Neural Computing and Applications (2019) 31:4331–4340
https://doi.org/10.1007/s00521-018-3345-0

ORIGINAL ARTICLE

Predicting hourly ozone concentrations using wavelets and ARIMA models

Ledys Salazar[1] · Orietta Nicolis[1] · Fabrizio Ruggeri[2] · Jozef Kisel'ák[3,4] · Milan Stehlík[1,4]

Received: 9 October 2017 / Accepted: 4 January 2018 / Published online: 19 January 2018
© The Author(s) 2018. This article is an open access publication

Abstract
In recent years, air pollution has been a major concern for its implications on human health. Specifically, ozone (O_3) pollution is causing common respiratory diseases. In this paper, we illustrate the process of modeling and prediction hourly O_3 pollution measurements using wavelet transforms. We split the time series of O_3 in daily intervals and estimate scale and wavelet coefficients for each interval by the discrete wavelet transform (DWT) with Haar filter. Subsequently we apply cumulated autoregressive integrated moving average (ARIMA) to estimate the coefficients and forecast their evolution in future intervals. Then the inverse discrete wavelet transform is implemented for the reconstruction of the time series and the forecast in the near future. In order to assess the performance of the proposed methodology, we compare the predictions obtained by the DWT–ARIMA with those obtained by the ARIMA model. Several theoretical results are shown through a simulation study.

Keywords Ozone (O_3) · Discrete wavelet transform (DWT) · Haar wavelet · Autoregressive integrated moving average (ARIMA)

Mathematics Subject Classification 62M10 · 42C40 · 91B76

1 Introduction

The atmospheric pollution is a great concern for many countries of the world, as a result of studies that have verified the negative effects on human health. Santiago de Chile is one of the most polluted cities in the world due to its geographical location and emission sources.

In particular, the ozone O_3 is a highly reactive gas that can irritate lungs and also cause eye, nose and throat irritation. Long exposure to ozone pollutant can seriously worsen these effects. In Chile, the O_3 concentrations are continuously recorded by the national air quality information system (SINCA) in order to check the critical levels and take preventive measures. For this reason, it is very important to develop methods which can provide a good predictions.

Many approaches have been proposed in the literature; most of these are based on the use of neural networks and ARIMA models. Duenas et al. [1] propose ARIMA models for estimating the ground-level ozone concentrations in air at an urban and rural sampling points in Southeastern Spain, [2] analyzes hourly ozone concentrations with

✉ Milan Stehlík
milan.stehlik@jku.at

Ledys Salazar
ledys.salazar@uv.cl

Orietta Nicolis
orietta.nicolis@uv.cl

Fabrizio Ruggeri
fabrizio@mi.imati.cnr.it

Jozef Kisel'ák
jozef.kiselak@upjs.sk

[1] Institute of Statistics, University of Valparaíso, Av. Gran Bretaña, 1111, Valparaíso, Chile

[2] CNR IMATI, Via Alfonso Corti, 12, 20133 Milano, Italy

[3] Institute of Mathematics, Faculty of Science, P. J. Šafárik University in Košice, Košice, Slovakia

[4] Institute of Applied Statistics and Linz Institute of Technology, Johannes Kepler University Linz, Linz, Austria

multiple regression and multilayer perceptron models on observations of an urban area of the east coast of the Iberian Peninsula and [3] use ARIMA models for forecasting daily maximum surface ozone concentrations at the airport in Brunei Darussalam. A neural network approach has been used by [4–8] for predicting the intraday ozone concentrations. A comparison of neural network models with ARIMA and regression models has been provided by [9] for prediction of Houston's daily maximum ozone concentrations.

In this work, we propose a new methodology for predicting the ozone concentrations in Santiago de Chile given by the combination of the wavelet analysis with the accumulated ARIMA approach. In particular, we apply the wavelet transform using the Haar cumulated wavelet function to the original time series; then, we estimate an ARIMA model using the transformed series. Finally, the inverse wavelet transform is applied to the predicted wavelet coefficients obtained by the estimated ARIMA model. Thus the reconstructed time series should represent the future cumulated behavior of the ozone pollutant.

Wavelet analysis is a recent tool which allows to decompose a time series into time–frequency space (see [10]). The multiresolution analysis for discrete wavelet transform proposed by [11] permits to reconstruct a signal until a certain level of resolution. Although the wavelet analysis is a very efficient tool that can be used in several applications (time series analysis, signal processing, image processing, etc.), few works deal with time series forecasting. Soltani [12] proposed to combine the wavelet approach with the artificial neural networks (ANN) for predicting the sun spot and MacKey–Glass time series. Mabrouk et al. [13] used the autoregressive models and the wavelet decomposition for forecasting the sun spot time series. The idea of the latter works is to predict time series by applying linear (ARMAX) or nonlinear (ANN) models on simplified signal given by the reconstructed time series in each level of resolution. From the point of view of this approach, our work is novel since we propose to apply linear models to the scaling and wavelet coefficients obtained by the discrete wavelet decomposition using the wavelet Haar filter.

The paper is organized as follows. First we introduce some basic concepts about the Haar wavelet and the discrete wavelet transform (Sect. 2); then, we show theoretical results about the AR(1) process and the wavelet coefficients (Sect. 3). A simulation study is implemented in Sect. 4 to show the ability of the wavelet approach to estimate the AR(1) parameters. Results and conclusions are reported in Sects. 5 and 6, respectively.

2 Haar wavelet and discrete wavelet transform

2.1 Haar wavelet

The development of wavelets connects with the work of Alfred Haar in the early twentieth century. According to [14], the Haar mother wavelet is a mathematical function defined by

$$\psi(x) = \begin{cases} 1, & x \in \left[0, \frac{1}{2}\right), \\ -1, & x \in \left[\frac{1}{2}, 1\right), \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Two relevant characteristics are the oscillation (the Haar wavelet "goes up and down"; more mathematically this can be expressed by the condition that $\int_{-\infty}^{\infty} \psi(x)\mathrm{d}x = 0$, a property shared by all wavelets) and the compact support (not all wavelets have compact support, but they must decay to zero rapidly). Hence, wavelets are objects that oscillate but decay fast and hence are "little." Haar wavelets are defined over the [0, 1] interval, but we will consider their shifted versions on the intervals $[i-1, i]$, $i = 1, \ldots, M$. The application of the Haar wavelets in modeling continuous functions stems from the result that any continuous function $f \in \mathcal{C}([0, 1])$ can be approximated as

$$f(x) \approx f_n(x) = \langle \xi_0, f\rangle \xi_0(x) + \langle \xi_1, f\rangle \xi_1(x) + \cdots + \langle \xi_n, f\rangle \xi_n(x), \tag{2}$$

with $\langle \xi_i, f\rangle = \int \xi_i(x) f(x)\mathrm{d}x$,

$$\xi_0(x) = 1\{0 \le x \le 1\} \tag{3}$$

$$\xi_1(x) = 1\{0 \le x \le 1/2\} - 1\{1/2 \le x \le 1\} \tag{4}$$

$$\xi_2(x) = \sqrt{2}[1\{0 \le x \le 1/4\} - 1\{1/4 \le x \le 1/2\}] \tag{5}$$

$$\vdots$$

$$\xi_n(x) = 2^{j/2}[1I_k - 1J_k] \tag{6}$$

where $I_k := \{k2^{-j} \le x \le (k+1/2)2^{-j}\}, J_k := \{(k+1/2)2^{-j} \le x \le (k+1)2^{-j}\}$. We take $n = 2^j + k$, $j \ge 0, 0 \le k \le 2^j - 1$ and it can be seen that $\xi_n(x) = \xi_{jk}(x) = 2^{j/2}\xi_1(2^j x - k)$. In this paper, we will use $\xi_n(x)$ for simplicity of notation. It is worth observing that $\xi_0(x)$ (*scaling function*) describes the average behavior of the function $f(x)$, whereas the functions $\xi_n(x), n > 0$ represent the details of $f(x)$. The function $\xi_1(x)$ is the *wavelet*.

We consider the interval $[i-1,i)$, $1 \le i \le M$ and the Haar wavelets up to the $(N-1)$th level. Here we use $j=0$ to denote the scaling function, whereas $j=1,\ldots,N$ denotes the wavelet function at level $j-1$. Given any $x \in [i-1,i)$, there is just one wavelet at each level of containing point x. Therefore, we can write

$$f(x) = \sum_{j=0}^{N} d_j^{(i)}(x) I_j^{(i)}(x), \tag{7}$$

where $d_0^{(i)}(x)$ is the coefficient of the scaling function, $I_0^{(i)}(x)$ is the set function of $[i-1,i)$, $I_j^{(i)}(x), j=1,\ldots,M$, is the set function of the interval x belongs to, at the $(j-1)$th level, and $d_j^{(i)}(x)$ is the corresponding wavelet coefficient multiplied by $2^{(j-1)/2}$ (here we have considered this factor, part of the definition of the wavelet at $(j-1)$th level, within the coefficient).

2.2 Haar Discrete Wavelet Transformation (HDWT)

For more details about general discrete wavelet theory, see [15]. Here we look for two vectors u and v such that the set of their translates by even integers forms an orthonormal basis.

Definition 21 Suppose N is an even integer, say $N=2M$ for some $M \in \mathbb{N}$. An orthonormal basis for $\ell^2(\mathbb{Z}_N)$ of the form $\mathcal{H} = \{R_{2k}u\}_{k=0}^{M-1} \cup \{R_{2k}v\}_{k=0}^{M-1}$ for some $u, v \in \ell^2(\mathbb{Z}_N)$ is called a first-stage wavelet basis for $\ell^2(\mathbb{Z}_N)$. We call u and v the generators of the first-stage wavelet basis. We sometimes also call u the father wavelet and v the mother wavelet.

There is a simple condition in terms of the DFT that characterizes orthonormal bases; see [15]. Here $\hat{w}_n := \sum_{m=0}^{N-1} w_m \, \mathrm{e}^{-\frac{2\pi m n}{N}}$.

Lemma 21 *Let $\mathbf{w} \in \ell^2(\mathbb{Z}_N)$. Then $\{R_k \mathbf{w}\}_{k=0}^{N-1}$ is an orthonormal basis for $\ell^2(\mathbb{Z}_N)$ if and only if $|\hat{w}_n| = 1$ for all $n \in \mathbb{Z}_N$.*

3 Estimation and transformation of AR(1) models

Suppose that the data are observed at discrete time points t_i and are corrupted by noise, so that the observed data follow AR(1) model,[1] i.e.,

$$x_t = \varphi x_{t-1} + \varepsilon_t, \tag{8}$$

where "near zero" x_0 is given, ε_t is a Gaussian white noise process with diagonal variances σ_ε^2 and φ is a parameter, which has to be estimated. Notice that we admit only such φ for which $|\varphi|<1$ holds (it implies that x_t is stationary and ergodic). It is well known that $\mathbb{E}[x_t]=0$ and $\mathbb{V}ar[x_t] = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$.

Let

$$u = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, 0, \ldots, 0\right) \quad \text{and} \quad v = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0, 0, \ldots, 0\right).$$

Then, the first-stage Haar basis for $\ell^2(\mathbb{Z}_N)$ is $\mathcal{H}$ and one can trivially check (one can simply verify Lemma 21) that this wavelet basis is an orthonormal basis. Define the Haar transform matrix as $\mathbf{H}_{2N}$ as

$$(H_{jk}) = \begin{cases} (R_{2j}u)_k, & j = 0,\ldots,N-1 \\ (R_{2j}v)_k, & j = N,\ldots,2N-1 \end{cases} \tag{9}$$

Suppose that observations $\mathbf{x} \in \mathbb{Z}_{2N+1}$ follow model (8). The Haar matrix is in fact of $2N \times 2N$ order with the shortened form

$$\mathbf{H}_{2N} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_N \otimes [1,1] \\ \mathbf{I}_N \otimes [1,-1] \end{bmatrix}, \tag{10}$$

where $\mathbf{I}_N$ is the identity matrix and $\otimes$ denotes Kronecker product. After obtaining the measurements $\mathbf{x}$ for model (8), we transform the data by HDWT

$$\mathbf{y} = \mathbf{H}_{2N}\mathbf{x}. \tag{11}$$

Notice that we can shift N in both directions sufficiently far.

Lemma 31 *HDWT (11) transforms model (8) into*

$$y_t = \varphi^2 y_{t-1} + \sqrt{1+\varphi^2}\,\varepsilon_t \tag{12}$$

with $\mathbb{E}[y_t]=0$ and $\mathbb{V}ar[y_t] = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$.

Proof First of all, it is well known that an affine transformation of $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e., $\mathbf{y} = \mathbf{c} + \mathbf{B}\mathbf{x}$ has a multivariate normal distribution with expected value $\mathbf{c} + \mathbf{B}\boldsymbol{\mu}$ and variance $\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^{\mathrm{T}}$. Therefore, $\mathbf{H}_{2N}\boldsymbol{\varepsilon}_{2N}^0$ has a multivariate normal distribution with expected value $\mathbf{0}$ and variance $\mathbf{H}_{2N}(\sigma_\varepsilon^2 \mathbf{I}_{2N})\mathbf{H}_{2N}^{\mathrm{T}} = \sigma_\varepsilon^2 \mathbf{I}_{2N}$ due to orthogonality of $\mathbf{H}_{2N}$. Thus it is sufficient to transform values x_t. One can see that

$$y_t = \begin{cases} \frac{\varphi}{\sqrt{2}}(x_{2t-2} + x_{2t-1}) + \varepsilon_t, & \text{for } t = 1,\ldots,N, \\ \frac{\varphi}{\sqrt{2}}(x_{2t-2} - x_{2t-1}) + \varepsilon_t, & \text{for } t = N+1,\ldots,2N. \end{cases} \tag{13}$$

From (8), we have

[1] The well-known Ornstein–Uhlenbeck process can also be considered as the continuous-time analogue of the discrete-time AR(1) process.

$$y_t = \begin{cases} \frac{\varphi}{\sqrt{2}}(\varphi x_{2t-3} + \varepsilon_{2t-2} + \varphi x_{2t-2} + \varepsilon_{2t-1}) + \varepsilon_t, & \text{for } t = 1,\ldots,N, \\ \frac{\varphi}{\sqrt{2}}(\varphi x_{2t-3} + \varepsilon_{2t-2} - \varphi x_{2t-2} - \varepsilon_{2t-1}) + \varepsilon_t, & \text{for } t = N+1,\ldots,2N, \end{cases} \tag{14}$$

On the other hand, we know from (11) that

$$y_{t-1} = \begin{cases} \frac{\varphi}{\sqrt{2}}(x_{2t-3} + x_{2t-2}), & \text{for } t = 2,\ldots,N, \\ \frac{\varphi}{\sqrt{2}}(x_{2t-3} - x_{2t-2}), & \text{for } t = N+1,\ldots,2N, \end{cases} \tag{15}$$

which implies

$$y_t = \begin{cases} \varphi^2 y_{t-1} + \frac{\varphi}{\sqrt{2}}(\varepsilon_{2t-2} + \varepsilon_{2t-1}) + \varepsilon_t, & \text{for } t = 2,\ldots,N, \\ \varphi^2 y_{t-1} + \frac{\varphi}{\sqrt{2}}(\varepsilon_{2t-2} - \varepsilon_{2t-1}) + \varepsilon_t, & \text{for } t = N+1,\ldots,2N. \end{cases} \tag{16}$$

Thus (12) is proved.

Now, since $\varphi^2 < 1$ one can show that $\{y_t\}$ is stationary and ergodic. Therefore, $\mathbb{E}[y_t] = 0$ and from $y_t = \sqrt{1+\varphi^2}(\varepsilon_t + \varphi^2\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \ldots)$ we have

$$\mathbb{V}ar[y_t] = \mathbb{E}[y_t^2] = (1+\varphi^2)\mathbb{E}[(\varepsilon_t + \varphi^2\varepsilon_{t-1} + \varphi^4\varepsilon_{t-2}\ldots)^2] = \frac{1+\varphi^2}{1-\varphi^4}\sigma_\varepsilon.$$

The proof is complete. □

Notice that we do not know whether the distributions of x_t and y_t coincide. We only know that first two moments are equal and we can assume that errors stay unchanged in the sense of the next model. Therefore, we also assume model

$$y_t = \varphi^2 y_{t-1} + \varepsilon_t \tag{17}$$

(conditional) MLE estimation of φ for model (8) with $t \in \{1,\ldots,N\}$ is

$$\hat{\varphi} = \frac{\sum_{t=2}^{N} x_{t-1}\, x_t}{\sum_{t=2}^{T} x_{t-1}^2}$$

(notice here that the conditional MLE of φ is also the OLS estimator of φ, but this does not hold for the exact MLE). This implies that for model (12) with $t \in \{1,\ldots,N\}$ we have

$$\hat{\varphi}^2 = \frac{\sum_{t=2}^{N} y_{t-1}\, y_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

if we neglect the term $\sqrt{1+\varphi^2} \approx 1$. It is important to note that for a sample of size N in order to estimate an AR(1) process by conditional MLE, we will use only $N-1$ observations of this sample.

Then, we transform the data $\hat{\mathbf{y}}$ given by model (12) with estimated $\hat{\varphi}$ back by IHDWT to obtain

Fig. 1 Transformations and estimation

$$\hat{\mathbf{x}} = \mathbf{H}_{2N}^{-1}\hat{\mathbf{y}} = \mathbf{H}_{2N}^{T}\hat{\mathbf{y}},$$

since matrix $\mathbf{H}_{2N}$ is orthogonal. The following diagram summarizes the main idea (Fig. 1).

Here we define the MSE of an estimator $\hat{\theta}$ with respect to an unknown parameter θ in a standard way as $MSE(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta}-\theta)^2\right]$.

These theoretical results can be extended to other linear processes such as the ARMA (p,q) and ARIMA (p,d,q) (see [16] for details on the autocorrelation functions, estimation and forecasting of such models).

4 Simulation study

We have calculated estimates for different parameter values ϕ and x_0 (for fixed value of σ_ε) for data generated by model (8) in the following way. For each realization, we have computed (conditional) MLE for the original data (here we have used fixed times $t \in \{1,\ldots,N\}, N = 100$), which has been repeated K times. Afterward average value $\hat{\varphi}_{\mathbf{x}}$ and estimation of MSE$(\hat{\varphi})$ as average value of $\left[(\hat{\varphi}-\varphi)^2\right]$ were computed (i.e., computing empirical version of MSE$(\hat{\varphi})$). The same procedure has been done using transformed data $\mathbf{y}$.

Here we have used the assumption that errors stay unchanged. Therefore, we have used (conditional) MLE for model (17). We have obtained estimates $\hat{\varphi}_{\mathbf{y}}$. Tables 1, 2, 3 and 4 show achieved results.

From the results, it is obvious that for large number[2] of repetitions the estimate $\hat{\varphi}_{\mathbf{x}}$ is better in average than the estimate $\hat{\varphi}_{\mathbf{y}}$. Nevertheless, when the number of repetitions, e.g., $K = 20$ in our case, is small, the (conditional) MLE estimation $\hat{\varphi}_{\mathbf{y}}$ has better behavior; see Tables 3,4 and values of "average" MSE. This results suggest that in practice might be helpful to use transformed data in order to obtain an estimation of specific parameter, especially when we realize that we have typically only one realization measured.

On the other hand, if $\hat{\mathbf{x}}$ is a vector of predictions, and $\mathbf{x}$ is the vector of observed values corresponding to the inputs to the function which generated the predictions, then the MSE of the predictor can be estimated by MSE $= \frac{1}{n}\sum_{t=1}^{n}(\hat{x_t} - x_t)^2$. We have used the mean squared

[2] We have used $K = 1000$, but probably smaller value could be eligible.

Table 1 Estimation of parameter φ, $\sigma_\varepsilon = 1, K = 1000, N = 100, x_0 = 1$

$x_0 = 1$	$\varphi = 0.99$	MSE	$\varphi = 0.8$	MSE	$\varphi = 0.7$	MSE
Average $\hat{\varphi}_{\mathbf{x}}$	0.969821	1.5819	0.786203	4.3048	0.687516	5.6160
Average $\hat{\varphi}_{\mathbf{y}}$	0.954043	2.5735	0.775157	5.6511	0.676731	7.7051

Table 2 Estimation of parameter φ, $\sigma_\varepsilon = 1, K = 1000, N = 100, x_0 = 0.1$

$x_0 = 0.1$	$\varphi = 0.99$	MSE	$\varphi = 0.8$	MSE	$\varphi = 0.7$	MSE
Average $\hat{\varphi}_{\mathbf{x}}$	0.971516	1.4638	0.782047	4.2449	0.687871	5.3416
Average $\hat{\varphi}_{\mathbf{y}}$	0.955874	2.3352	0.772372	5.4061	0.679346	7.5531

Table 3 Estimation of parameter φ, $\sigma_\varepsilon = 1, K = 20, N = 100, x_0 = 1$

$x_0 = 1$	$\varphi = 0.99$	MSE	$\varphi = 0.8$	MSE	$\varphi = 0.7$	MSE
Average $\hat{\varphi}_{\mathbf{x}}$	0.968205	0.0368	0.803024	0.0741	0.672014	0.1725
Average $\hat{\varphi}_{\mathbf{y}}$	0.960740	0.0396	0.802245	0.0729	0.675860	0.15026

Table 4 Estimation of parameter φ, $\sigma_\varepsilon = 1, K = 20, N = 100, x_0 = 0.1$

$x_0 = 0.1$	$\varphi = 0.99$	MSE	$\varphi = 0.8$	MSE	$\varphi = 0.7$	MSE
Average $\hat{\varphi}_{\mathbf{x}}$	0.969563	0.0174	0.794335	0.0658	0.699146	0.0858
Average $\hat{\varphi}_{\mathbf{y}}$	0.960106	0.0289	0.788669	0.0827	0.700260	0.0693

prediction error, which measures the expected squared distance between what our predictor predicts for a specific value and what the true value is: $\mathrm{MSPE(L)} = \mathbb{E}\left[\sum_{i=1}^{n}(x_i - \hat{x}_i)^2\right]$. See results in Tables 5, 6, where two types of errors were computed. Clearly the inversion of $\mathbf{H}_{2N}$ complicates whole situation. Nevertheless, we have computed mean variances for specific models (8), (12) and (17); see Table 7. Surprisingly the nearest estimation has model (17).

5 Application to Ozone data

In this section, we apply the HDWT to the ozone time series; then, we estimate ARIMA model on the resulting coefficients (detail and scaling coefficients); finally, we apply the IHDWT on the predicted coefficients by ARIMA models in order to reconstruct the future behavior of the considered cumulated time series. Description of used models follows in the next subsection.

5.1 Models

5.1.1 Autoregressive AR(p)

According to [17], an autoregressive model of order p, abbreviated $AR(p)$, has the form

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t, \tag{18}$$

where x_t is stationary, and $\phi_0, \phi_1, \phi_2, \ldots, \phi_p$ are constants ($\phi_p \neq 0$). Although it is not necessary, we assume for the sake of simplicity that ε_t is a Gaussian white noise with zero mean and variance σ_w^2, unless otherwise stated.

Table 5 MSPE for $\sigma_\varepsilon = 1, K = 20, N = 100$ and ℓ_2 error

ℓ_2 error	$\varphi = 0.7$	x_0	MSPE (8)	MSPE (17)	MSPE (12)
$\hat{\varphi}_{\mathbf{x}} = 0.699146$	$\hat{\varphi}_{\mathbf{y}} = 0.700260$	1	0.0007	319.2027	383.8687
$\hat{\varphi}_{\mathbf{x}} = 0.672014$	$\hat{\varphi}_{\mathbf{y}} = 0.675860$	0.1	0.7611	319.3887	376.3713

Table 6 MSPE for $\sigma_\varepsilon = 1, K = 20, N = 100$ and ℓ_∞ error

ℓ_∞ error	$\varphi = 0.7$	x_0	MSPE (8)	MSPE (17)	MSPE (12)
$\hat{\varphi}_{\mathbf{x}} = 0.699146$	$\hat{\varphi}_{\mathbf{y}} = 0.700260$	1	0.0067	4.9695	5.4833
$\hat{\varphi}_{\mathbf{x}} = 0.672014$	$\hat{\varphi}_{\mathbf{y}} = 0.675860$	0.1	0.2129	4.8487	5.2666

Table 7 Estimation of variance for $\sigma_\varepsilon = 1, K = 1000, N = 100$

$\mathbb{V}ar[x_t] = \frac{\sigma_\varepsilon^2}{1-\varphi^2}$	est. for (8)	est. for (17)	est. for (12)
1.9607	1.8192	1.9182	1.2874

5.1.2 Moving average model MA(q)

According to [17], a moving average model of order q, abbreviated $MA(q)$, has the form

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \ldots + \theta_q w_{t-q}, \tag{19}$$

where there are q lags in the moving average and $\theta_1, \theta_2, \ldots \theta_q$ $(\theta_q \neq 0)$ are parameters. Although it is not necessary yet, we assume that w_t is a Gaussian white noise series with mean zero and variance σ_t^2, unless otherwise stated.

5.1.3 Autoregressive integrated moving average ARIMA(*p,d,q*)

According to [18], in order to make a time series stationary, one must differentiate it d times and then apply an ARMA (p,q) model. Thus the original series is ARIMA (p,d,q), where p denotes the number of autoregressive terms, d is the number of times the series must be differentiated and q is the number of terms of the invertible moving average.

The ARIMA model is given by

$$\begin{aligned} X_t^d = c + \phi_1 X_{t-1}^d, +\ldots, \phi_p X_{t-p}^d + \varepsilon_t \\ + \theta_1 \varepsilon_{t-1}^d, \ldots, \theta_q \varepsilon_{t-q}^d + \varepsilon_t^d \end{aligned} \tag{20}$$

Expressed in terms of the backshift operator, we have

$$\phi(L)(1-L)^d X_t = c + \theta_L \varepsilon_t \tag{21}$$

where X_t^d is the series of differences of order d, ε_t^d is a white noise process, and $c, \phi_0, \phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the model.

In the present article, we use ARMA and ARIMA models on the different vectors of scale coefficients and wavelets of the DWT.

5.2 Data analysis

The data used in this work have been recorded by the SINCA (Chile). In particular, we consider the hourly ozone concentrations (O_3) gathered in the providence monitoring station of the Metropolitan Region of Santiago, Chile, during the period from October, 1, 2002, to December 19, 2002.

The methodology is illustrated in Fig. 2. Firstly the time series from SINCA database are transformed by the DWT using the Haar filter; then, an AR or ARIMA model is identified through the partial and total auto-correlation coefficients (ACF and PACF) for each sequence of wavelet coefficients corresponding to each level of resolution. Then we apply the ARIMA models to the resulting coefficients in order to predict the future coefficients, namely 1, 2, 3, 7 and 9 levels. Finally we implemented the IHDWT on the predicted wavelet coefficients in order to obtain the future concentrations of the O_3 pollutant.

Since the wavelet algorithm works on time series which length is power of two, we extracted 16 hours for each day (from 6:00 to 21:00) (see Fig. 3).

The determination of the level of decomposition is fundamental since it provides the number of vectors with scaling coefficients and wavelets. For example, if we use three levels of decomposition, we obtain three vectors of scale coefficients and three vectors of wavelet coefficients, where the vectors of wavelet coefficients represent the detail of the series (see Fig. 4). For the reconstruction of the original series, the detailed vectors and the first scale in correspondence with the level of decomposition are used.

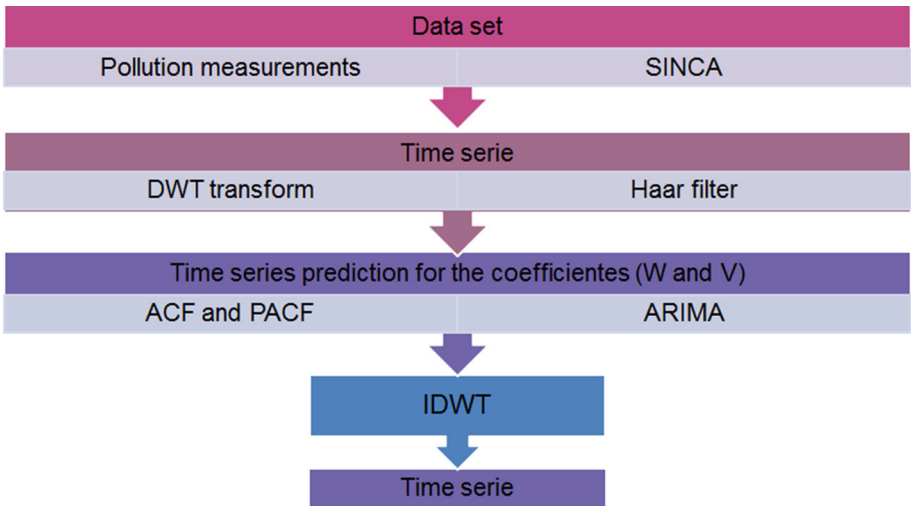

Fig. 2 Methodology

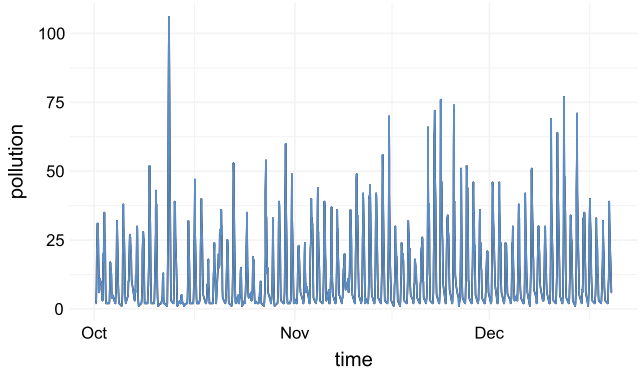

Fig. 3 Hourly Ozone concentrations from October, 1, 2002 to December 19, 2002, from 6:00 to 21:00

Different levels of resolution have been used in the reconstruction step. In particular, to reconstruct the series two levels of the DWT have been applied.

With respect to the two levels, we used the prediction of the coefficients in a cumulative way, that is, step by step, updating the time series and obtaining the predicted values with the ARIMA model in correspondence with the appropriate time interval.

In Figs. 5 and 6, we show the time series with the HDWT for a number of nine and two levels of decomposition, respectively. The first sequence of coefficients represents the scaling coefficients and the others the detailed coefficients of the HDWT. The original time series is at the bottom of Figs. 5 and 6.

Then we performed the prediction of the wavelet coefficients and the reconstruction of the ozone time series for the next 10 days, that is, from December 20, 2002, to December 29, 2002. By using three and seven levels of decomposition, we obtain the prediction represented in Fig. 7.

We did not observe the significant difference between the two reconstructed series. Now let us consider one level

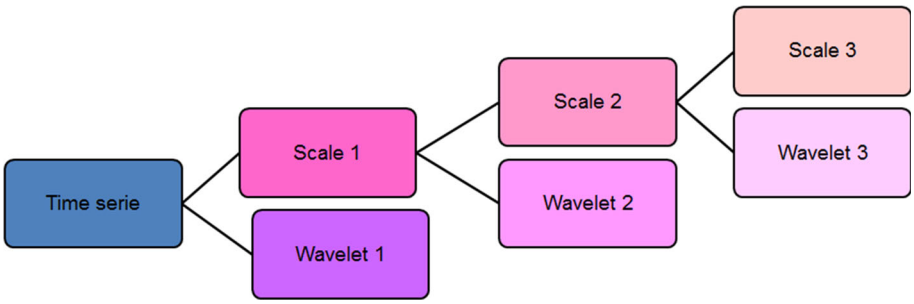

Fig. 4 Wavelet decomposition

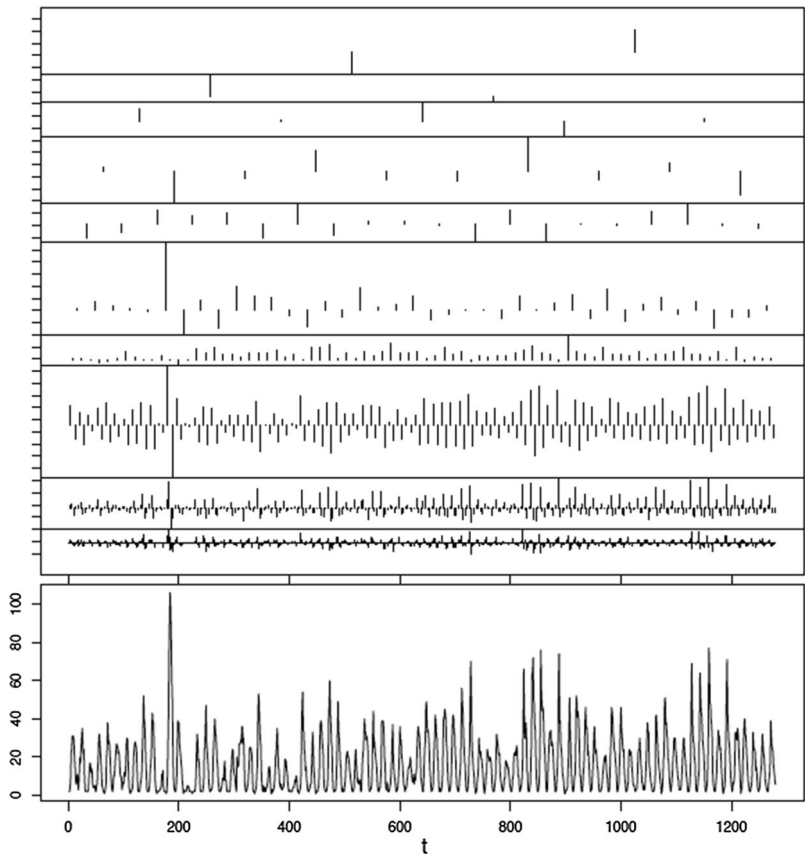

Fig. 5 Nine levels of decomposition

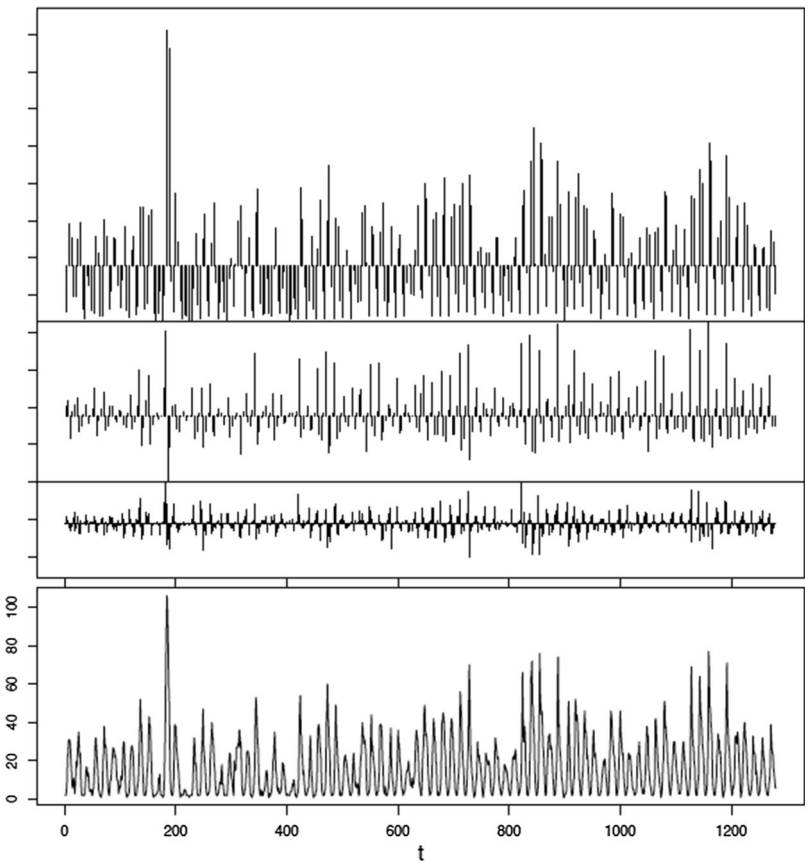

Fig. 6 Two levels of decomposition

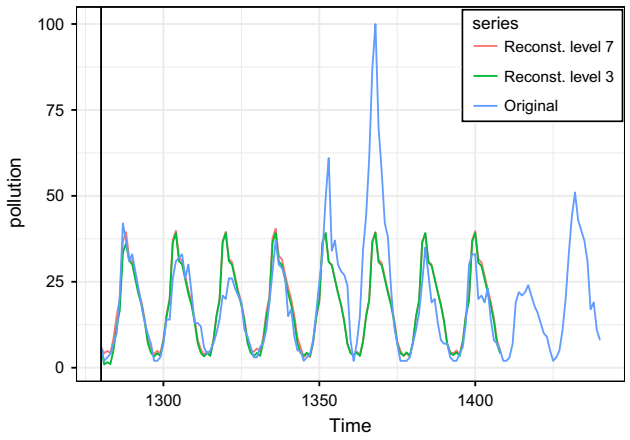

Fig. 7 Ozone prediction using seven and three levels of resolution

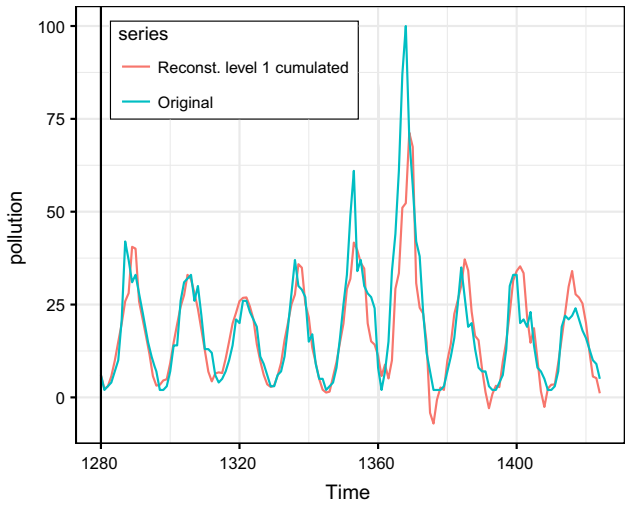

Fig. 8 Ozone prediction using one cumulated level of resolution

in cumulative way, i.e., performing the prediction one step ahead and updating the data at each step. Figure 8 shows that the prediction is much better when using only one accumulated level of resolution.

Figure 9 compares the ozone prediction obtained by applying the ARIMA models on the transformed series (with two and three levels of resolutions) cumulated on the original ozone series. Differently to the classical application of the ARIMA models, the wavelet approach allows to predict the extreme events in cumulated ozone time series. Such extremes are very important from environmental and population health points of view.

Figure 10 plots the reconstructed time series with level 2 of accumulated reconstruction accompanied with the upper and lower limits of the 95 % confidence interval. Such confidence interval is important for further statistical analysis of the series.

6 Conclusion

We shown in this paper that the predictions obtained by proper combination of HDWT and ARIMA models are better than those obtained by the simple application of the ARIMA models to the original data. This is mainly because of complexity and non-linearity intrinsically present in ozone pollution data. Moreover the 95% confidence interval contains all the points of the reconstructed series. Thus the proposed methodology provides a new tool for improving the prediction of the ozone concentrations which could allow the competent authorities to take corrective and preventive actions in order to decrease the

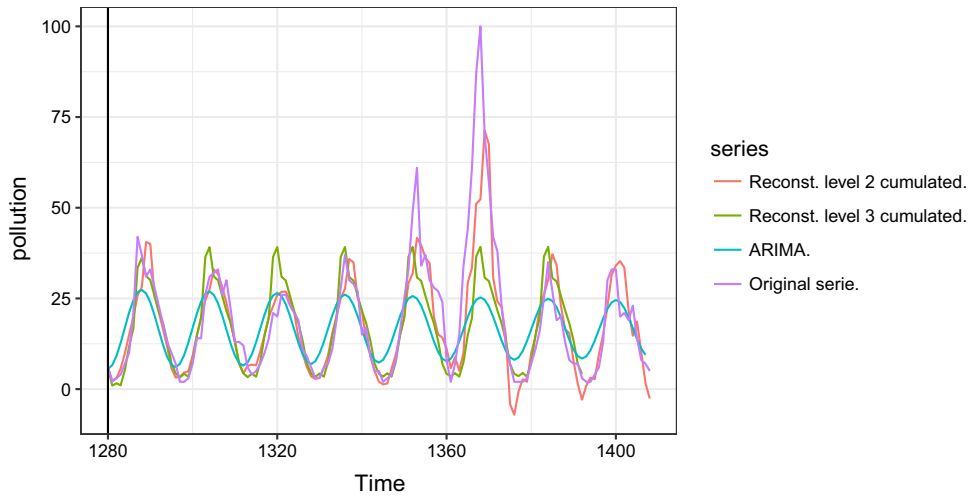

Fig. 9 Application of ARIMA models on original and transformed time series: The purple line illustrates the original time series, the pink line the prediction using two cumulated levels of resolution, the line green using 3 cumulated levels of resolution, and the blue line is the resulting of the ARIMA prediction without the application of the HDWT (color figure online)

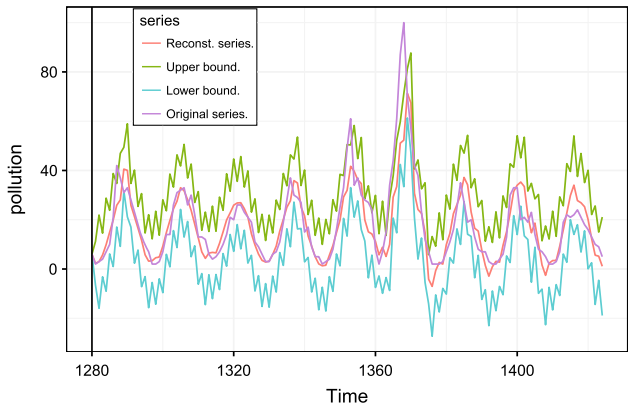

Fig. 10 Predicted ozone time series with its confidence interval and original ozone time series

pollutant levels and consequently reduce the negative impact on the human health.

Acknowledgements Open access funding provided by Johannes Kepler University Linz. We acknowledge LIT-2016-1-SEE-023 project "Modeling complex dependencies: how to make strategic multicriterial decisions?/MODEC" and FONDECYT Regular N1151441. The authors are very grateful to the Editor, Associated Editor and the Reviewers for their valuable comments.

Compliance with ethical standards

Conflict of interest On behalf of all authors, I hereby attest that there are no conflicts of interest regarding financial relationships, intellectual property or any point mentioned under the publishing ethics.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Duenas C, Fernández M, Canete S, Carretero J, Liger E (2005) Chemosphere 61(10):1379. https://doi.org/10.1016/j.chemosphere.2005.04.079. http://www.sciencedirect.com/science/article/pii/S0045653505006533
2. Capilla C (2016) Int J Sustain Dev Plan 11(4):558
3. Kumar K, Yadav A, Singh M, Hassan H, Jain V (2004) J Air Waste Manag Assoc 54(7):809. https://doi.org/10.1080/10473289.2004.10470949
4. Basurko EA, Anta A, Barron LJR, Albizu M (2007) In: Borrego C, Brebbia C (eds) Air pollution XV, WIT transactions on ecology and the environment, vol 101, pp 109–118
5. Coman A, Ionescu A, Candau Y (2008) Environ Model Software 23(12):1407. https://doi.org/10.1016/j.envsoft.2008.04.004. http://www.sciencedirect.com/science/article/pii/S1364815208000650
6. Paoli C, Notton G, Nivet ML, Padovani M, Savelli JL (2011) In: 10th International conference on environment and electrical engineering, pp. 1–4. https://doi.org/10.1109/EEEIC.2011.5874661
7. Faris H, Alkasassbeh M, Rodan A (2014) Pol J Environ Stud 23(2):341
8. Luna A, Paredes M, de Oliveira G, Correa S (2014) Atmos Environ 98(Supplement C):98. https://doi.org/10.1016/j.atmosenv.2014.08.060. http://www.sciencedirect.com/science/article/pii/S1352231014006621
9. Prybutok VR, Yi J, Mitchell D (2000) Eur J Oper Res 122(1):31. https://doi.org/10.1016/S0377-2217(99)00069-7
10. Percival DB, Walden AT (2006) Wavelet methods for time series analysis. Cambridge University Press, Cambridge
11. Mallat SG, Trans IEEE (1989) Pattern Anal Mach Intell 11(7):674. https://doi.org/10.1109/34.192463
12. Soltani S (2002) Neurocomputing 48(1–4):267. https://doi.org/10.1016/S0925-2312(01)00648-8
13. Ben Mabrouk A, Ben Abdallah N, Dhifaoui Z (2008) Appl Math Comput 199(1):334. https://doi.org/10.1016/j.amc.2007.09.067
14. Nason GP (2008) Wavelet methods in statistics with R. Springer, New York
15. Frazier MW (1999) An introduction to wavelets through linear algebra. Springer, New York. https://doi.org/10.1007/b97841
16. Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2016) Time series analysis. Forecasting and control, 5th edn. Wiley, Hoboken
17. Shumway RH, Stoffer DS (2017) Time series analysis and its applications. With R examples, 4th edn. Springer, Cham. https://doi.org/10.1007/978-3-319-52452-8
18. Villavicencio J (2014) Introducción a series de tiempo. http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D