A spectral analysis of meteorological data for weather forecast applications

Eros Pasero[1], Alfonso Montuori[1], Giovanni Raimondo[1]
[1]Department of Electronics, Polytechnic of Turin
Corso Duca degli Abruzzi 24, 10129 Turin, Italy
Phone: +39-011-564-4063, Fax: +39-011-564-4099, E-mail: giovanni.raimondo@polito.it

ABSTRACT

The great majority of natural signals are not stationary and they undergo transients which include a large range of frequencies in a short period of time. Fourier Transform is not sufficient to process this kind of signals because all the information on time location of a certain frequency is lost in the analytical process. The complexity of climate variability on all time scales requires the use of several refined tools to unravel its primary dynamics from observations. The analysis of the temporal series has shown that climatic variations are extremely irregular in the time-space domain. This feature makes them difficult to be foreseen if no particular mathematical tools are used. The aim of this paper is to describe how to analyze and represent climatic signals in order to foresee their future short-time evolution for weather nowcasting. A spectral analysis of the meteorological data is described. More precisely the purpose of the spectral analysis is the determination of the optimum sampling frequency for a set of climatic not stationary signals and the description of an algorithm for the automatic adjustment of such a frequency.

Keywords: climatic data spectral analysis nowcasting.

1. INTRODUCTION

A great variety of applications, regarding technological fields very different from each other, require the processing of signals that represent, in general, the trend of the physical quantities under analysis. In many cases such signals are electric signals: bioengineering, telecommunications, mechanics, the study of seismic phenomena are some samples of scientific fields in which the information to be processed (the heart beat, television's images, a mechanical vibration, a seismic movement) is first of all transduced (by means of the right transducers) into an electric signal and then processed in this form. With the spreading of computers and microprocessors the signals are often represented as sequences of numerical values that in the continuous or discrete time domain, can be analyzed with numerical techniques and mathematically modeled as random processes (see [1] and [5]).

Climatic signals in the meteorological field can be analyzed with the same techniques: like the great majority of natural signals climatic signals are not stationary and they experience transients that include a large range of frequencies in a short period of time. These signals can be represented by representations with expansions on bases of functions possibly complete, not redundant and orthogonal or orthonormal. In other words we can represent a time-continuous signal (but the same procedure is easily extensible to discrete-time signals) by means of a sequence (possibly a finite one) of coefficients (see [1] and [6]). This representation may be exact or only approximated, and the corresponding basis shall be chosen by means of a trade-off between accuracy and simplicity. Given a finite energy, real signal x(t), $t \in R$, the Fourier Transform X(f) of the signal, can be interpreted as a particular kind of scalar product:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt = < e^{j2\pi ft}, x(t) >$$

The functions $e^{-j2\pi ft}$ are an infinite set of complex orthogonal functions, in fact it is easy to show that $< e^{-j2\pi f_i t}, e^{-j2\pi f_j t} >= 0 \;\; \forall \; i \neq j$.

We assume to have a sampled signal $x(nT_0)$ where $f_s=1/T_0$ is the sampling frequency; we further assume that $x(nT_0)$ is meaningfully different from zero only in the interval $t \in (0,T)$.

So we can build a periodical version $\dot{x}(t)$ of the signal $x(t)$:

$$\dot{x}(t) = \sum_{n=-\infty}^{+\infty} x(t - nT)$$

$$\dot{X}(f) = \sum_{n=-\infty}^{\infty} X(f - nB)$$

where $\dot{X}(f)$ is the amplitude spectrum of $\dot{x}(t)$ and $B= 1/T_0$ is another way to write the sampling frequency. And we have $N=T/T_0$ samples on which we calculate the transform. The coefficients of the Discrete Fourier Transform (DFT) of the sequence $\dot{x}(nT_0)$, for $n \in [0, N-1]$, can be written in the following way:

$$\dot{X}(nf_0) = \frac{T}{N}\sum_{m=0}^{N-1}\dot{x}(mT_0)e^{-jn\frac{2\pi}{N}m}$$

The Fourier Transform can be adjusted to analyze only a little portion of the signal each time: this technique is called windowing of the signal (Short-Time Fourier Transform, STFT) [10]. The STFT represents the signal as a bidimensional function of time and frequency. It provides information concerning when and at which frequencies an event present on the signal analyzed happened. Anyway the accuracy of such information is bounded by the window's dimension. The disadvantage of such a technique is that once you choose the width of the window this remains the same for all frequencies. This fact affects the accuracy with which we can identify in time and frequency possible transients present on the signal. If the sphere of sinusoidal signals is exited there are bases more suitable than Fourier's basis for the analysis of transients and not stationary signals. Such bases have to be local in time and frequency and simple to obtain. The wavelet analysis represents a windowing technique with variable dimensions. Such technique allows to use a long time interval if we want to get from the signal information on its contents at low frequencies, and short time intervals if we want to investigate on the contents of the signal at high frequencies. We can define the Continuous Wavelet Transform (CWT) of a signal x(t) as in the following:

$$CWT_x(a,b) = \int_{-\infty}^{\infty} x(t)\psi_{ab}^{*}(t)dt$$

$$= < \psi_{ab}(t), x(t) >$$

$$\psi_{ab}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right)$$

The parameter *a* is called scaling factor and it is strictly greater than zero, *b* is the time translation. The prototype function $\psi(t)$ is called mother wavelet, it has to have a zero mean value and a good localization in time; so it oscillates in a damping way and extinguishes as a little wave, this is why it is called wavelet. An acceptable wavelet $\psi(t)$ shall have a zero mean value and pass-band characteristics. Many methods and results of meteorological data spectral analysis already exist in the literature ([2-3],[8-9],[12]), but their aim is the study of the trends of global/medium scale meteorological data in the medium and long term (on time scale) while the analysis described in this paper is useful for local weather nowcasting applications (see [11]): forecasting with up to three hours in advance of the local climatic data trend. The spectral analysis described in this paper was carried out in two steps. Firstly a Fourier Analysis (Fast Fourier Transform, FFT [5], a much faster algorithm to compute the DFT) of the climatic data was performed (Section III). The outcome of the Fourier analysis was the determination of the optimum sampling frequency on the whole observation interval. Then many questions arose: does the sampling frequency have to be held at its optimum value all the day? In which periods of time is each signal more stationary? In which others does it undergo fast transients? Is it possible to find an algorithm for the automatic adjustment of the sampling frequency? To answer all these questions the wavelet analysis (Discrete Wavelet Transform, DWT, see [6], [10]) of the same data was carried out in order to detect potential drifts and transients with daily periodicity (Section IV).

2. CLIMATIC SIGNALS

A meteorological station and a sensor for air temperature and humidity measurements have been installed at the Department of Electronics of Polytechnic of Turin. A data acquisition unit is connected to the sensors enabling the recording of the relevant values as shown in Table 1. The data used for the spectral analysis refer to the period of time 09/15/2003-02/17/2003. In the following sections are shown the results of the spectral analysis for air and soil temperature and humidity.

Sensor	Data	Unit of measurement	Sampling period (T_0)	Kind of Measure
Termometer	**Air temperature (2 m high)**	° C	15 min	Istantaneous value
Termometer	**Soil's temperature**	° C	15 min	Istantaneous value
Hygrometer	**Relative humidity**	%	15 min	Istantaneous value
Rain-Gauge	**Quantity of precipitation**	mm	15 min	Cumulative value
Gonioanemometer	**Wind direction**	Grades from North	15 min	Istantaneous value
Tacoanemometer	**Wind intensity**	m/s	15 min	Istantaneous value
Barometer	**Atmospheric pressure**	hPa	15 min	Istantaneous value
Pyranometer	**Solar radiation**	W/m^2	15 min	Istantaneous value

Table 1: Climatic data analyzed.

3. FOURIER ANALYSIS RESULTS

We performed the analysis on overlapping windows centered at the noon of each day. The parameters of the Fourier Transform were chosen as shown below:

- Number of window samples: N=256 (approximately 2.6 days);
- Sampling frequency: fs=1/(15 min)= 1/900 Hz = 1.111 10^{-3} Hz;
- Estimated bandwidth of the data: B=5.555 10^{-4} Hz;
- Frequency resolution: fs/N=4.34 10^{-6} Hz.

First of all the Fourier Amplitude Spectrum was calculated for each signal in each day in which all the data were significant (in the proper range of values). The proper definition of the bandwidth (B) of a signal was also chosen as the interval of frequencies in which the 99.9% of the signal energy (calculated as the sum of the coefficients of the Energy density Spectra) is located. The energy density spectrum of each signal was truncated at the minimum of these two values: the 99.9% energy value and the value for which the signal rebuilt from the truncated spectrum compared to the original one showed an absolute error equal to a certain threshold, Err_abs, determined by the precision of the corresponding sensor (Err_abs thresholds are shown in Table 2 for air temperature, soil temperature and humidity).

Signal	**Err_abs**
Air temperature	**0.3 °C**
Soil temperature	**0.3 °C**
Humidity	**1%**

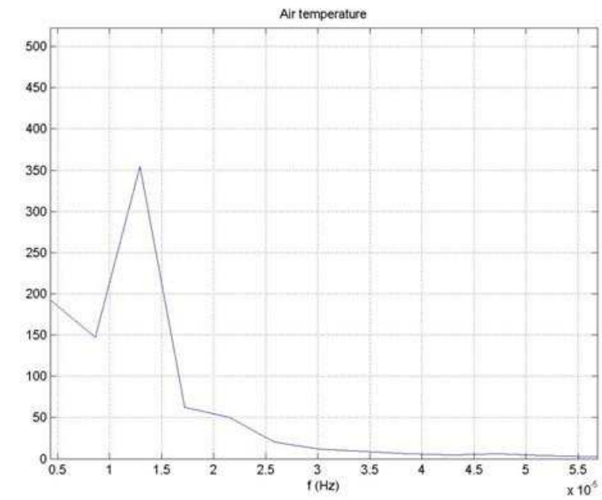
Table 2: Truncation thresholds of climatic data.

The highest daily values of such bandwidth were assumed to be correlated to the optimum sampling frequency fs (they are shown in Table 3 for air temperature, humidity and rain).

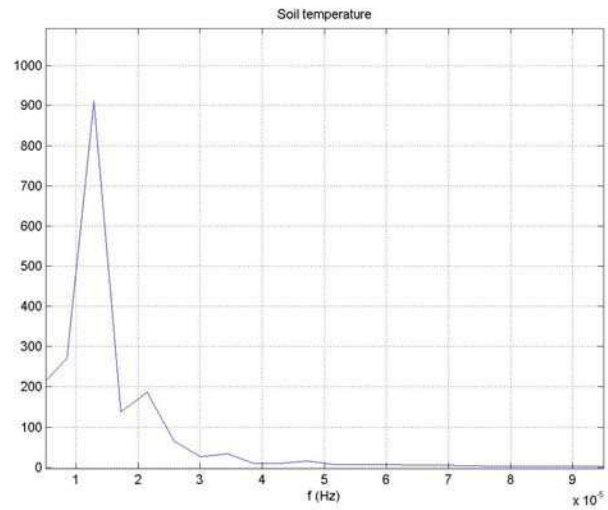
Signal	fs	T_0
Air temperature	**$8.25 \cdot 10^{-4}$ Hz**	**20.2 min**
Soil temperature	**$8.94 \cdot 10^{-4}$ Hz**	**18.6 min**
Humidity	**$6.18 \cdot 10^{-4}$ Hz**	**27 min**

Table 3: Optimum sampling frequencies (highest daily values).

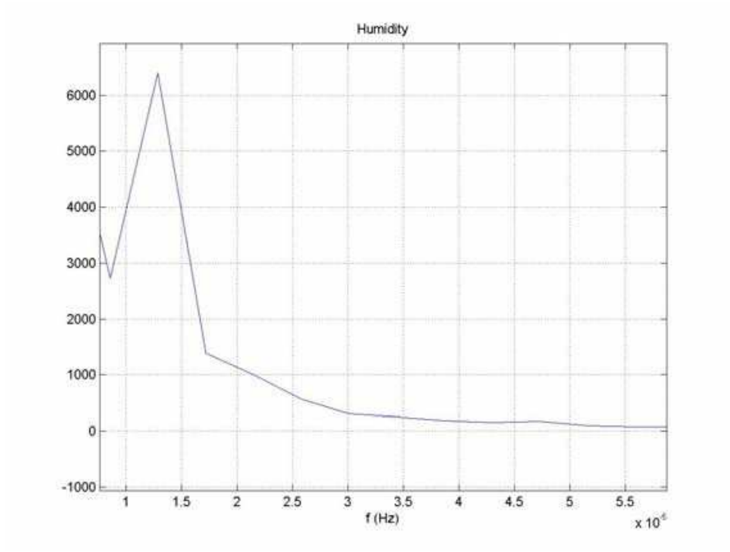
Note that for the Nyquist's criterion the band can not be greater than half the sampling frequency (fs=1.111 10^{-3} Hz). So if the optimum sampling frequency assumes a value very near to the measurement frequency it is assumed that the corresponding parameter is downsampled. Then the average spectrum on the useful days was calculated. It has to be noted that the principal peak of the average energy density spectrum for specific signals, like air and soil temperature and humidity, is on the fourth coefficient at $f=3*(1.111*10^{-3})/256$ Hz$=1.3*10^{-5}$Hz (see Picture 1, Picture 2 and Picture 3). This value indicates a frequency of one cycle per day and it's in agreement with the daily pseudoperiodicity of the signals. Other signals with not negligible high-frequencies components, such as velocity and direction of the wind, the rain and the solar radiation, have broader spectra and should be considered as downsampled. On the other hand pressure behaves nearly as a constant and has a spectrum with a high peak in the origin and a value close to zero almost everywhere else. Even if the long-term high-frequency energy contents of the climatic data are small, the short-term high-frequency energy contents could affect the forecasting capability of the nowcasting tools [11], if no time-dependent sampling frequency's adjustment is performed. Consequently the wavelet analysis was applied to the climatic data.



Picture 1: Average Energy Density Spectrum of the air temperature.



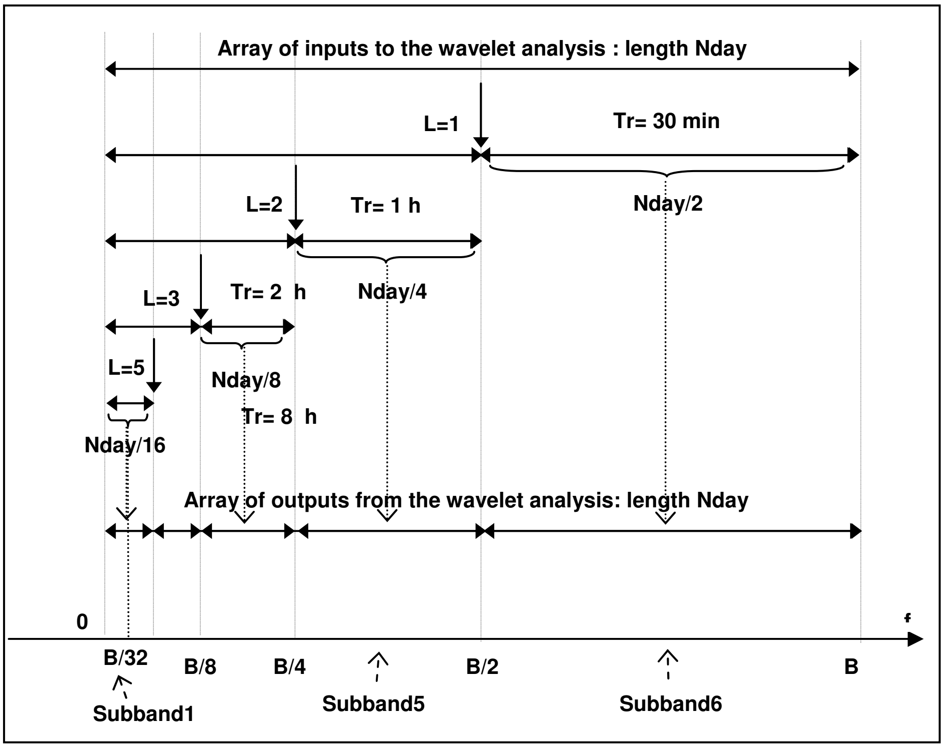
Picture 2: Average Energy Density Spectrum of the soil temperature



Picture 3: Average Energy Density Spectrum of humidity.

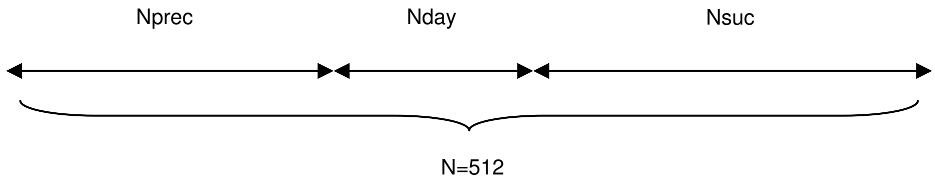
4. WAVELET ANALYSIS RESULTS

The Matlab-compliant software packet Wavelab [4] was used to carry out the Wavelet analysis. An analysis was performed using Haar orthonormal basis (see [10]). In fact we were interested in time location of dynamical transients and the analysis with the highest time resolution had to be used (see [10]). The output of the wavelet transform is set up by a number of levels set by the user. We chose 5 levels of analysis. Each level is relative to a given subband of the original signal and it has its own time resolution. The output coefficients of the wavelet transform represent the energy components of the input signal in the subband identified by the level of the transform (if B is the band of the input signal: *subband6*= $B/2$-B for level L=1, *subband5*= $B/4$-$B/2$ for level L=2 and so on) and in the time interval identified with an accuracy equal to the time resolution, Tr, of the level (Tr = 30 min for level L=1, Tr = 1 h for level L=2, etc.). Picture 4 shows the data structure of the wavelet analysis described in this paper.

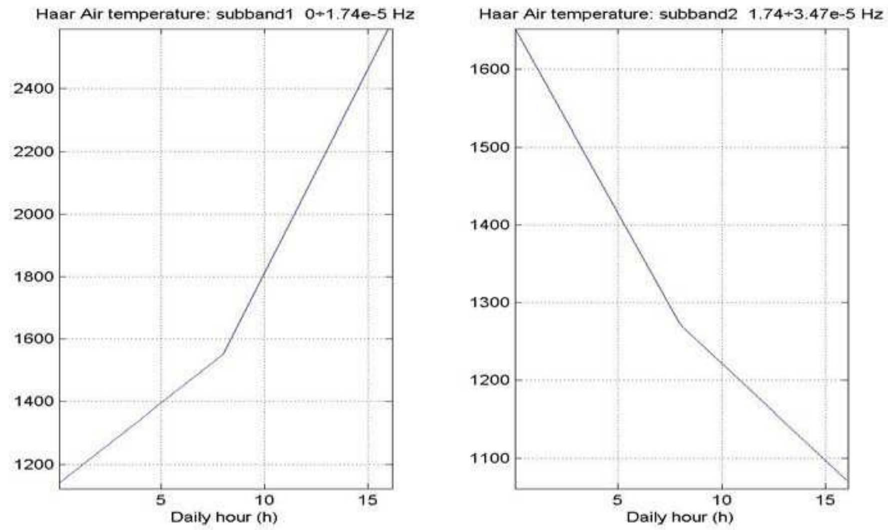


Picture 4: Data structure for the Wavelet Analysis (L=5 levels of analysis).

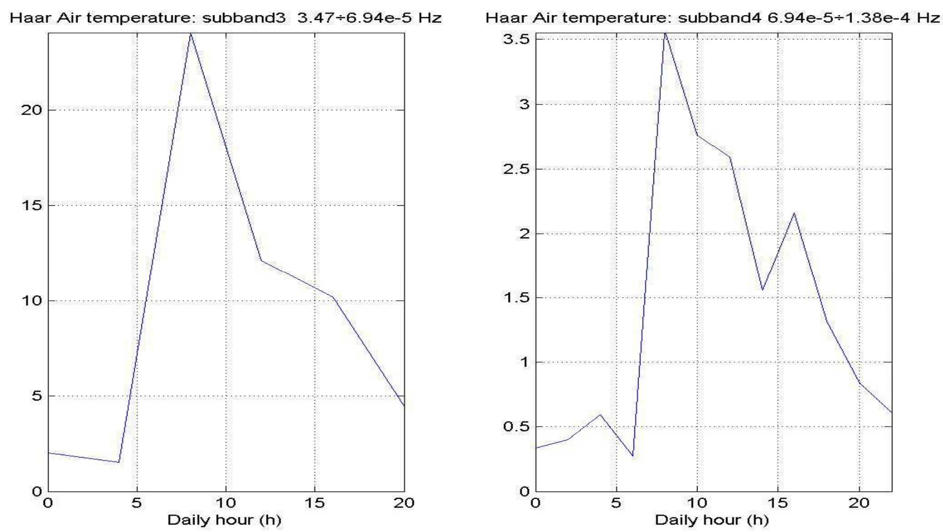
The number of samples for the transform must be a power of 2. It was chosen to be $512=2^9$. The samples were grouped as follows:



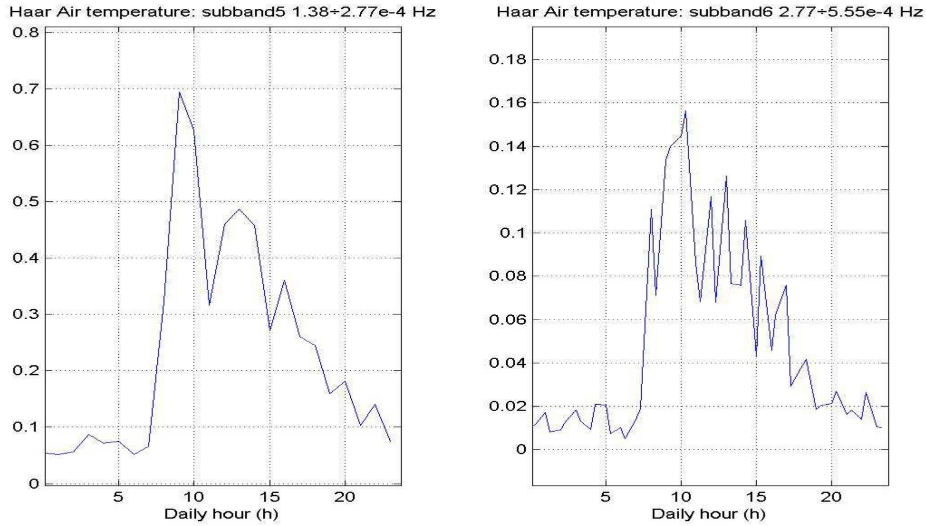
Nday are the 96 samples belonging to the day under analysis, *Nprec* (160) and *Nsucc* (256) are respectively the 160 *lookbehind* and 256 *lookforward* samples necessary to calculate the wavelet transform at the edges of each day. In this case the coarsest level of the transform, in the worst case, has valid coefficients also close to the edges of the time interval considered. Picture 5, Picture 6 and Picture 7 are the graphs relative to the air temperature, which as soil temperature and humidity shows, during the preventive analysis of data, the presence of well localized transients during the day. Such signals may need the fine adjustment of the sampling frequency during the transients' time intervals in order to be properly analyzed. As pictures show, the low-frequency energy contents of the signal (in *subband1* and *subband2*) are two orders of magnitude greater than the energy contents in *subbands 3, 4* and *5* ($f \leq 2.77*10^{-4}$ Hz) and the energy contents in such *subbands* are ten times greater than those in *subband6*.



Picture 5: DWT coefficients for the Air Temperature subband1 and subband2 (Haar basis).



Picture 6: DWT coefficients for the Air Temperature subband3 and subband4 (Haar basis).



Picture 7: DWT coefficients for the Air Temperature subband5 and subband6 (Haar basis).

From the pictures we can observe that the DWT's coefficients are significant only at low frequencies ($f \leq 2.77 * 10^{-4}$ Hz). This frequency band is lower than half the optimum sampling frequency calculated in Section III. Consequently there is no need for the adjustment of the sampling frequency during the transients. Table 4 shows the results of the wavelet analysis for air and soil temperature, and humidity: air and soil temperature have their energy peaks (not stationary dynamical behaviour) at 8 a. m. (principal peak) and 4 p. m. (secondary peak) and humidity has its peak at 8 a. m.

Parameter	Principal peak	Secondary peak
Air temperature	8 a.m.	4 p.m.
Soil temperature	8 a.m.	4 p.m.
Humidity	8 a.m.	-

Table 4: Wavelet transform peaks for some climatic data.

One of the many possible algorithms for the automatic adjustment of the sampling frequency for each climatic *datum* with transients at a certain peak hour could be for example:
$f_s \leftarrow$ Optimum sampling frequency from Fourier analysis always but in the transient hour when $f_s \leftarrow$ twice the bandwidth of the transient calculated with the wavelet transform analysis.

5. CONCLUSIONS

In this paper a particular spectral analysis and representation of climatic signals, which are based upon the use of the Fourier and Wavelet Transforms, have been discussed. A successful solution strategy for the automatic adjustment of the sampling frequency for the set of climatic data has been presented, and the approach is generally applicable to sets of not stationary signals. The outcome of the Fourier analysis was the determination of the optimum sampling frequency for each climatic signal under analysis and on the whole observation interval. The wavelet analysis gave some hints on the time localization of the daily transients for each signal and on the determination of an algorithm for the adjustment of the optimum sampling frequency during the day. Further activities of analysis may be the seasonal analysis of the interesting data over a certain number of years and the analysis of an historical database to investigate the possible presence of not-stationary variabilities (seasonal, annual, astronomical, and so on).

6. REFERENCES

[1] Benedetto, S., Biglieri, E. and Castellani V. 1987. *Digital Transmission Theory*. Prentice Hall Inc.
[2] Colombo, F. 2000. *L'analisi spettrale dei dati climatici*. Rivista di Meteorologia Aeronautica.
[3] Gao, W. and. Li, B.W. 1993. *Wavelet analysis of coherent structures at the atmosphere forest interface*. J. Appl. Meteorol. 32. 1717-1725.
[4] http://www-stat.stanford.edu/~wavelab/
[5] Lo Presti, L. and Neri, F. 1991. *L'Analisi dei Segnali*. Edizioni C.L.U.T. Torino.
[6] Mallat, S. 1998. *A Wavelet tour of signal processing*. Academic Press.

[7] Mallat, S. 1989. *A theory for multiresolution signal decomposition: the wavelet representation*. IEEE Trans. Pattern Anal. Mach. Intell.. 11(7). 674-693.
[8] Mann, M. E. and Lees, J.M. 1996. *Robust estimation of background noise and signal detection in climatic time series*. Clim. Change 33. 409-445.
[9] Meyers, S., Kelly, B. and O'Brien, J. 1993. *An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves*. Mon. Wea. Rev. 121. 2858-2866.
[10] Olmo, G. 2000. *Dispense di TDS II*. Torino,.
[11] Pasero, E. and Moniaci, W. July 2004. *Artificial Neural Networks for Meteorological Nowcast*. IEEE International Conference on Computational Intelligence for Measurement Systems and Applications Boston. MA USA. 36-39.
[12] Yiou, P., Baert, E. and Loutre, M.F. 1996. *Spectral analysis of climate data*. Surveys in Geophisics 17. 619-663.

ACKNOWLEDGEMENT
The authors would like to thank Motorola for its financial support for this research.