Wavelets Modelling of Water, Energy and Electricity Consumption in Spain

C. GONZÁLEZ-CONCEPCIÓN, M.C. GIL-FARIÑA, C. PESTANO-GABINO

Department of Applied Economics
University of La Laguna (ULL),
Campus de Guajara, s/n, 38071 La Laguna, SPAIN

cogonzal@ull.es, mgil@ull.es, cpestano@ull.es

Abstract: Wavelet techniques possess many desirable properties, some of which are useful in economics and finance, but many of which are not. In this paper we use wavelets to reveal the different cycles that are hidden by the data. So, we consider some of the advantages and drawbacks of this time series multiresolution analysis technique for the purpose of obtaining simultaneous time and frequency information in determining the periodicity and the modelling of the water consumption in Santa Cruz de Tenerife (Canary Islands, Spain), energy and electricity consumption in Spain.

Key-Words: Wavelets, Time series, Scale domain, Water, energy and electricity consumption.

1 Introduction

The literature on wavelets has expanded rapidly over the past 20 years [1, 4…], which have seen many articles and papers published using this methodology in a wide variety of disciplines. Applications using wavelets are extensive in astronomy, engineering, medicine, physics and many other fields of study, including finance and economics in recent years [5, 7, 8, 13…].

Wavelets are certain families of orthogonal or quasi-orthogonal functions that possess many desirable properties, some of which are useful in economics and finance, but many of which are not. For example, in [14], Ramsey and Lampart explore four ways in which wavelets might be used to enhance the empirical toolkit in economics and finance (exploratory analysis, density estimation and local inhomogeneity, time scale decomposition, forecasting). The drawbacks involve the sample characteristics and potential numerical instabilities. Wavelet-based techniques rely on equally-spaced data, a condition which does not always hold in economic series. Even when it does, the cycles over which the economic activity takes place may not be homogeneous with respect to the type of data. Moreover, some techniques based on wavelets require dyadic samples and a certain number of initial values in order to begin the calculation processes. Certain instabilities can also arise when attempting to decompose a signal using very high-order polynomials, due to the requirement to calculate their roots.

The goal of this paper is to illustrate the application of the wavelet technique, bearing the above drawbacks in mind, in the multiresolution analysis for modelling water and energy consumption to obtain time and frequency information simultaneously.

The examples chosen reveal the importance of being able to decompose the original signal into a trend component and various detail components, which allow the different frequencies and their time model to be observed separately.

The type of data involved (daily, weekly, monthly, annual, or other frequency schedules) is vital when drawing conclusions. These can be modelled using wavelet decompositions so as to yield conclusions regarding frequencies that cannot be observed directly in the original series. When highly disaggregated series (data with an hourly or higher frequency) are available, however, it is also of interest to aggregate data in order to obtain another type of periodicity (weekly, monthly, etc.).

The following sections offer a summary of the methodology we employ in the data to establish the contrast among the three domain types: time, frequency and scale. The wavelet technique, when applied to this last domain, can be understood as a necessary expansion of the Fourier and short-time Fourier techniques. Given their practical interest, we focus on discrete wavelet filters, specifically those devised by Daubechies and Meyer [1, 4, 6], chosen specifically for their theoretical properties. The calculations are done using MATLAB 7.0 and focused on the interpretation of the methodology in some practical cases.

Finally we present those conclusions drawn from the empirical study using data for water consumption in Santa Cruz de Tenerife and for total energy and electricity consumption in Spain.

2 From Filter in the Time and Frequency Domains to the Scale Domain

The Fourier basis functions (sines and cosines) are very appealing when working with stationary time series. However, restricting ourselves to stationary time series is not very desirable since most economic/financial time series exhibit rather complicated patterns over time. The Fourier transform cannot efficiently capture these events. In fact, if the frequency components are not stationary and are able to change over time, traditional spectral tools, such as the Fourier analysis, may miss such frequency components.

The Fourier transform is an alternative representation of the original data as a function of frequency without any time information. In this sense, the Fourier representation is the opposite of how we observe the original time series. The short-time Fourier transform was developed to provide a balance between time and frequency, applying the Fourier transform to pieces of the time series of interest.

However, wavelets are, by definition, small waves. Wavelet analysis has various points of similarity and contrast with Fourier analysis. One of the drawbacks of Fourier transforms, though, is that they assume that the frequency content of the function is stationary along the time axis. Wavelets are not homogenous over time, being instead functions with more irregular waves than those of sines and polynomials. As such, the wavelet transform can be considered a scale transform because it use a basic function that is adapted to capture features that are local in both time and frequency. This makes the wavelet transform a better tool than the Fourier transform for studying general time series. In this paper, we basically use the so-called multiresolution analysis, which separates the observed data into timescales using multiple time series associated with the original time series.

There are numerous families of wavelet functions, but not all are equally appealing from a financial application standpoint. We focus here on the discrete case and on the Daubechies (denoted dbN) and Meyer (denoted dmey) wavelets. The first are commonly used in many practical cases because they are the generalization of the simplest case, the Haar wavelets. The second one has better properties (symmetry, smoothness ...) but their computational cost is higher.

On the one hand, the Daubechies wavelets [6, 10], named after Ingrid Daubechies, are a family of orthogonal wavelets defining a discrete wavelet transform that provide the maximum number of vanishing moments for some given support, which generates an orthogonal multiresolution analysis. This family includes the Haar wavelet, written db1, the simplest wavelet imaginable and certainly the earliest. In general, the Daubechies wavelets dbN are chosen to have the highest number N of vanishing moments for a given support width 2N-1. dbN are not symmetrical and for some, the asymmetry is very pronounced. The regularity increases with the order and the functions are more regular at some points than at others. The analysis is orthogonal. The wavelet transform is also easy to put into practice using the fast wavelet transform. Daubechies wavelets are widely used to solve a broad range of problems, e.g. self-similarity properties of a signal or fractal problems, signal discontinuities, etc. These wavelets are not defined in terms of the resulting scaling and wavelet functions; in fact, they cannot be written in closed form, except for db1. They are generated by using the cascade algorithm an appropriate number of times.

On the other hand, the Meyer wavelets [10, 12] and their scaling functions are defined in the frequency domain, starting with an auxiliary function. These wavelets ensure an orthogonal analysis as well. The function does not have finite support, but decreases to zero faster than any inverse polynomial. This property also holds for the derivatives and the Meyer wavelet is infinitely differentiable.

We refer the reader to [5, 8, 13] for the theoretical considerations on this type of wavelet, whose most important properties, as taken from the MATLAB 7.0 Manual, are summarized in Table 1.

Property	dbN	dmey
Arbitrary Regularity	*	
Compact support	*	
Symmetry		*
Asymmetry	*	
Number of vanishing moments	* (N)	
Orthogonal analysis	*	*
Exact reconstruction	*	
Explicit expression		
DWT	*	*
CWT	*	*
Fast algorithm	*	*

Table 1

3 Multiresolution decomposition

Our contribution focus on the practical interpretation of the multiresolution decomposition (MRD) of a discrete signal for the economic data selected.

In general, the multiresolution decomposition of the variable or signal s_t at the level N is given by $\{a_N, d_N, d_{N-1},..., d_1\}$ such that $s_t=a_N+d_N+d_{N-1}+...+d_1$.

$$s_t = \underbrace{a_1}_{\underbrace{a_2}_{\substack{\vdots \\ a_N + d_N}} + d_2} + d_1$$

a_N is called the trend signal at level N and d_i is the detail signal at level i constructed using the *father* (or scaling function, representing the smooth, trend or low-frequency) and *mother* (representing the detailed or high frequency) wavelets, respectively [5].

In economics, s_t is normally a discrete real time signal. Then, we can directly use a discrete wavelet transform (DWT) or the continuous version of the wavelet transform (CWT), if we assume a continuous signal x(t) approximating s_t.

In general, the CWT is a function of two variables W(U,s) and is obtained by projecting the function x(t) onto a particular wavelet ψ via

$$W(u,s) = \int_{-\infty}^{\infty} x(t)\psi_{u,s}(t)dt$$

where $\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right)$

is the translated (via u) and dilated (via s) version of the original wavelet function. Then, we may view the DWT as a discretization of the CWT through sampling specific wavelets coefficients. We can see that Fourier transform is a particular case when $\psi_{u,s}(t) = \psi_u(t-u) = e^{-2\pi tui}$ $(s=1, i=\sqrt{-1})$. The classical DWT is obtained from CWT using wavelets with $s=2^{-j}$ and $u=k2^{-j}$ where j and k are integers (dyadic basis) representing the set of discrete translation and discrete dilations. The value

$$\psi_{u,s(t)} = s^{j/2}\psi\left(2^j t - k\right)$$

for all integers j, k produces an orthogonal basis with a dyadic dilation.

The interpretation of the MRD using the DWT and its inverse (IDWT) is of interest in terms of understanding the frequency at which activity in the time series occurs. For example with an annual, monthly, weekly, daily, or hourly time series, the interpretation of the different scale d_i with a dyadic dilation is given in Table 2.

Scale	**Annual Frequency Resolution between**	**Monthly Frequency Resolution between**	**Weekly Frequency Resolution between**	**Daily Frequency Resolution between**	**Hourly Frequency Resolution between**
d_1	2-4 years	2-4 months (including quarterly cycle)	2-4 weeks	2-4 days	2-4 hours
d_2	4-8 years	4-8 months (including bi-annual cycle)	4-8 weeks (including *monthly* cycle)	4-8 days (including *weekly* cycle)	4-8 hours
d_3	8-16 years (including decade cycle)	8-16 months (including annual cycle)	8-16 weeks (including *bi-monthly* and *quarterly* cycle)	8-16 days (including *bi-weekly* cycle)	8-16 hours (including *half-day* cycle)
d_4	-	16-32 months	16-32 weeks (including *bi-annual* cycle)	16-32 days (including *monthly* cycle)	16-32 hours (including *daily* cycle)
d_5	-	32-64 months	32-64 weeks (including *annual* cycle)	32-64 days (including *bi-monthly* cycle)	32-64 hours
d_6	-	64-128 months (including decade cycle)	64-128 weeks	64-128 days (including *quarterly* cycle)	64-128 hours
d_7	-	-	128-256 weeks	128-256 days (including *bi-annual* cycle)	128-256 hours (including *weekly* cycle)
d_8	-	-	256-512 weeks	256-512 days (including *annual* cycle)	256-512 hours (including *bi-weekly* cycle)
d_9	-	-	512-1024 weeks (including *decade* cycle)	512-1024 days	512-1024 hours (including *monthly* cycle)
d_{10}	-	-	-	1024-2048 days	1024-2048 hours
d_{11}	-	-	-	2048-4096 days (including *decade* cycle)	2048-4096 hours (including *quarterly,* cycle)
d_{12}	-	-	-	-	4096-8192 hours (including *bi-annual* cycle)

Table 2

The following annotations are of practical interest:

- The number of observations T dictates the choice of wavelet and number N of scale d_j that can be produced with the condition $T \geq 2^N$.
- The calculation of a large decomposition can imply correlations in the results.
- The calculation of a decomposition using high-degree wavelets can yield numerical instabilities, though if these are avoided, a more regular asymptotic behaviour is achieved.
- Wavelet MRD analysis assumes that data are sampled at equally-spaced intervals.
- Usually, in the MRD for economic data, the activity lasts for a decade at most. For example, for yearly data, the decomposition makes sense at level 3 at most and for monthly data at level 6 at most.
- The interpretation of the scales when using more highly disaggregated series, such as those involving weekly, daily or hourly data, for example, is more delicate (hence the use of italics in the above table), since the number of hours/day, hours/week, weeks/month, days/month, hours/month, weeks/year, days/year, hours/year, etc., does not have a set whole value. Moreover, depending on the type of activity, each sample may contain data associated with only certain days or hours (e.g. daytime or night-time activity, activity on certain days of the week). This would require deciding whether the data were lost or whether such gaps were intrinsic to the activity in question. The stock market, for example, is not open every day; education only takes place on school days, etc.

4 Working with real data

The examples to be considered involve series of varying natures which exhibit a certain periodic behaviour. We use wavelets to reveal the different cycles that are hidden by the data. The calculations and graphs were done on MATLAB 7.0, WAVELETS 1-D. Only the graphs are shown here.

Two cases are considered, one on water consumption in Santa Cruz de Tenerife (hourly data), and another on total energy and electricity consumption in Spain (monthly data).

Case 4.1 Water consumption data in the city of Santa Cruz de Tenerife (hourly data)

The water demanded by the consumers is distributed from several reservoirs, which divide the urban map into zones. Figure 1 shows the initial part of the series for one of these reservoirs, Cueva Roja, from a real data sequence compiled at 1-hr intervals over 31 days: 754 observations, from 00:00 on 11 October, 1996, to 09:00 on 11 November, 1996. The data, which were provided by the Santa Cruz de Tenerife water utility, ENMASA, reflect overall consumption, including both demand and losses due to leakage in the water collection and distribution system.

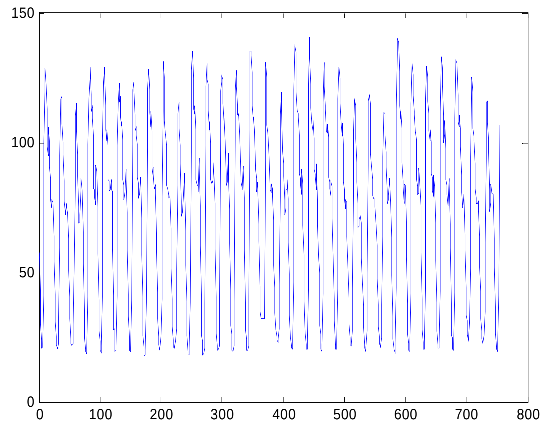


Figure 1

The daily and half-day periodicities were already revealed by the Fourier transform [2], as shown in Figure 2.

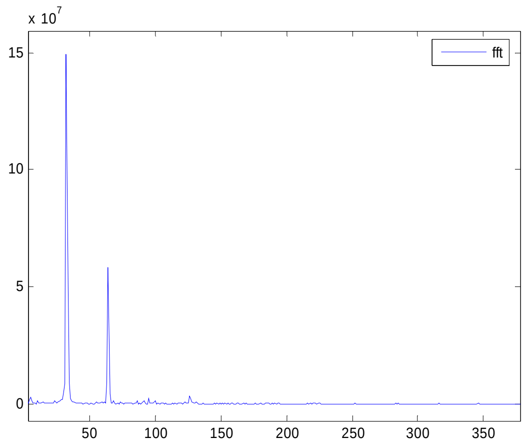


Figure 2

Next, we decompose the signal using wavelets up to, at a minimum, level 4 in order to observe these periodicities. The available data allows for decomposition up to level 9.

Using first-level Daubechies db3 wavelet decomposition yields a *trend* signal a_1 (in blue, Figure 3) and a *detail* signal d_1 (in green, Figure 3).

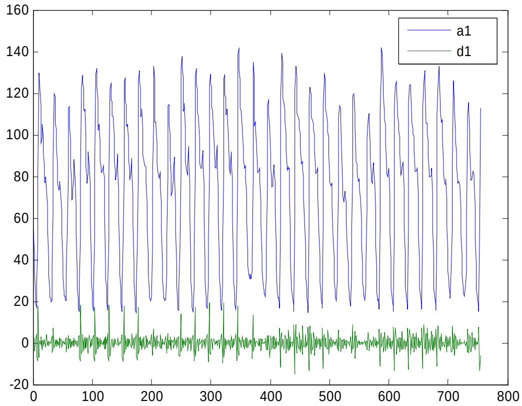
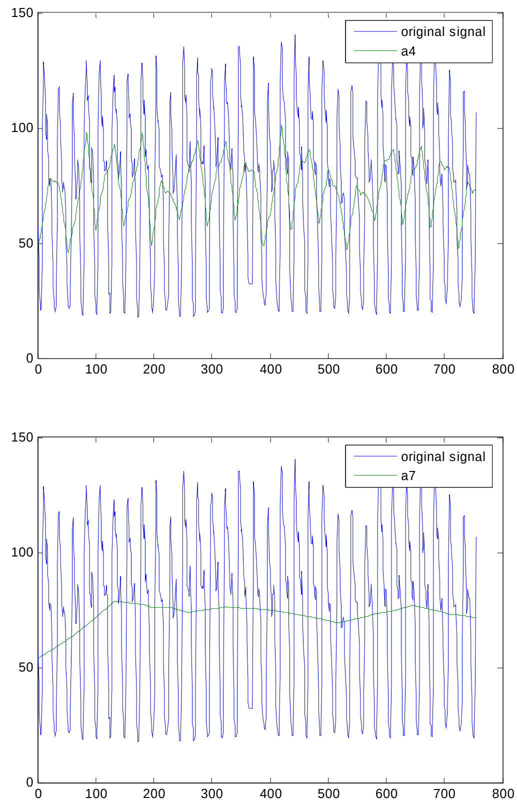


Figure 3

Note the similarity between a_1 and the original signal, and how d_1 reveals what could be considered local irregularities, which might be interpreted as *noise*.

If signal a_1 is decomposed again, another *detail* signal result, yielding information on another frequency, and so on successively. a_3 is the first signal decomposition from which relevant information is extracted in d_3. If details are calculated down to level 7, we obtain a trend component signal a_7 and seven *detail* signals, d_1, d_2 ... d_7, some of which are shown in Figure 4. At this level of decomposition all of the changes inherent in the signal are compiled in details d_i, i=1…7.



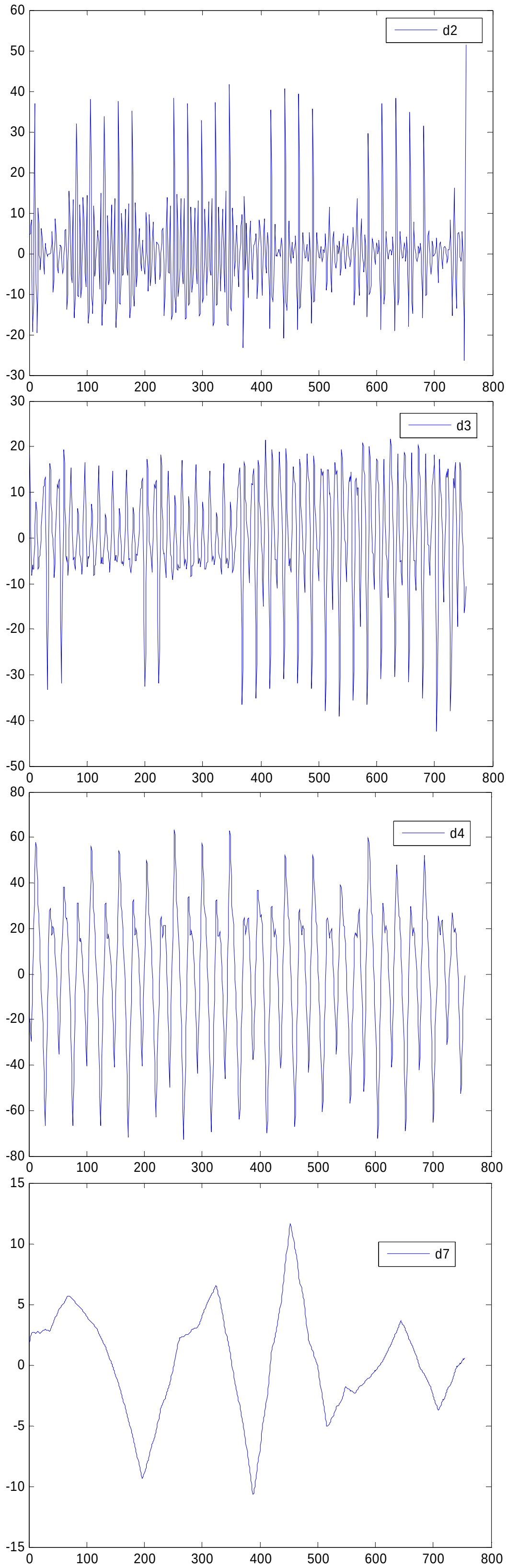


Figure 4

The detail signal d_2, like d_1, shows local irregularities. Note how d_3 and d_4 are the most regular details of all and reveal the periodicities inherent in the data, that is, the 12-hr (half day) and 24-hr (daily) periodicities, respectively; d_4 being greater in magnitude. This fact coincides with the findings obtained through the use of the Fourier transform [2]. With the wavelet technique more information is obtained since said frequencies are associated with certain time periods, meaning the maxima are localized in the morning hours and the minima in the evening. d_3 reveals the minima associated with 7:00 a.m. on Saturdays and Sundays in the first segment, and every day in the last segment. d_6 reveals once more the maxima during the morning hours and the minima in the evening hours.

So as to more precisely calculate the periodicities, Figure 5 shows the spectrum for each detail in order from d_1 to d_8. Note how the one with the greatest magnitude is, in fact, d_4, followed by d_3. The numerical results associated with the more relevant frequency indices are listed in Table 3.

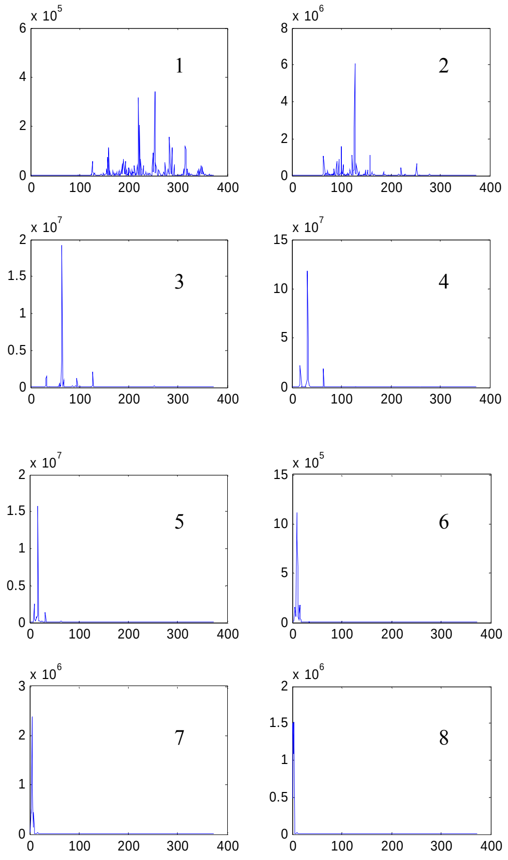
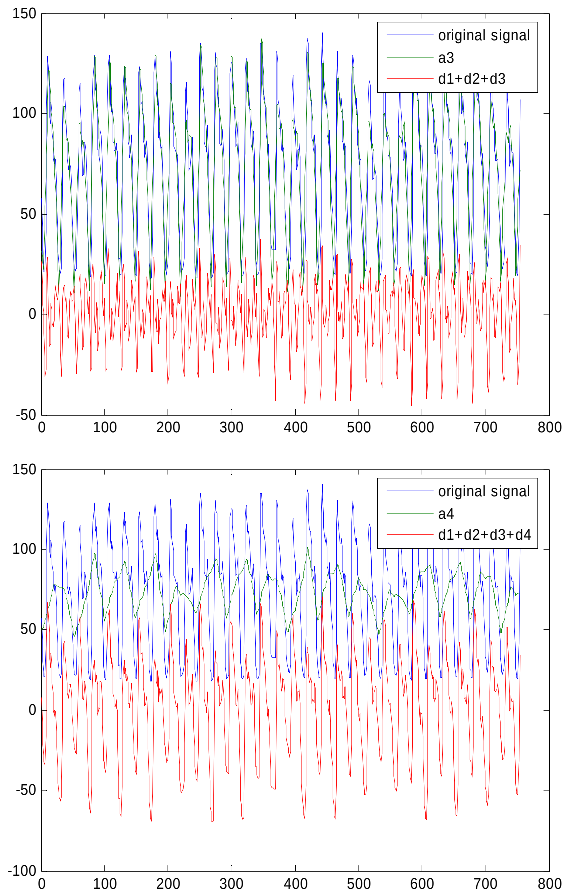


Figure 5

db3	**Most Relevant Frequency Index**	**Hourly Period**
d_1	253, 220, 283, 315	2-4 hours
d_2	127, 100, 158	6-8 hours
$\mathbf{d_3}$	**64**	**12 hours**
$\mathbf{d_4}$	**32**	**24 hours (1 day)**

Table 3

As for a_1 and a_2, these are similar to the original signal. As shown by MATLAB, it is reasonable to stay on the fourth level of decomposition, that is, with approximation a_4 and details d_1 to d_4, which completes the results obtained by the Fourier transform in terms of frequency. The original signal is shown in Figure 6, along with the superimposed trends from a_3 to a_7, as well as the sum of the details extracted from level 1 to level i-th, respectively. This allows us to observe the progression of the *trend* signal and the signal that includes the sum of the *details*.



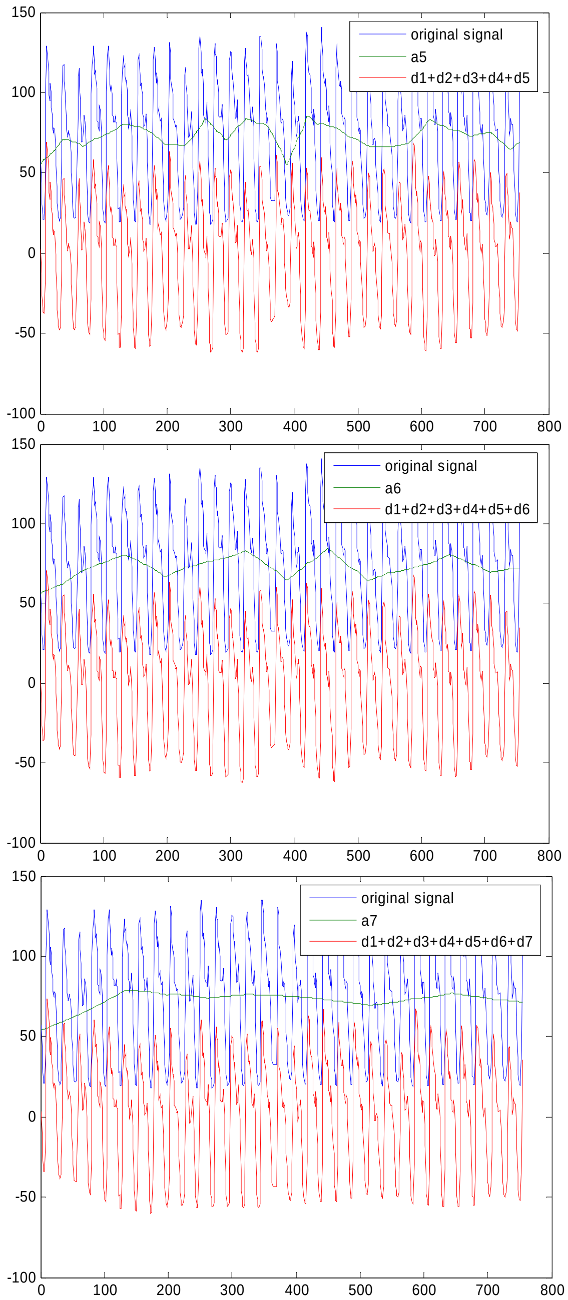


Figure 6

Case 4.2: Total energy consumption in Spain (monthly data from January 1992 to September 2008)

The data were obtained from the Monthly Statistics Bulletin (National Statistics Institute, INE in Spanish) at www.ine.es (201 observations in millions of PET (petroleum equivalent tonnes). They include both demand and losses due to distribution system. They do not include renewable energies since monthly data for this energy source were not available.

The results of the decomposition analysis using db3 and dmey are summarized in Figures 7 and 8, respectively. In each figure, the first graph refers to real data, the second to the *trend* signal a_3 and the subsequent figures to the *detail* signals, d_1, d_2 and d_3, respectively, each shown with its corresponding spectrum.

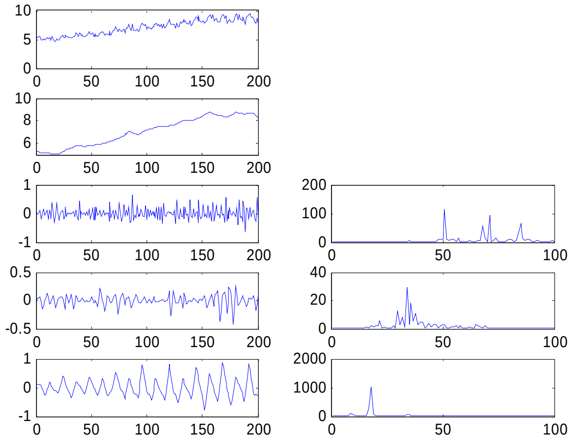


Figure 7

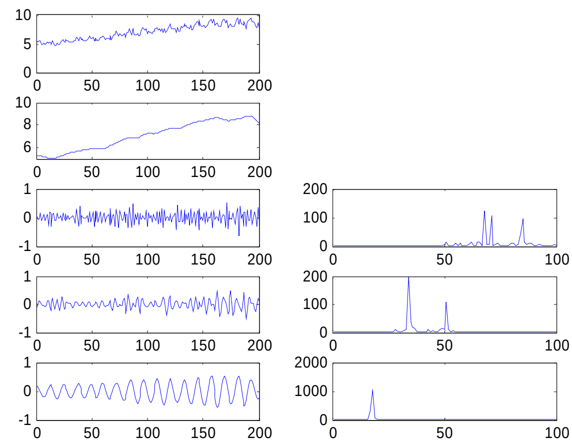


Figure 8

Tables 4 and 5 list the most relevant frequency indices for each d_i.

db3	**Most Relevant Frequency Index**	**MonthlyPeriod**
d_1	51	4 months
d_2	**34**	**6 months**
d_3	**18**	**11 months (≈1 year)**

Table 4

dmey	**Most Relevant Frequency Index**	**Monthly Period**
d_1	68	3 months
d_2	**34**	**6 months**
d_3	**18**	**11 months (≈1 year)**

Table 5

Case 4.3: Electric energy consumption in Spain (monthly data from January 1992 to September 2008)

These data were also obtained from the Monthly Statistics Bulletin (INE) at www.ine.es (201 observations in thousands of PET), and do not include electricity from renewable sources since no monthly data were available.

As above, Figures 9 and 10 summarize the results obtained from the decomposition analysis using db3 and dmey, respectively. Tables 6 and 7 show the most relevant frequency indices.

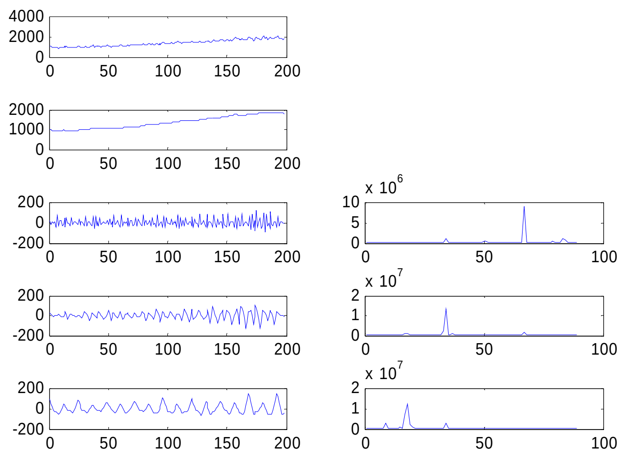


Figure 9

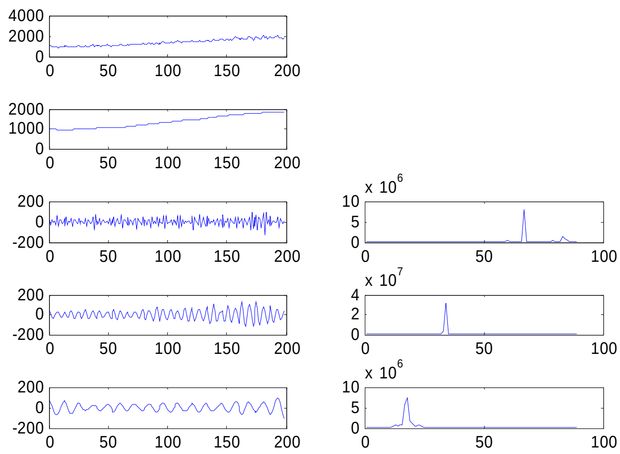


Figure 10

db3	Most Relevant Frequency Index	Monthly Period
d_1	67	3 months
d_2	34	6 months
d_3	18	11 months (≈1 year)

Table 6

dmey	Most Relevant Frequency Index	Monthly Period
d_1	67	3 months
d_2	34	6 months
d_3	18	11 months (≈1 year)

Table 7

Case 4.4: Electric energy consumption in Spain (hourly data from 1st January 1998 1:00 to 28th January 1998 24:00 (672 observations))

These data were obtained from the Operador del Marcado Eléctrico (OMEL) at www.omel.com (35064 observations in megawatts). Alternatively, we can obtain longer data series.

If we decompose a long time series (hourly data from one year or longer than) using wavelets db3, we would see that d_3 shows four time sectors with maximum and minimum values in the winter and summer, respectively. This implies that there is a frequency associated with the seasons. d_3 and d_4 once again contain relevant information on half-day and daily periodicities, as does d_7 on the weekly periodicity. Displaying these periodicities, however, requires zooming in by areas, since the data cloud is very dense. These periodicities are basically those analyzed by [11] through the use of spline techniques. Also, we can see some results on time series electricity consumption in other pays obtained by, for example, [9…] using splines and [3…] using wavelets.

On the other hand, we have chosen a short time series so as to compare the results with those obtained in previous cases involving water, total energy and electricity consumption.

The Fourier Transform (Figure 11) shows that the most relevant indices are 29 and 57, which correspond to daily (672/29, approx. 24 hours) and half-day (672/57, approx. 12 hours) periodicities in this order of importance, similar to the previous case on water and total energy consumption.

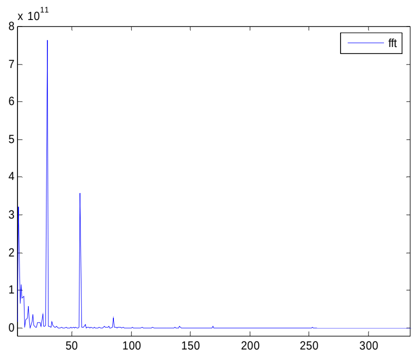


Figure 11

Let us consider the additional information offered by the wavelets technique. Using Daubechies db3 reveals the same frequencies as those obtained with the Fourier transform, namely, half day in d_3 and daily in d_4 (Figure 12). New information is also added at other frequency levels (Figure 14). In fact, a_1 and a_2 are very similar to the initial signal, meaning that d_1 and d_2 could be regarded as noise corrections. a_4 is a good approximation of the initial signal once it is decomposed to level 4, and its shape is reminiscent of a weekly frequency that might be observable in d_8 if the signal decomposition were to continue, which, given the amount of data available, is viable. Nonetheless, a_8 is a very smooth approximation that is not a characteristic representation of the initial signal, but of the trend. If we decompose it, we see that d_6 and d_7 detect abnormally low values in observations close to data 400 and 600, respectively.

The Fourier Transforms for each detail are shown in Figure 13 (from d_1 to d_4) and Figure 15 (d_5 to d_8). Table 8 illustrates the most relevant frequency indexes.

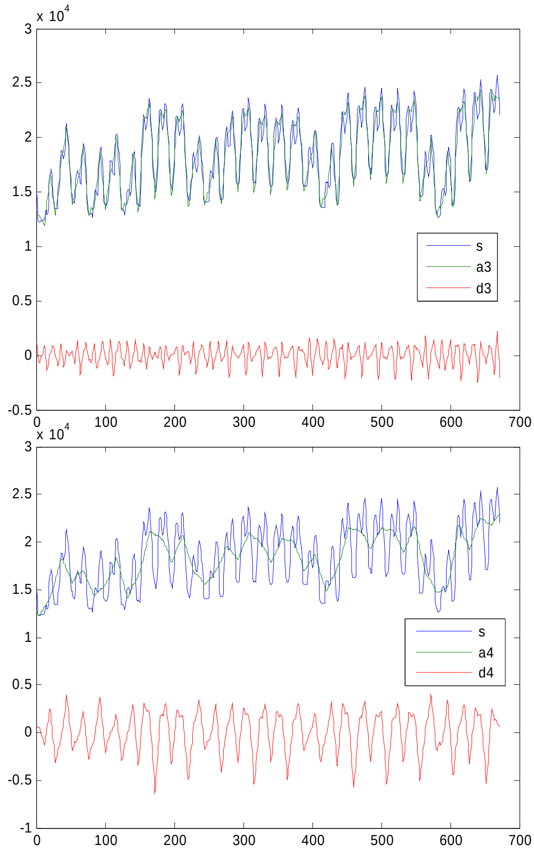


Figure 12

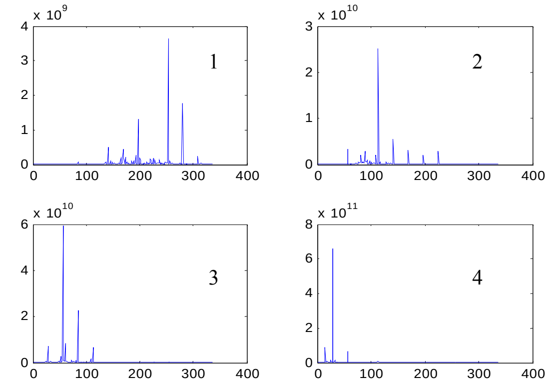


Figure 13

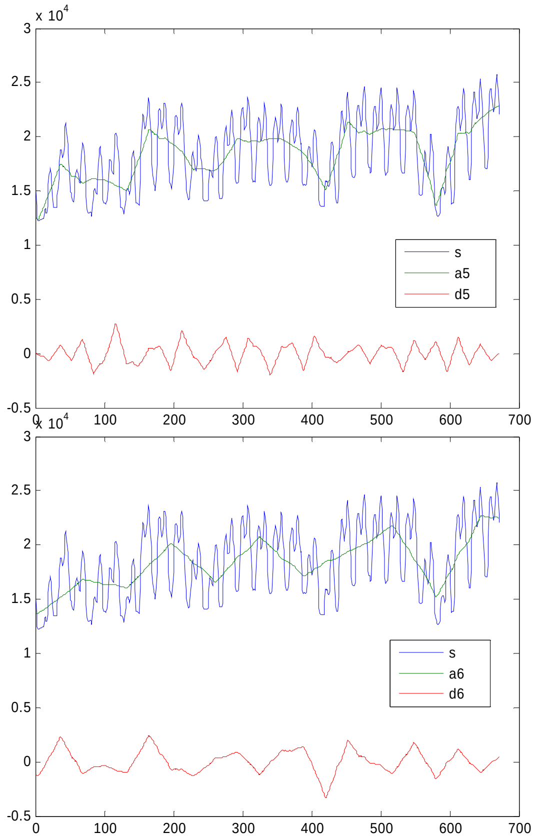


Figure 14

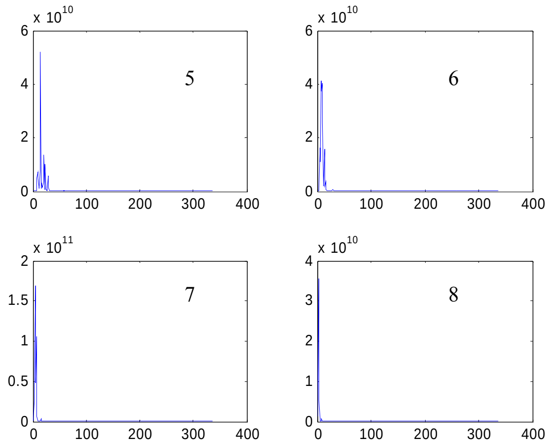


Figure 15

db3	Most Relevant Frequency Index	Hourly Period
d_1	253, 281, 197, 141, 169	2-4 hours
d_2	113	6 hours
d_3	57	12 horas
d_4	**29**	**24 hours (1 day)**
d_5	15	44 hours (≈2 days)
d_6	7,9	96-74 hours (≈4-3 days)
d_7	**5**	**134 hours (≈ 6 days)**
d_8	3	224 hours (≈ 9 days)

Table 8

5 Conclusions

In this paper we present the numerical and graphical results obtained using wavelet-based multiresolution analysis as applied to two data types relevant to economics, water, total energy and electricity consumption.

We note the usefulness of this technique to expand studies conducted on time and frequency domain data series in that it provides us with simultaneous information in both domains through the so-called scale domain.

Depending on the data type and due to their possible inhomogeneity, drawbacks may arise in the interpretation of frequency indices associated with the different scales. Also noted were certain differences in the *smoothness* of the series obtained as a function of the type of wavelet utilized.

The data series studied exhibited significant periodicity components. The use of data for which this property does not hold would be of interest, as might be the case with financial data, so as to better observe the possible interpretation of the multiresolution analysis in different time segments within the series.

Acknowledgement: This research was partially funded by Plan Nacional Project no.MTM2008-06671, MTM2006-14961-C05-03 and the Centro de Investigación Matemática de Canarias (CIMAC).

References

[1] Abry, P., Ondelettes et turbulence, Diderot ed., Paris, 1997.

[2] Alonso, S. and González-Concepción, C., Linear approximation in the discrete frequency domain, Journal of Computacional and Applied Mathematics, 126, 2000, 255-267.

[3] Benaoudaa, D., Murtaghb, F., Starckc, J.-L. and Renaudd, O., Wavelet-based nonlinear multiscale decomposition model for electricity load forecasting, Neurocomputing, 70, 1-3, 2006, 139-154.

[4] Chui, Ch. K., An Introduction to Wavelets, Academic Press, Inc., 1992.

[5] Crowley, P.M., A Guide to Wavelets for Economists, Journal of Economic Surveys, Vol. 21, No. 2, 2007, pp. 207-267.

[6] Daubechies, I., Ten lectures on wavelets, 61, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

[7] Fay, D. and Ringwood, J., A Wavelet Transfer Model for Time-Series Forecasting, International Journal of Bifurcation and Chaos, 17, 10, 2007, 3691-3696.

[8] Gencay, R., Faruk, S. and Whitcher B., An Introduction to Wavelets and Other Filtering Methods in Finance and Economics, Academic Press, Elsevier, 2002.

[9] Harvey, A.C. and Koopman, S.J., Forecasting hourly electricity demand using time-varying splines, Journal of the American Statistical Association, 88 (424), 1993, 1228-1236.

[10] MATLAB Software, The Language of Technical Computing, The MathWorks, 2007.

[11] Martín-Rodríguez, G. and Cáceres-Hernández, J.J., Modelling the hourly Spanish electricity demand, Economic Modelling, 22, 2005, 531-569.

[12] Meyer, Y., Wavelets: Algorithms and Applications, Society for Industrial and Applied Mathematics Translation, 1993.

[13] Ramsey, J., Wavelets in Economics and Finance: Past and Future. Studies in Nonlinear Dynamics & Econometrics, Vol. 6, No. 3, 2002, 1-27.

[14] Ramsey, J. and Lampart, C., The Decomposition of Economic Relationships by Time Scale Using Wavelets: Expenditure and Income. Studies in Nonlinear Dynamics and Econometrics, Vol. 3, No. 1, 1998, 23-41.