(i)$ ,  $\bar{\xi}$  is the mean of the time series during the forecasting period of time. If  $E = 0$ , the predictions are perfect;  $E = 1$  indicates that the performance is no better than a constant predictor  $\xi_{\text{pred}}(i) = \bar{\xi}$ .

### 3 Chaotic analysis of the experimental time series

Figure 2 shows the time series of values of cloud coverage  $\zeta(t)$  obtained from the Meteosat during the last four years (from August 1995 to August 1999). The



**Fig. 2.** First row: cloud absorption  $\zeta$  series of semi-hourly measurements in north-western Spain (Galicia) obtained from the Meteosat and consisting of 73 000 points of data (approximately 4 years). The time series begins on August 1995, where day 1 is 1 January 1995.

Second row: from left to right; power spectrum of the full series, the autocorrelation coefficient, showing evidence of “recurrence” for periods of one year, and evidence of stationarity of the full time series

chaotic character of the time series is revealed by the broad band form of the power spectrum. Our results suggest that the time evolution of the data is governed by a strong deterministic component. Although, the high-frequency part of the spectrum is dominated by the daily frequency  $w_1 = 1/24 \text{ h}^{-1}$ , it is also possible to observe a continuous band of frequencies which departs from the low-frequency part of the spectrum. This is where the annual frequency ( $w_0 = 1/8760 \text{ h}^{-1}$ ) of our data lies, which forms the large “sine-wave” variation observed in the time series. These periodic behaviors are not so dominant as to lead us to consider the cloud coverage series as an annual and daily cycle with some small background chaos or noise imposed on it. Instead, we have some dynamics driven by periodic external solar forcing. Note as well that the time series reveals that the winter of 1995/1996 was particularly bad since cloud coverage reached its higher values, mainly because the European area was blocked for most of the winter with wind coming from the east, and very cold temperatures.

The autocorrelation function  $A(\tau)$  of our series, i.e.,

$$A(\tau) = \frac{1}{N} \sum_{i=1}^N (\bar{\xi}_i \cdot \bar{\xi}_{i+\tau}) , \quad (7)$$

where  $\bar{\xi}_i = \xi_i - \frac{1}{N} \sum_{i=1}^N \xi_i$ , does not decay very rapidly supporting the idea that our data contains a noise component (Bountis *et al.*, 1993). Besides, the time scale associated with the first zero crossing of the autocorrelation function gives a time at which variations of the cloud coverage  $\xi$  are linearly decorrelated ( $\tau = 83$  days). A time delay reconstruction of the phase space using this large time scale as the lag would quite likely reveal little, since this is clearly too long for any dynamical

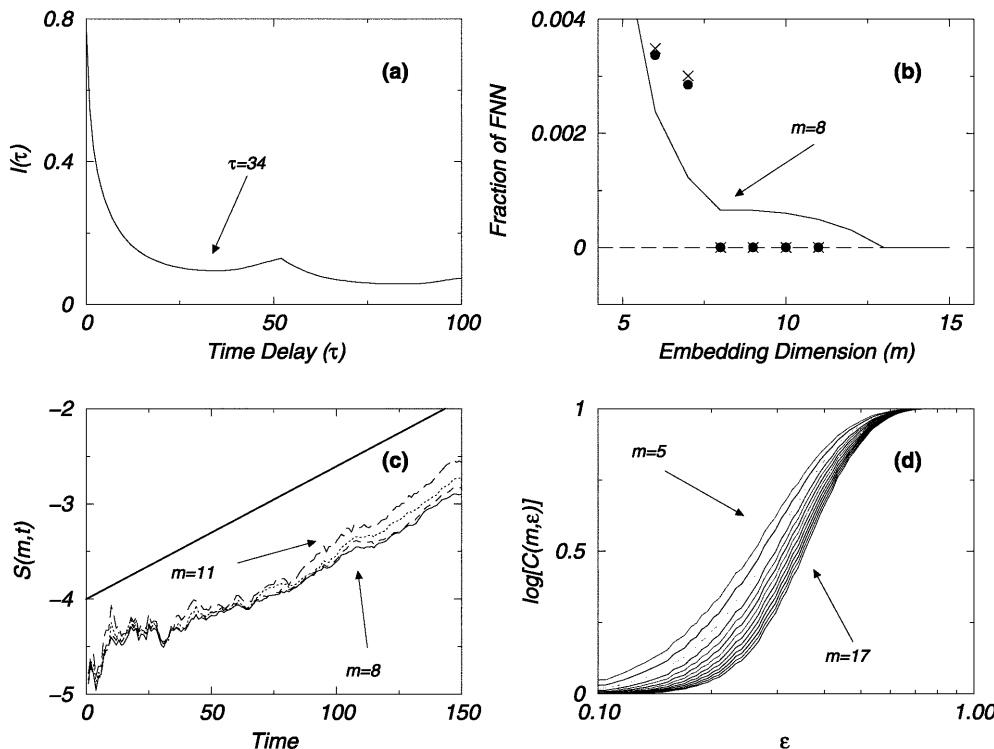
correlations in this system to persist. In sharp contrast, the mutual information  $I(\tau)$  discussed later will reveal new possibilities.

Stationarity is a necessary condition when the time series corresponds to the dynamic evolution on a strange attractor. Hence, the stationarity of the time series is tested by plotting the probability density function for the first 60 000, 65 000, 70 000 and the whole set of 73 000 points, with the average  $\bar{\xi}$  and the standard deviation  $\sigma_\xi$  of the data given in Table 1. Note that there are no significant time variations in that quantity.

The mutual information  $I(\tau)$  and the fraction of false nearest neighbors are shown in Fig. 3a, b, respectively. Mutual information  $I(\tau)$  shows a clear first minimum for  $\tau = 34$  or times of about 17 hours. Using a time lag  $\tau = 34$ , we use the whole data set to find out the fraction of false nearest neighbors (FNN) and to determine the minimum embedding dimension. Note from Fig. 3b that there is a sharp drop at  $m = 8$  after which the fraction of FNN remains equal to zero. The strength of this conclusion is enhanced when we see that varying the Theiler window (Hegger *et al.*, 1999) over the range  $[0, 2000]$  changes nothing but the detailed number of FNN at dimensions  $m < 8$ . For  $m \geq 8$  the fraction of

**Table 1.** Average value of the cloud absorption  $\bar{\xi}$  and the standard deviation  $\sigma_\xi$  for different time series lengths  $N$

$N$	$\bar{\xi}$	$\sigma_\xi$
60 000	0.481	0.079
65 000	0.485	0.078
70 000	0.486	0.077
73 000	0.487	0.076



**Fig. 3.** First row (a, b): the average mutual information  $I(\tau)$  and the fraction of false nearest neighbors (FNN). Here, three values of the Theiler window were used;  $w = 0$  (solid line),  $w = 100$  (○) and  $w = 1000$  (×). Second row (c, d): estimation of the maximal Lyapunov exponent  $\lambda_{\max}$  and the correlation dimension  $D_2$ . In the first case, the straight line indicates  $\lambda_{\max} \approx 0.014$  and the convergence of the Lyapunov exponent for  $m \geq 8$  is shown. In the same way, the correlation dimension is shown to converge to the same value for  $m \geq 8$ . Parameter values:  $R_t = 10$  and the time window to suppress correlated neighbors has been set equal to 100

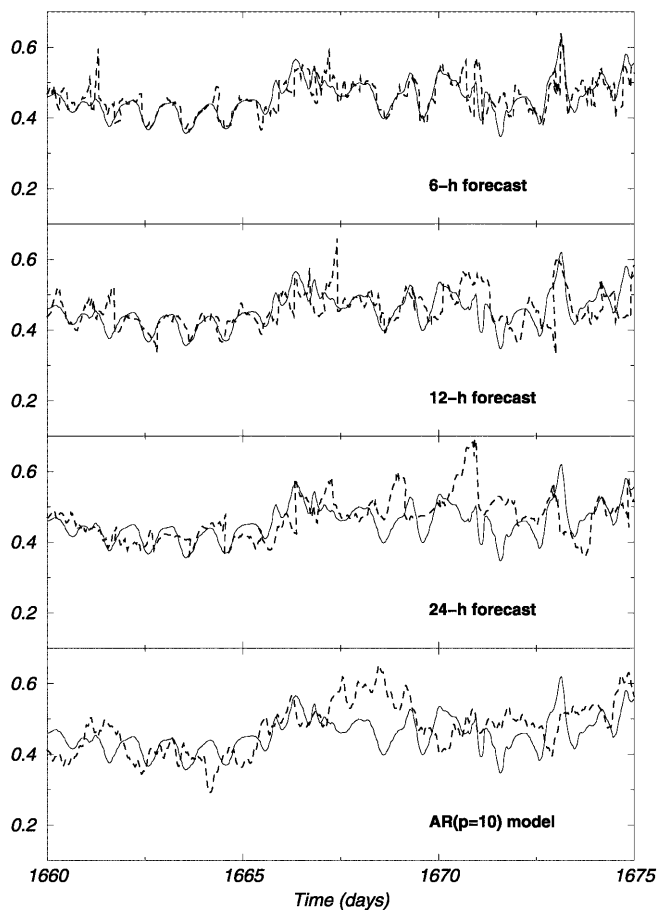
FNN is always zero except when the Theiler window approaches the zero value then, the fraction of FNN remains nearly zero within the interval  $8 \leq m \leq 13$ . In order to calculate the FNN of a given point in the embedded phase space, the distances between phase points of the attractor must be calculated over a set of points (usually the whole time series) drawn independently according to the invariant measure of the attractor. Successive elements of a time series are not usually independent. Theiler suggested in 1990 that, in order to remove this spurious effect, all points whose time indices differ by less than the Theiler-window  $w$  (where  $w$  should be chosen generously) should simply be ignored. If the time delay  $\tau = 83$  days suggested by the autocorrelation statistic is used, a very high dimension would be predicted for the embedding. This is clearly a numerical artifact of the very large  $\tau$  and comes from the intrinsic instabilities in the system which are associated with the presence of chaos.

The final characteristic of the data we determine is the largest Lyapunov exponent  $\lambda_{\max}$  and the correlation dimension  $D_2$  shown in Fig. 3c–d, respectively. For sufficiently small length scales  $\epsilon$  and when the embedding dimension  $m$  exceeds the correlation dimension of the attractor  $D_2$ ,  $C(m, \epsilon) \propto \epsilon^{D_2}$ . Note in Fig. 3d that the slope of  $\log[C(m, \epsilon)]$  versus  $\log(\epsilon)$  tends to converge for  $m$  up to 8, in agreement with previous results obtained with the fraction of FNN. The slope  $D_2(m)$  saturates as  $m$  increases at the value  $D_2 = 2.9$ . On the other hand, according to the embedding theory, the saturation value of the slope in the scaling region, as  $m$  increases, supports the existence of a low-dimensional attractor with fractal dimension  $D_2 = 2.9$  and  $m > 2D_2$  (Sauer *et al.*, 1991).

Chaotic time series taken from experimental systems are always contaminated by some (usually unknown) level of noise, both observational and dynamical. With a finite amount of data, the estimated correlation dimension could be at least equal to that obtained for the original series. We need then to distinguish chaos from noise, or in other words to *detect* a nonlinear structure in the time series by showing that the time series is inconsistent with a linear stochastic process. One important method in this context is the method of surrogate data (Theiler *et al.*, 1992; Bountis *et al.*, 1993). The main idea of this method consists of comparing results, for example the calculated correlation dimension, for the original time series to randomized data subject to a given constraint. Then, surrogate data can be obtained by permuting the data randomly without replacement or, to maintain a given spectrum, by permuting the random phases and taking the inverse Fourier transform. In both cases, for a large number of realizations of the random processes, we observe that the correlation dimension of the surrogate data behaves differently from the original series; it shows no tendency to saturate but continues to increase with increasing  $m$ . This clearly implies that the original time series shown in Fig. 2 corresponds to a nonlinear structure of low dimensionality, even though it contains a small noise component.

The estimated values of  $\lambda_{\max}$  [shown as the slopes of  $S(m, t)$  versus  $t$  in Fig. 3c] also converge for  $m > 8$  to the value  $\lambda_{\max} \approx 0.014$  in inverse units of the sample time equal to half an hour. Since the predictability time is about  $\lambda_{\max}^{-1}$  this means that models for the cloud coverage factor should allow prediction for about 1.5 days from any given time before the intrinsic instabilities of the system mask any such ability.

The analysis of the cloud absorption time series showed the chaotic characteristics of our data, so we were able to construct a predictor map for our series using the method described in Section 2. This method forecasts the daily amount of cloud absorption several hours in advance from the time series shown in Fig. 2. Then, the resulting forecast is used as an input to the meteorological model that is used to calculate the wind, temperature and humidity fields in the area of interest. In the present study, we have selected the last and first weeks of July and August 1999 respectively (from days 1660 through 1675), for research purposes. The first three panels of Fig. 4 show the predicted values of  $\xi$  (dashed line) compared with the measured values



**Fig. 4.** Forecasting values of cloud coverage (*dashed line*) compared with the real time series obtained from the Meteosat (*solid line*). The *three upper panels* correspond to predictions done with the nonlinear model depicted in Sect. 2 when predictions are made 6, 12 or 24 hours in advance, while the *last panel* corresponds to predictions done with the AR model ( $p = 10$ ), Eq. (8). Parameter values for the nonlinear model:  $\tau = 34$ ,  $m = 8$  and  $N_B = 3(m + 1)$  nearest neighbors

obtained from the Meteosat (continuous line) when predictions are made 6, 12 or 24 hours ahead. Most of the large oscillations in the time series were reproduced by our nonlinear forecasting method independently of the forecasting period. Note that for the same time step, the predicted values of  $\zeta$  can change depending on the forecasting period  $T$ , since the coefficients of the model,  $\alpha_0, \dots, \alpha_m$  in Eq. (5), are recalculated at each forecasting step using the new predicted values of the cloud absorption.

The calculated RMSE of the cloud absorption forecasting [Eq. (6)] increases with the number of hours ahead the prediction is needed as expected for a chaotic system. Thus, the histograms shown in Fig. 5 reveal that for predictions made 6 hours ahead, the percentage of predictions whose forecasting error is lower than 0.25 is about 75%, while this value is drastically diminished when the forecasting is made 24 hours ahead. In the following Sections, we will investigate the use of these predictions in a meteorological model for mesoscale weather forecasting.

At this point, it is worth to figure out if the deterministic approach followed above is the most appropriate to explain irregularities in the time series. The traditional approach given by the time series literature is that external random influences may be acting on the system. The external randomness explains the irregularity, while linear dynamical rules may be sufficient to explain the structure found in the sequence. One of the most general linear (univariate) models is the autoregressive AR( $p$ ) process (Box and Jenkins, 1968, 1970), given by,

$$\zeta_n - \bar{\zeta} = \sum_{i=1}^p \alpha_i (\zeta_{n-i} - \bar{\zeta}) + \eta_n \quad (8)$$

where  $\eta_n$  is a purely random process with mean zero and  $\bar{\zeta} = 0.487$  is the mean value of the time series  $\zeta$ . The last panel of Fig. 4 shows the cloud coverage prediction compared with the real one after fitting the first  $N = 73\,000$  data of the time series to the polynomial function<sup>3</sup> given by Eq. (8) with  $p = 10$ . In this case, the obtained RMSE was equal to 1.05, so the forecast was no better than a constant predictor, as happens when using the average value of the time series. Note from the figure, that still the AR model prediction is able to follow the trend of the series but the nonlinear deterministic behavior is lost. Thus, for example, the diurnal cycle in the cloud coverage dynamics is lost when using this model, while this did not happen with the nonlinear model whose results are shown in the first three panels of Fig. 4.

#### 4 Surface temperature model

A non-hydrostatic meteorological model was applied to A Coruña city area (north-west of Spain, near the

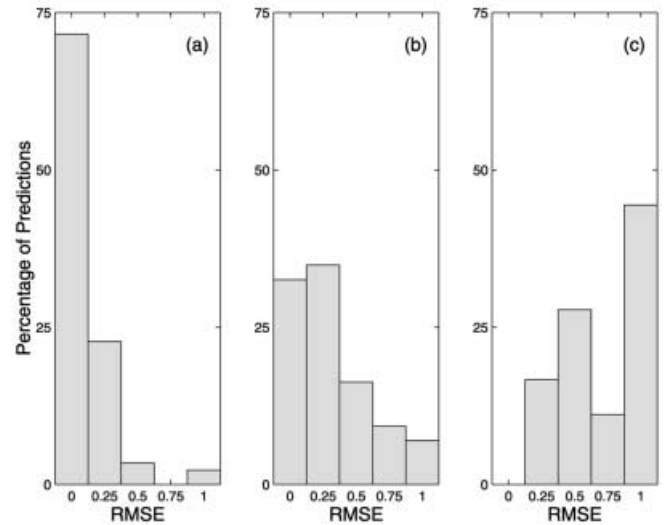
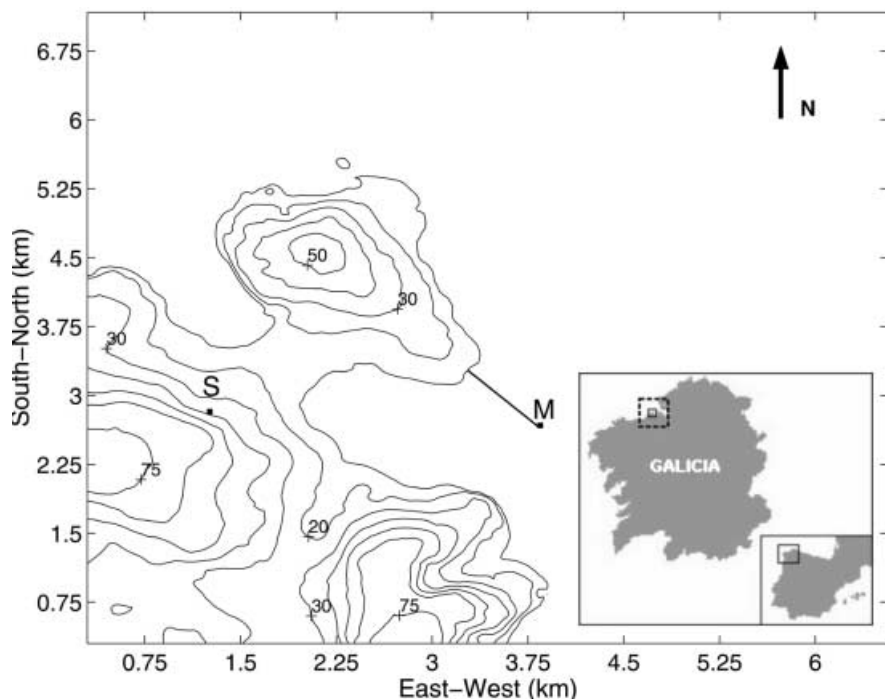


Fig. 5. Histograms of the RMSE when the forecastings are made 6 (a), 12 (b) and 24 (c) hours ahead, for the periods of days shown in Fig. 4

Atlantic Ocean). Figure 6 shows the topography of the area ( $7.0 \text{ km} \times 6.5 \text{ km}$ ). The center of the computational domain is located at  $43^\circ 22' \text{N}$  and  $8^\circ 25' \text{W}$ . The region of interest is mainly influenced by northeastern and southwestern winds. The meteorological model provides three-dimensional fields of wind, temperature, humidity, and turbulence. The mesoscale meteorological model used is a modified version of ARPS (Xue *et al.*, 1995) (Advanced Regional Prediction System) developed by CAPS (Center for Analysis and Prediction of Storms) in Oklahoma (USA). A specific sea-surface physics parameterization (Pérez-Muñuzuri, 1998) has been incorporated in order to achieve a better description of the circulating flows at the coastline. The Durran and Klemp (1983) open boundary condition option is chosen for the lateral boundaries. The 1.5-order Turbulent Kinetic Energy (TKE) closure scheme is used for subgrid-scale turbulence parameterization. The horizontal grid spacing is 300 m and 25 vertical levels following the terrain are used. Vertical model domain extends to 5 km with grid stretching. The minimum  $\Delta z$  used near ground is 3 m. The horizontal grid spacing smooths the complex orography of the city (streets, squares, etc) and prevents the model from using detailed parameterizations for a city environment. Nevertheless, the main air-flow characteristics over this region are preserved under this assumption.

Cloud cover affects the radiation balance, which is a determining factor for a good description of the meteorological variables. Although the microphysics parameterization implemented in ARPS model was turned on, an explicit model for cloud prediction was found to be unsatisfactory due to the small size of the domain and its high resolution. Then, to describe the effect of clouds over the domain, cumulus effects must be parameterized to limit computational costs, so the forecasting model for cloud formation based on nonlinear chaotic predictions described above was coupled

<sup>3</sup> For  $p > 10$  the residual sum of squares does not give better results than for  $p = 10$ .



**Fig. 6.** Topography of the numerical domain corresponding to A Coruña city. Labels *S* and *M* (Dique Barrié de la Maza) correspond to the location of the rawsonde and surface meteorological station from the National Weather Institute and the surface meteorological station within the A Coruña Port facilities, respectively. The *x*- and *y*-axes are oriented in the west-east and south-north directions, respectively

to the ARPS model. At this point, it is interesting to note that the cloud coverage time series was obtained from the Meteosat infrared images for a larger area (dashed-line box in Fig. 6) than the one considered for the ARPS model. Thus, the predicted cloud coverage is an average value of cloud cover inside and outside the boundaries of the model.

The ARPS is initialized at 0Z with radiosonde data provided by the Spanish Meteorological Office obtained at site location *S* in Fig. 6, within the computational area. The computations extended for 24 hours and the time step used in the meteorological model was 5 s.

Soil temperature was initialized uniformly using the surface value given by the sounding and the sea-surface temperature was assumed to be two degrees higher. The initial moisture within the soil was 0.6 times that of the saturated value for each type of soil. The surface temperature and moisture are updated by the time-dependent soil-vegetation model based on Noilhan and Planton (1989) and Pleim and Xiu (1995). In this model, the surface net radiative flux for the ground-air  $R_n^{g \leftrightarrow a}$  (Deardorff, 1978; Xue *et al.*, 1995) and the sea-air  $R_n^{s \leftrightarrow a}$  (Portela and Neves, 1994; Pérez-Muñuzuri, 1998) interfaces are given by,

$$R_n^{g \leftrightarrow a} = (1 + A_1^g \zeta^2(t) + B_1^g \zeta(t)) \times (1 - \alpha_g) \tau_{rg} \tau_{wv} S_0 (a/r)^2 \cos Z + \sigma (\epsilon_a (1 + A_2^g \zeta^2(t) + B_2^g \zeta(t)) T_a^4 - \epsilon_g T_g^4) \quad (9)$$

$$R_n^{s \leftrightarrow a} = (1 + A_1^s \zeta^2(t) + B_1^s \zeta(t)) (1 - \alpha_s) A_S S_0 \cos Z + \sigma (10^{-5} (1 + A_2^s \zeta^2(t) + B_2^s \zeta(t)) T_a^6 - T_s^4) \quad (10)$$

where the first and second terms in both equations account for the shortwave and longwave radiation terms, respectively. Term  $\zeta$  is the sum of the cloud-top

albedo and in-cloud absorption by cloud droplets;  $\zeta$  can change with time and it is obtained from the time series described in Sect. 2. Clouds are assumed to be in one layer at a height which is the average height weighted by the amount of each layer observed from the surface. Table 2 shows the values of parameters  $A_i^{g,s}$  and  $B_i^{g,s}$  ( $i = 1, 2$ ) in Eqs. (9, 10) obtained after fitting the predicted surface temperature with data obtained from a surface meteorological station (dot labeled *M* in Fig. 6) for the most typical meteorological situations occurring in the area of interest.

Finally, the ground- and sea-surface temperatures are calculated by the following equations,

$$\frac{\partial T_g}{\partial t} = C_T^g (R_n^{g \leftrightarrow a} - Q_H^g - Q_E^g) - \frac{2\pi}{Y} (T_g - T_M) \quad (11)$$

$$\frac{\partial T_s}{\partial t} = \frac{C_T^s}{H} (R_n^{s \leftrightarrow a} - Q_H^s - Q_E^s) \quad (12)$$

where the last term in Eq. (11) includes the effect of conduction from the ground below the interface. It avoids the problem of computing temperatures at a number of levels beneath the surface. Parameters  $Q_H^{g,s}$  and  $Q_E^{g,s}$  are the sensible and latent heat fluxes respectively, and parameterizations given by Xue *et al.* (1995) and Pérez-Muñuzuri (1998) were used. Besides Eqs. (11, 12), the soil-vegetation model used by ARPS includes other equations for soil-surface, deep-soil and canopy moistures (Xue *et al.*, 1995).

**Table 2.** Coefficients used in Eqs. (9, 10) for Fig. 8

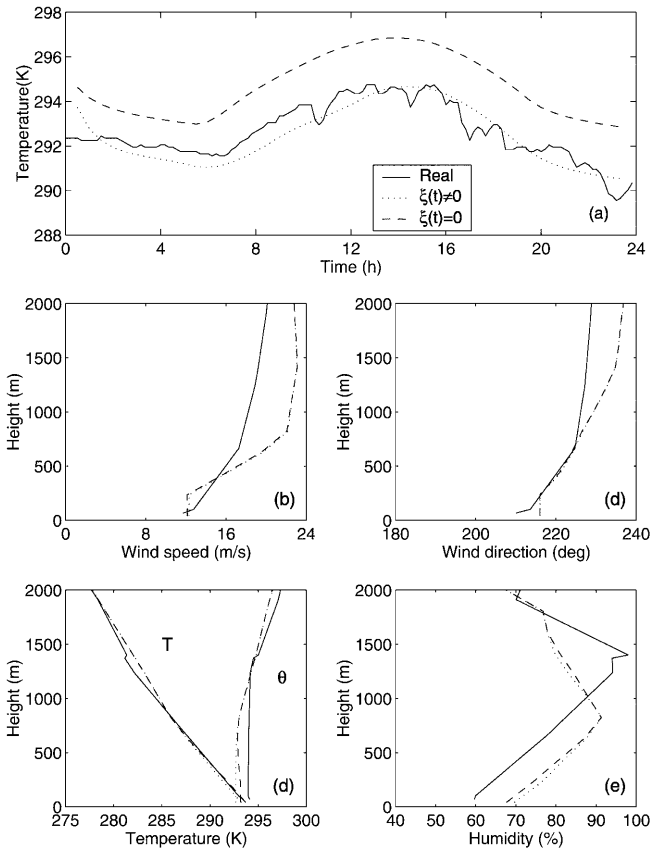
Surface type	$A_1$	$B_1$	$A_2$	$B_2$
Ground	-1	-1	-1	-1
Sea	-0.65	0	0.17	0



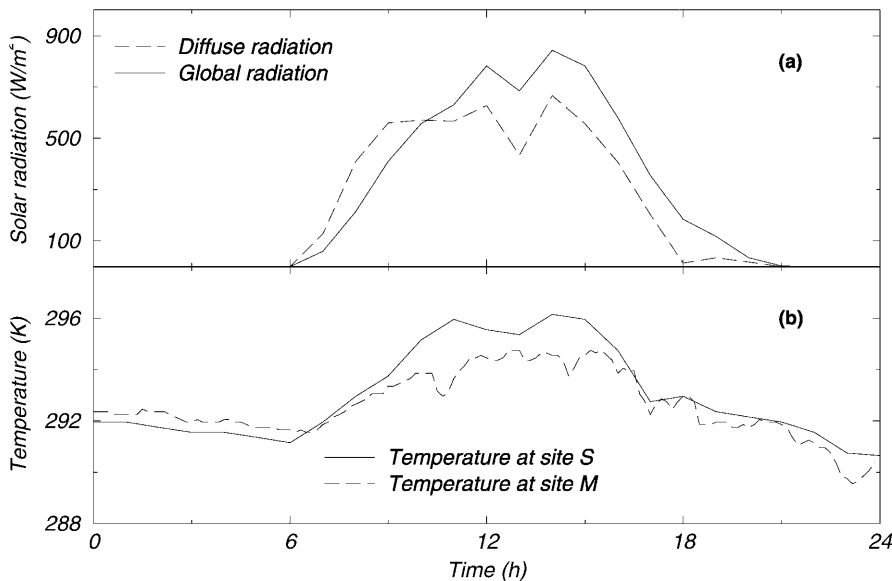
**5 Meteorological results**

In this Section, we have selected August 8th, 1999 from the time series shown in Fig. 4 to avoid redundancy in the presentation of the meteorological results. During this day, moderate temperatures (around 21 °C during all of the day) and overcast conditions with some light rain affected the region of interest. As a consequence of a low-pressure area located near Great Britain, SW winds affected northwestern Spain during all of the day. Figure 7 shows the solar radiation and 2-m height surface temperatures measured at sites *S* and *M* (shown in Fig. 6). Note that during the afternoon, there was a difference of approximately one degree between the sea coast *M* and the city center *S*. Besides, the amount of cloud became more pronounced at approximately 13Z as is shown by the solar radiation and temperature time series. So, the cloud dynamics was not exactly a classical diurnal cycle as could be expected for the summer season, but it showed up some “irregularities” as was the case at noon.

The results of the ARPS model are compared in Fig. 8 with real data obtained from the surface meteorological station *M* and the sounding made at 12Z from point *S* in Fig. 6, respectively. Model results are obtained when the time series of measured cloud coverage data obtained from the Meteosat are taken into account. These results are also compared with those obtained by a different run of the model where the effect of cloud cover is not taken into account [ $\xi(t) = 0$  in Eqs. (9, 10)]. Note the good agreement between our model results and the meteorological data obtained from the surface station when the cloud coverage is taken into account. The model reproduces well the diurnal change in temperature due to the surface heating and cooling which is also shown by the observational data. To describe the accuracy of the method, the standard deviation error (STD) was calculated between the measured 2-m height surface temperature at the



**Fig. 8.** Measured surface temperature at the meteorological station Dique Barrié de la Maza (*solid line*) compared with the predicted temperatures using the cloud cover time series (*dotted line*) and without cloud absorption [ $\xi(t) = 0$ ] (*dashed line*) (a). Vertical profiles, with (*dotted line*) and without (*dashed line*) taking into account the cloud cover dynamics, of wind speed (b), wind direction (c), temperature (both virtual *T* and potential  $\theta$ ) (d) and relative humidity (e) compared with the vertical sounding (*solid line*) performed at 12Z at the *S* position in Fig. 6 for 8 August, 1999



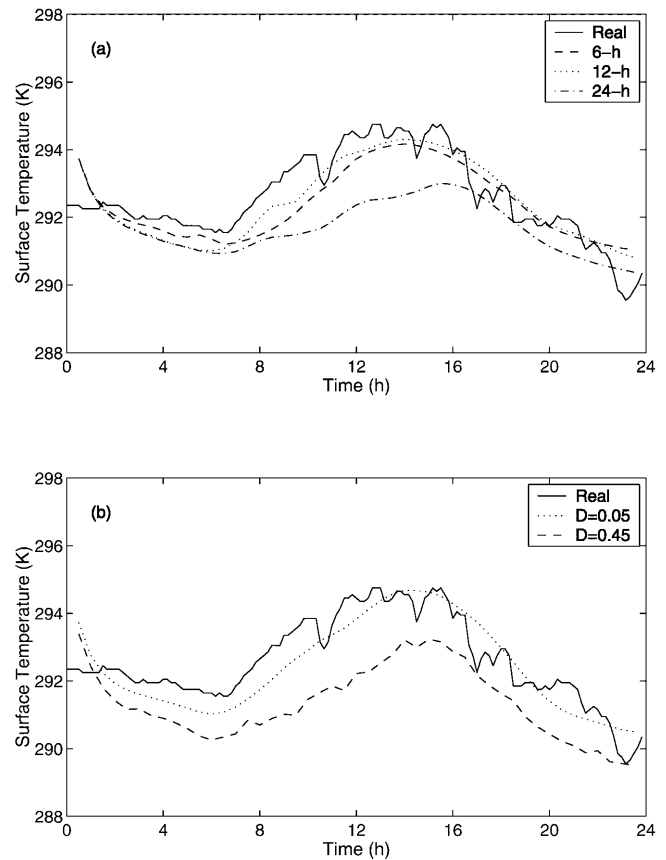
**Fig. 7.** Global and diffuse solar radiation at site *S* (a) and 2-m height surface temperature measured at sites *S* and *M* (b) (shown in Fig. 6)

meteorological station Dique Barrié de la Maza and the predicted one. Thus, when the cloud coverage is considered, the measured error is 0.56, while for the uncovered situation the obtained error is 2.07, nearly four times greater.

Cloud coverage mainly modifies the vertical profiles of temperature and humidity within the mixed boundary layer, while wind speed and direction do not show important variations in their profiles. In any case, the behavior of all the represented profiles Fig. 8b–e does not vary substantially whether cloud coverage is considered or not, and both modeled and measured profiles show a good qualitative agreement. For example, the potential temperature  $\theta$  profile (Fig. 8d) obtained from ARPS shows a mixed layer depth approximately equal to 900 m, which is smaller than the rawinsonde measured depth ( $\approx 1200$  m). The differences between the mixed layer depths observed at the humidity profiles (Fig. 8e) are similar to those seen at the  $\theta$  profile. Thus, in the case of an atmospheric dispersion model coupled to the meteorological model, the behavior of pollutants within the mixed layer depends on whether cloud absorption is considered or not to model the heat budget (Pérez-Muñuzuri, 1998).

In order to test the validity of our nonlinear forecasting method when considered for meteorological modeling, the effect of the forecasting period on the prediction of the surface temperature was analyzed when predictions are made 6, 12 or 24 hours ahead (see Fig. 9a). The following standard deviation errors were obtained for each period: 0.62, 0.55 and 1.19, respectively. In any case, these errors were smaller than those obtained without cloud coverage. Note that for intermediate forecasting periods, the error is smaller than for smaller and greater ones, and approximately equals the value obtained above (Fig. 8a) when the real time series from the Meteosat is used (STD=0.56). Nevertheless, for all the forecasting periods considered the trend of the curve is preserved as the nonlinear prediction of the cloud cover factor  $\xi(t)$  accounts for the 24-h periodicity that is an essential part of the cloud cover and the boundary layer dynamics. In this sense, trying to predict beyond the instability horizon, that is for times greater than  $\lambda_{\max}^{-1} \approx 1.5$  days, our predictions should rapidly lose accuracy and may not even account for the diurnal cycle. This limit could be reduced if, for example, the time series is contaminated by noise. The computations reported above for the 24-h forecasting period support this idea as the RMSE was found to be greater than one.

To examine the effects of noises on the nonlinear forecasting, we performed another experiment consisting of modifying the cloud cover time series used to forecast the results shown in Fig. 8 by adding a white Gaussian noise  $\eta(t)$  with  $\langle \eta(t) \rangle = 0$  and  $\langle \eta(t)\eta(t') \rangle = 2D\delta(t-t')$ , and  $D$  the noise dispersion. In this case, the surface temperature routine in the ARPS model was modified to account for the effect of noise so equations are then solved by using a modified Euler method (García-Ojalvo and Sancho, 1999). Figure 9b shows the forecasted surface temperatures for two different values of the noise dispersion  $D$  in comparison with real data



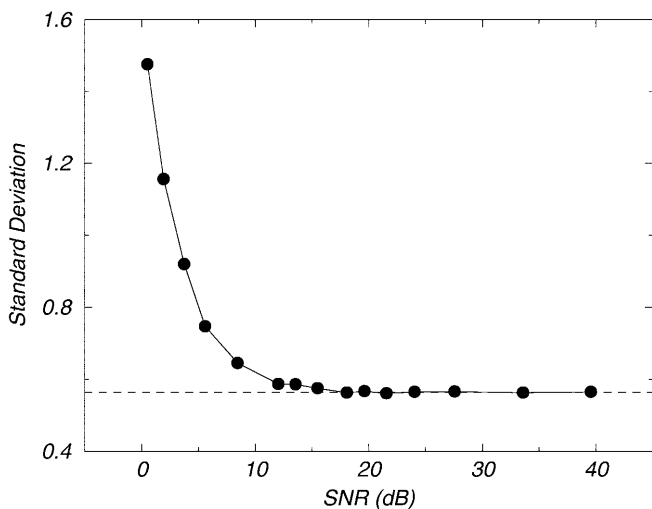
**Fig. 9.** Comparison of the measured surface temperature at the meteorological station Dique Barrié de la Maza (*solid line*) with the predicted temperatures using a 6, 12 and 24 hours cloud coverage forecasting period (a) and using a noisy time series of cloud absorption values with  $D = 0.05$  (SNR = 19.6) and  $D = 0.45$  (SNR = 0.52) (forecasting period  $T = 12$  h) (b) for 8 August, 1999

obtained from the meteorological station within the domain. The obtained values of the standard deviation as a function of the signal-to-noise ratio<sup>4</sup> (SNR) are shown in Fig. 10. For SNR > 10, the STD remains approximately constant and equal to 0.56, while for SNR < 10, the STD increases nearly exponentially as SNR decreases. Thus, even using a noisy time series, for cloud coverage forecasting, the ARPS model is able to reproduce the meteorological conditions above some threshold of the SNR.

## 6 Conclusions

A four-year long time series of cloud absorption values obtained from semi-hourly Meteosat infrared images has been used for meteorological forecasting. Our cloud absorption time series has passed a number of tests for a low-dimensional chaotic attractor which confirm that the presence of noise does not overwhelm the intrinsic nature of its deterministic dynamics. The results presented in this paper support this idea. Clearly, the low-

<sup>4</sup> The SNR is calculated as  $10 \log_{10}(\sum_t \xi^2(t) / \sum_t \eta^2(t))$ .



**Fig. 10.** Standard deviation error (STD) as a function of the signal-to-noise ratio (SNR) for the predicted surface temperatures given in Fig. 9b. The dashed line corresponds to a value of  $\text{STD} = 0.56$ . Forecasting period  $T = 12$  h

frequency part of the power spectrum containing the daily 24-h period plays an important role in the low-dimensional, deterministic appearance of the dynamics, which turns out to be of high importance for its use in meteorological weather forecasting. The coupling between the nonlinear forecasting model and the non-hydrostatic meteorological model ARPS has been very successful in predicting the temporal evolution of the meteorological variables. Furthermore, the extension of the forecasting period and a certain amount of noise added to the cloud coverage time series did not affect the meteorological predictions.

On the other hand, the nonlinear forecasting method presented here reveals the possibilities of short-term predictions of atmospheric parameters whose intrinsic dynamics would make it very expensive, in terms of computational resources, to solve a prognostic equation.

Finally, we wish to emphasize that the entire system of forecasting models described here has been successful in forecasting the most important plume impacts around As Pontes Power Plant (northwestern Spain) since 1995 (Pérez-Muñuzuri, 1998; Souto *et al.*, 1998) and it is now being used in the A Coruña city area to forecast the pollutant dynamics near a harbor.

*Acknowledgements.* We want to thank M.J. Souto for fruitful discussions and E. Canitrot for providing us with data for Fig. 7. This work was partially supported by Norcontrol S.A. and Autoridad Portuaria de A Coruña (Spain), as well as by the ERDF Grant No. 1FD97-0118-C02-01.

Topical Editor J.-P. Duvel thanks V. Casse and another referee for their help in evaluating this paper.

## Appendix: notations

$S_0$	Effective solar constant, $1376 \text{ W m}^{-2}$
$A_S$	Absorption by atmospheric gases
$Z$	Solar zenith angle

$\alpha_{g,s}$	Albedo of ground ( $g$ ) and sea ( $s$ ) surfaces.
$\alpha_s$	$\alpha_s = -0.0139 + 0.0467 \tan Z$
$Q_H^{g,s}$	Sensible heat fluxes for ground ( $g$ ) and sea ( $s$ ) surfaces
$Q_E^{g,s}$	Latent heat fluxes for ground ( $g$ ) and sea ( $s$ ) surfaces
$\sigma$	Stefan-Boltzmann constant, $5.67 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$
$(a/r)^2$	Squared ratio of average distance of the Earth from the Sun to its actual distance at any time of the year
$C_T^{g,s}$	Ground ( $g$ ) and sea ( $s$ ) surface heat capacities
$\tau_{rg}$	Transmittance after Rayleigh scattering and absorption by atmospheric gases
$\tau_{wv}$	Water vapor transmittance
$\epsilon_a$	Emissivity of the atmosphere
$\epsilon_g$	Emissivity of the ground
$H$	Mean depth of the sea coast ( $m$ )
$\Upsilon$	Length of the day, 24 h
$T_M$	Deep soil temperature (K)
$T_a$	Air temperature (usually calculated at the first level of the model) (K)
$T_g$	Ground-surface temperature (K)
$T_s$	Sea-surface temperature (K)

## References

- Abarbanel, H. D. I., *Analysis of observed chaotic data*, Springer-Verlag, Berlin Heidelberg New York Tokyo, 1996.
- Abarbanel, H. D. I., R. Brown, J. J. Sidorowich, and L. Sh. Tsimring, The analysis of observed chaotic data in physical systems, *Rev. Mod. Phys.*, **65**, 1331–1392, 1993.
- Bountis, T., L. Karakatsanis, G. Papaioannou, and G. Pavlos, Determinism and noise in surface temperature time series, *Ann. Geophys.*, **11**, 947–959, 1993.
- Box, G. E. P., and G. M. Jenkins, Some recent advances in forecasting and control, I, *Appl. Stat.*, **17**, 91–109, 1968.
- Box, G. E. P., and G. M. Jenkins, *Time series analysis, forecasting and control*, Holden-Day, San Francisco, 1970.
- Casdagli, M., Nonlinear prediction of chaotic time series, *Physica D*, **35**, 335–356, 1989.
- Deardorff, J. W., Efficient prediction of ground-surface temperature and moisture, with inclusion of a layer of vegetation, *J. Geophys. Res.*, **83**, 1889–1903, 1978.
- Donner, L. J., A cumulus parameterization including mass fluxes, vertical momentum dynamics, and mesoscale effects, *J. Atmos. Sci.*, **50**, 889–906, 1993.
- Durrán, D. R., and J. B. Klemp, The effects of moisture on trapped mountain lee waves, *J. Atmos. Sci.*, **39**, 2152–2158, 1983.
- Farmer, J. D., and G. G. Siderowich, Predicting chaotic time series, *Phys. Rev. Lett.*, **59**, 845–848, 1987.
- Fraser, A. M., and H. L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A*, **33**, 1134–1140, 1986.
- García-Ojalvo, J., and J. M. Sancho, *Noise in spatially extended systems*, Springer-Verlag, Berlin Heidelberg New York Tokyo, 1999.
- Gimson, N. R., Pollution transport by convective clouds in a mesoscale model, *Q.J.R. Meteorol. Soc.*, **123**, 1805–1828, 1997.
- Grassberger, P., and I. Procaccia, Measuring the strangeness of strange attractors, *Physica D*, **9**, 189–208, 1983.
- Grassberger, P., T. Schreiber, and C. Schaffrath, Non-linear time sequence analysis, *Int. J. Bif. Chaos*, **1**, 521–547, 1991.
- Hegger, R., and H. Kantz, Improved false nearest neighbor method to detect determinism in time series data, *Phys. Rev. E*, **60**, 4970–4973, 1999.
- Hegger, R., H. Kantz, and T. Schreiber, Practical implementation of nonlinear time series: the TISEAN package, *Chaos*, **9**, 413–435, 1999.
- Kantz, H., and T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press, Cambridge, 1997.
- Kennel, M. B., R. Brown, and H. D. I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, *Phys. Rev. A*, **45**, 3403–3411, 1992.

- Mölders, N., H. Hass, H. J. Jakobs, M. Laube, and A. Ebel**, Some effects of different cloud parameterizations in a mesoscale model and a chemistry transport model, *J. Appl. Meteorol.*, **33**, 527–545, 1994.
- Noilhan, J., and S. Planton**, A simple parameterization of land surface processes for meteorological models, *Mon. Weather Rev.*, **117**, 536–549, 1989.
- Pérez-Muñuzuri, V.**, Forecasting of chaotic cloud absorption time series for meteorological and plume dispersion modeling, *J. Appl. Meteorol.*, **37**, 1430–1443, 1998.
- Pleim, J. E., and A. Xiu**, Development and testing of a surface flux and planetary boundary layer model for application in mesoscale models, *J. Appl. Meteorol.*, **34**, 16–32, 1995.
- Portela, L. I., and R. Neves**, Modelling temperature distribution in the shallow Tejo estuary, in *Advances in Water Resources Technology and Management*. Eds. G. Tsakiris, M. A. Santos, Balkema, Rotterdam, 457–463, 1994.
- Sauer, T., J. Yorke, and M. Casdagli**, Embedology, *J. Stat. Phys.*, **65**, 575–581, 1991.
- Schreiber, T.**, Interdisciplinary application of nonlinear time series methods, *Phys. Rep.*, **308**, 1–64, 1999.
- Souto, J. A., V. Pérez-Muñuzuri, M. DeCastro, M. J. Souto, J. J. Casares, and T. Lucas**, Forecasting and diagnostic analysis of plume transport around a power plant, *J. Appl. Meteorol.*, **37**, 1068–1083, 1998.
- Takens, F.**, *Detecting strange attractors in turbulence*, (Lecture notes in mathematics, vol. 898) Springer-Verlag, Berlin Heidelberg New York Tokyo, 1981.
- Theiler, J.**, Estimating fractal dimension, *J. Opt. Soc. Am. A*, **7**, 1055–1073, 1990.
- Theiler, J., S. Eubank, A. Longtin, B. Galdrikin, and J. D. Farmer**, Testing for nonlinearity in time series: the method of surrogate data, *Physica D*, **58**, 77–94, 1992.
- Wolf, A., J. B. Swift, H. L. Swinney, and J. Vastano**, Determining Lyapunov exponents from a time series, *Physica D*, **16**, 285–317, 1985.
- Xue, M., K. K. Droegemeier, V. C. Wong, A. Shapiro, and K. Brewster**, *ARPS Version 4.0 Users Guide*, CAPS, Oklahoma, 1995.