Atmos. Chem. Phys., 11, 7905–7923, 2011
www.atmos-chem-phys.net/11/7905/2011/
doi:10.5194/acp-11-7905-2011

Atmospheric Chemistry and Physics

Technical Note: Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series

J. P. Musial[1,*], **M. M. Verstraete**[1], **and N. Gobron**[1]

[1]European Commission, Joint Research Centre, Institute for Environment and Sustainability, 21027 Ispra (VA), Italy
[*]current address: Geographisches Institut der Universität Bern (GIUB), 3012 Bern, Switzerland

Received: 26 February 2011 – Published in Atmos. Chem. Phys. Discuss.: 10 May 2011
Revised: 26 July 2011 – Accepted: 29 July 2011 – Published: 4 August 2011

Abstract. Geophysical time series often feature missing data or data acquired at irregular times. Procedures are needed to either resample these series at systematic time intervals or to generate reasonable estimates at specified times in order to meet specific user requirements or to facilitate subsequent analyses. Interpolation methods have long been used to address this problem, taking into account the fact that available measurements also include errors of measurement or uncertainties. This paper inspects some of the currently used approaches to fill gaps and smooth time series (smoothing splines, Singular Spectrum Analysis and Lomb-Scargle) by comparing their performance in either reconstructing the original record or in minimizing the Mean Absolute Error (MAE), Mean Bias Error (MBE), chi-squared test statistics and autocorrelation of residuals between the underlying model and the available data, using both artificially-generated series or well-known publicly available records. Some methods make no assumption on the type of variability in the data while others hypothesize the presence of at least some dominant frequencies. It will be seen that each method exhibits advantages and drawbacks, and that the choice of an approach largely depends on the properties of the underlying time series and the objective of the research.

1 Introduction

Time series analysis finds applications in a wide range of disciplines, from science to engineering and from marketing to econometrics; it naturally plays a critical role in geophysics, meteorology, hydrology, or the exploitation of remote sensing data. A time series is a finite, ordered set of couples of numerical expressions $\{(t_i, x_i); i = 0, 1, \ldots, n\}$, one providing a time reference and the other corresponding to the value of a measurement or observation acquired at that time. For conciseness, the sequence $\{x_i\}$ is often referred to as being the time series. Records collected by analog instruments typically yield continuous time series, but most frequently these series exist as finite sets of discrete records, either because they have been acquired in this way or because a continuous record has been digitized at a given temporal resolution. This paper only considers discrete time series.

Analyzing time series is simplified when the temporal sampling occurs at equally spaced time steps, and a host of techniques have been developed for complete and regular series. Researchers may also want to analyze related but independently acquired time series, and thus need to resample them on a common timeline, e.g., Mahecha (2010). Yet, actual time series turn out to be incomplete or unsuitable for standard analyses, either because some of the records may be missing (e.g., due to instrument failure or inadequate observing conditions), or because the records were originally acquired at unevenly distributed times. In addition, one might be interested in determining the likely value of the variable of interest at a time that may not coincide with a particular measurement or observation. For these reasons, it is useful to be able to generate reasonable estimates of the values of the variable of interest for arbitrary time references, including to replace missing values.

Multiple processes may simultaneously influence the values x_i recorded in the time series, although not all of them may be of interest. In many (but not all) practical cases, the broad, slow variations that offer some degree of predictability are of greater interest than the fast changes, which

Correspondence to: J. P. Musial
(jan.musial@giub.unibe.ch)

often appear as random, unpredictable events of lesser consequence, such as uncertainties in the measurements. By analogy with such fields as acoustics and radar, the interesting variations in the time series are called the "signal" and all other variations are referred to as "noise". It is clear that the presence of noise can interfere with the goal of accurately filling the gaps in a time series.

The standard approach to estimate the values of the variable of interest at arbitrary times, to separate the signal from the noise or to understand past or forecast future values of the series, calls for the determination of a mathematical model that captures the essential (physical or statistical) properties of the system. Although each of these three issues might be addressed separately, using different tools, it is apparent that the determination of an optimal underlying model should prove beneficial to address all these issues in a systematic and coherent manner.

The work described below has been motivated by interest in describing the phenology of terrestrial vegetation over wide areas, using satellite remote sensing measurements in the solar spectral region as the main source of information. Nowadays, such global data sets have been accumulated daily or weekly for periods of up to one or more decades. The accuracy of these measurements has improved in time, thanks to technological advances, nevertheless geophysical processes such as the ubiquitous cloud cover or the limited availability of solar radiation at high latitudes in winter seasons still result in a significant patchiness in the records.

Various researchers have addressed aspects of these questions (see, e.g., Moffat et al., 2007), but recent advances in the treatment of irregular time series (see, e.g., Hocke and Kämpfer, 2009; Kondrashov and Ghil, 2006) suggested to conduct an evaluation of some of the methods recently published or updated before pursuing a particular approach and investing considerable resources in the processing of large satellite databases. The purpose of this paper is thus to compare the performance of a few published modern methods to deal with the presence of gaps and noise in satellite data records and to report on such findings, which might be of interest to a wider scientific audience.

2 Outline of published approaches

2.1 Choosing an approach

Estimating the likely values of a time series at arbitrary times, for instance to replace missing data, is a particular case of the general problem of interpolation. The simplest approach might be to fit piece-wise linear functions between successive values of the time series. However, this method yields a very jagged series that may be continuous but not differentiable at each point in the original time series. It is also unlikely to provide reliable estimates: the values generated in this manner are always strictly bounded by the existing values in the time series and therefore tend to underestimate the "true" values, on average.

Another simple approach consists in fitting the Lagrange form of the interpolation polynomial through every record in a time series. For each data value x_i this process involves definition of a basis polynomial function which matches that point at given t_i and it is equal to 0 for all remaining t. All basis functions are then summed into a final form of the polynomial that provides a unique, smooth, differentiable solution everywhere. However, when the number of points in the time series increases, so does the order of the polynomial, which starts fluctuating wildly, not only between the observations but also outside the range of the time series, thereby making it inappropriate for most applications, including forecasting. In this case, the interpolated values may not be realistic and could take arbitrarily large values.

In both of these approaches, the interpolation problem has a solution and it is unique, but severe undesirable side effects limit or void its applicability. These simple underlying models (piece-wise linear functions or Lagrange polynomials) force the solution to match exactly each original record, which might excessively constrain the problem, especially given that original measurements or observations always include some level of uncertainty (e.g., due to the finite precision and accuracy of the instruments, calibration limitations, human errors, etc.).

A natural response to this issue is to relax the requirement on the model to match existing records and only insist that it takes on values that are "reasonably close" to these records whenever they are available, and to use a relatively smooth model formulation to catch the bulk of the variability of the time series. In the context of polynomials, this means using low-order functions. This approach clearly requires defining a measure of "goodness of fit" and a criterion to decide how close is "close enough". Also, since it might be unrealistic to globally fit a long time series exhibiting arbitrary fluctuations with a single smooth function, the interpolation may be performed on a local basis. Cubic splines have been developed and used in this context; their performance will be evaluated below.

An advantage of the methods discussed so far is that they make no assumptions about the underlying nature of the processes responsible for the variability exhibited in the time series. As a result, they can be applied to series of arbitrary complexity and work equally well if these underlying processes themselves change in time. The price to pay for this flexibility is that these approaches do not "learn" from the available records what might be the nature and properties of the processes responsible for the variations and thus exhibit little or no inherent forecasting skill.

An entirely different approach to this problem then consists in assuming that each of the relevant underlying processes can be represented by its own model, and that the entire time series can be reconstructed by a combination or superposition of these elementary models. To guarantee the

uniqueness of the solution, it is generally sufficient to select those constituent models from amongst a set of mutually orthogonal functions. Fourier (1822) appears to be one of the first researchers who developed the solution of a physical problem (the propagation of heat in a condensed medium) in the form of a superposition of trigonometric functions, opening the way to what is now known as spectral analysis. This method has proven extremely powerful and has been successfully applied in many fields of science, but works best to analyze time series that are clearly combinations of elementary periodic signals. When the fluctuations are aperiodic, and especially when they include random or unique events, the number of frequencies required to represent the time series becomes very large and the approach loses some of its appeal.

This drawback can be overcome, however, by selecting the elementary functions from a different set (or base), such as Legendre or Tchebicheff polynomials, or even as Empirical Orthogonal Functions (EOFs), which are an extension of the so-called Principal Component (or Factor) Analysis of the time series. In this latter case, the elementary functions are not explicitly prescribed a priori but are derived directly from the dataset.

Significant progress has been achieved over the last decade, so a modern approach in each of these categories will be tested below. The Lomb-Scargle method, specifically designed to retrieve the periodogram of time series acquired at unequally distributed instants is a modern application of the Fourier approach to arbitrary time-dependent records. It estimates the power spectrum of the time series without requiring the original data to be provided on a regular time grid or to be complete in any sense of the word. This method has been recently updated and applied to geophysical (or astrophysical) problems by Hocke and Kämpfer (2009). The Singular Spectrum Analysis (SSA) employed by Kondrashov and Ghil (2006) is a modern example of an approach capitalizing on the exploitation of orthogonal functions (EOFs, in this case) derived from the data themselves rather than imposing at the outset the form of the base models (e.g., trigonometric functions).

A key comparative advantage of these latter methods is that by "learning" about the underlying processes that control the evolution of the system and thus of the time series, these approaches may be quite suitable and efficient to predict the future evolution of that system, assuming of course that the same underlying processes will continue to play a similar role in the future. It will be seen that these methods are computationally much more demanding than the simpler approaches mentioned earlier.

2.2 The smoothing spline method

The polynomial smoothing spline method provides an attractive way of smoothing noisy data values observed at n arbitrarily located points over a finite time interval (Hutchinson and de Hoog, 1985). Described by Reinsch (1967), it is an extension of Whittaker (1923) spline. This method makes no assumptions on the underlying causes of the variations or on the mathematical structure of the series.

The smoothing spline constructs a continuous curve from segments of cubic polynomials joined together at knot points in such a way that the first and second derivatives of the resulting curve are continuous throughout. This method is applicable to a wide range of datasets because it is both flexible (i.e., it makes few assumptions) and adjustable through a single smoothing parameter λ, which controls the "stiffness" or "flexibility" of the spline curve. For small values of λ, the spline remains close to the data points, and in the limit case $\lambda \to 0$, the function simply interpolates the data. A contrario, larger values of λ increase the "stiffness" of the curve and in the limit case $\lambda \to \infty$, the spline becomes a linear least square fit. This simple method is robust and computationally inexpensive, so it is suitable to process large data sets.

Craven and Wahba (1979) proposed an objective method to determine an "optimal" value of the smoothing parameter, based on the minimization of the Generalized Cross Validation (GCV) procedure, which is a direct measure of the predictive error of the fitted line. GCV is calculated by removing each data point in turn, and forming a weighted sum of the square of the discrepancy of each omitted data point from a line fitted to all other data points (Hutchinson, 1998). The weights are evaluated as the inverse of the standard deviation applicable at each data point. To ensure reliable results with the GCV procedure, the time series should include at least 25 to 30 observations, according to Wahba (1990), and the noise level should not be highly correlated with the signal (Hutchinson, 1998). In this study, the smoothing parameter was evaluated dynamically using the IMSL (International Mathematics and Statistics Library) routine CSSMOOTH, as implemented in the IDL (Interactive Data Language) environment. This routine utilizes the GCV procedure proposed by Craven and Wahba (1979) and was used across all experiments.

2.3 The Singular Spectrum Analysis method

Kondrashov and Ghil (2006) proposed an approach to fill gaps in time series based on the Singular Spectrum Analysis (SSA) technique originally developed by Broomhead and King (1986) and Broomhead et al. (1987). This method incorporates elements from a wide range of mathematical fields including classical time series analysis, multivariate statistics and geometry, dynamical systems, as well as signal processing (Golyandina et al., 2001). It aims at describing the structure of the time series as a sum of simpler, elementary series describing features such as a trend, various oscillations and noise. The workflow of the SSA gap-filling and smoothing algorithm proceeds in four phases:

1. The first phase of the process, called embedding, involves the transformation of a one-dimensional scalar time series $\{x_i\}; i = 1, 2, \ldots, n$, into a multidimensional trajectory matrix of lagged vectors $\mathbf{X} = [X_1, \ldots, X_{n'}]$, where $n' = n - m + 1$ and each lagged vector is defined as $X_j = (x_j, \ldots, x_{j+m-1})^T$; $j = 1, \ldots, n'$. Each one of these vectors corresponds to a partial view of the original time series, seen through a window of length m. Choosing the most appropriate value for m, $(1 \leq m \leq n)$, is a matter of balancing the retrieval of information on the structure of the underlying time series, which would require larger values, and the degree of statistical confidence in the results, which is enhanced by using shorter but more numerous windows that repeatedly capture the notable features of the series (Ghil et al., 2002). The trajectory matrix $\mathbf{X}$ is thus a rectangular Hankel matrix of the form

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_{n'} \\ x_2 & x_3 & x_4 & \ldots & x_{n'+1} \\ x_3 & x_4 & x_5 & \ldots & x_{n'+2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ x_m & x_{m+1} & x_{m+2} & \ldots & x_n \end{pmatrix} \quad (1)$$

2. The second step consists in the Singular Value Decomposition (SVD) of the trajectory matrix $\mathbf{X}$ of size $m \times n'$, which is "decomposed" into a product of matrices $\mathbf{X} = U \Sigma V^T$ where U is a unitary matrix of size $m \times m$, Σ is a rectangular diagonal matrix of size $m \times n$ and V is a unitary matrix of size $n \times n$. The elements of Σ, called singular values, are the square roots of the eigenvalues of the covariance matrix $\mathbf{C} = \mathbf{X}\mathbf{X}^T$ of size $m \times m$. The rows of U are the eigenvectors of $\mathbf{X}\mathbf{X}^T$ and are often referred to as the left singular vectors or the Empirical Orthogonal Functions (EOFs) of the matrix $\mathbf{X}$. The columns of V^T are the eigenvectors of $\mathbf{X}^T\mathbf{X}$. If all eigenvalues are distinct, the solution is unique. Furthermore, if the eigenvalues are organized in decreasing order of magnitude, then any subset of the d eigenvectors (or EOFs), $1 \leq d \leq m$, for which the eigenvalues are strictly positive provides the best representation of the matrix $\mathbf{X}$ as a sum of matrices X_k, $k = 1, \ldots, d$ (Golyandina et al., 2001). The triplets composed of an eigenvalue and its associated left and right eigenvectors are called eigentriples of the trajectory matrix $\mathbf{X}$.

3. The third step involves the partitioning of these d eigentriples into p disjoint subgroups and summing them within each group, such that $\mathbf{X} = \sum_1^p \mathbf{X}_p$, where, ideally, the matrices $\mathbf{X}_p$ also have the structure of a Hankel matrix and thus correspond to the trajectory matrices of the hypothesized simpler series that combine to make the original time series. If these component series can each be described by distinct subsets of eigentriples, they are said to be separable by the SVD. In this case, the original time series can be described as a superposition of a trend, some harmonic oscillations and noise, for instance (Golyandina et al., 2001).

4. In practice, such an ideal situation rarely occurs and the component time series do not exactly match completely separate subsets of the eigenvectors of $\mathbf{X}$. The last step of the SSA algorithm, known as "diagonal averaging", aims at transforming the matrices $\mathbf{X}_p$ into Hankel matrices, which then become the trajectory matrices of the underlying time series, in such a way that the original time series can be reconstructed as a sum of these components. The entire procedure aims at defining in some optimal way what those components are.

The SSA gap filling method can be generalized to process spatio-temporal data or to regenerate missing values in multivariate time series. Here, only univariate time series were considered. We have implemented the code written in R by Lukas Gudmundsson, available from https://r-forge.r-project.org/projects/simsalabim/, and processed the time series described below using different window lengths and a variable number of leading EOFs.

The first step of SSA iterative gap filling algorithm includes centering the original time series on zero by subtracting the mean value of all its elements and zeroing the missing data values.

The inner loop of the SSA procedure (decomposition, grouping and reconstructing) is performed first on this centered, zero-filled time series. The missing values are replaced by computed values of the leading EOF and on this basis the first estimate of the first reconstructed component is generated. At the next iteration, the SSA algorithm is performed again to produce a second estimation of the first component on the basis of the new time series with missing values. replaced by the first estimation of the first leading component. Missing values replaced by the first estimate are now replaced by the second estimate of the first leading component. The convergence test between current estimation of the first component and the previous estimate is then carried out. If this test is positive then the inner loop stops and the first reconstructed component is returned.

In the outer loop the next leading EOF is added to the first reconstructed component. Then again the inner loop is performed until the convergence criterion is met and the best estimate of the second reconstructed component is returned. The third leading EOF is added in the same way and this process is carried out until the outer loop reaches the fixed number of analyzed EOFs.

Two main parameters are necessary to implement the SSA gap filling algorithm: window length m and maximum number of leading EOFs η, which create the η-th reconstructed component. The optimum combination of these parameters can be obtained by the cross-validation procedure, in which a fixed amount of available data is removed, then the SSA algorithm is performed and the RMSE (Root Mean Square Error)

between original dataset and each of the reconstructed components is calculated. This experiment is repeated several times with the same set of parameters to obtain mean values of RMSE over all experiments. Then the entire procedure is repeated with the same number of leading EOFs, but with different values of the window length (Kondrashov and Ghil, 2006). The values of parameters m and η corresponding to the case with the smallest RMSE among all cross-validation experiments is deemed optimal for the purpose of regenerating missing data.

The SSA gap filling algorithm is suitable for reconstructing time series with highly anharmonic oscillation shapes (Vautard et al., 1992) or nonlinear trends (Ghil et al., 2002). It can be economical in the sense that a small number of SSA eigenmodes may be sufficient to reconstruct the original time series. This is an advantage over traditional spectral methods based on the classical Fourier analysis, which typically require many trigonometric functions with different phase and amplitudes to provide a credible result. On the other hand, the high computational requirements of the SSA gap-filling algorithm may be a drawback in operational applications involving large numbers of time series. Other limitations of this method have been reported when the gaps in time series are long and continuous (Kondrashov and Ghil, 2006).

2.4 The Lomb-Scargle method

Hocke and Kämpfer (2009) used the Lomb-Scargle method to compute the periodogram of unevenly sampled time series and reconstructed the missing values in an astrophysical series from the amplitude and phase information of the dominant frequencies.

In practice, the first two steps of this procedure involve removing the mean value of original time series from each individual observation and applying a Hamming window to enhance spectral information. The Lomb-Scargle periodogram is then calculated, yielding a result equivalent to a linear least-squares fit of sine and cosine model functions to the observed time series (Lomb, 1976, Press et al., 1992, Hocke and Kämpfer, 2009).

Once the periodogram has been retrieved, the signal is reconstructed by considering only its most significant components, i.e., those associated with a power larger than a given threshold. The latter can be estimated either on the basis of a confidence level analysis or simply set to a fixed fraction of the largest peak in the periodogram. The reverse Hamming window procedure is applied and the final result is of course a continuous and complete series, which can be resampled at any desired frequency.

While the Hamming window procedure improves the performance of the algorithm in the bulk of the time series, it also results in poorer results near either end of the series than away from these borders. This can be remedied, however, by applying a Kaiser-Bessel window instead (Harris, 1978), as it features an adaptable shape parameter.

The optimal value of this shape parameter is obtained iteratively by calculating the RMSE for each smoothed and gap filled time series against the original one and selecting the parameter value corresponding to the smallest RMSE. For the purpose of this analysis, we have converted Hocke's 2007 MatLab code (available as on-line supplement from http://www.atmos-chem-phys.org/9/issue12.html) to the IDL language. According to (Hocke and Kämpfer, 2009), this approach should be suitable to process either periodic or non-periodic time series.

3 Methodology

3.1 Choosing a quality fit criterion

An important methodological issue that requires careful attention is the selection of a measure of "goodness of fit" between the models and the data (time series), and of a criteria to judge when this measure is "good enough" for the stated purpose.

The root mean square error (RMSE, or deviation RMSD) has traditionally been used in this context because it enjoys well-understood and desirable statistical properties. This measure is defined as the square root of the mean square error, or the square root of the sum of the squares of the differences between the model predicted values $y_i = y(t_i)$ and the observations $x_i = x(t_i)$ recorded in the time series:

$$\mathrm{RMSE} = \left[\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2 \right]^{1/2} \tag{2}$$

where n is the number of points in the time series. The square of the deviations between the model and the records prevents errors of different signs to compensate each other, and enhances the role of large deviations compared to smaller ones. However, Willmott and Matsuura (2005) have argued that the Mean Absolute Error (MAE) provides a better indicator of the quality of the fit or the performance of the model in representing a given data set. MAE is defined as follows:

$$\mathrm{MAE} = \left[\frac{1}{n} \sum_{i=1}^{n} |y_i - x_i| \right] \tag{3}$$

Both statistics measure the difference between modeled values and the corresponding observations, assuming the latter are reliable, with larger values indicative of a worst fit; they differ in the emphasis they give to particular situations: RMSE penalizes large individual differences while MAE focuses on the mean overall performance. More importantly, RMSE values are not representative of mean or typical errors only: they range between MAE and $n^{1/2} \times \mathrm{MAE}$, increase non-monotonically with MAE and vary with the square root of n (Willmott and Matsuura, 2005). MAE will thus be used in this evaluation.

Another criteria that characterizes the fit of a model to observations is the mean bias error (MBE) or mean error (ME). It assess on average if the model tends to overestimate or underestimate reconstructed values. It is defined as a mean value of residuals:

$$\mathrm{MBE} = \left[\frac{1}{n} \sum_{i=1}^{n} y_i - x_i \right] \tag{4}$$

The chi-square test as opposed to MAE and MBE is a categorical score which determines whether the difference in distributions of observed and predicted values in one or more classes are statistically significant. Predicted values are based on particular theoretical distributions which fulfill a null hypothesis. In an univariate case, the categorization of data can be achieved by computation of a cumulative distribution function (CDF) or a histogram of values. The minimum frequency (number of elements) in each class should be greater than 5. The chi-square statistic is defined as:

$$\chi^2 = \left[\sum_{i=1}^{c} \frac{(f_o - f_e)^2}{f_e} \right] \tag{5}$$

where c is number of classes, f_o is observed frequency, and f_e is expected frequency. The probability level p associated with χ^2 is derived from the theoretical chi-squared distribution accounting for number of degrees of freedom df defined as: $df = c-1$. Commonly, p-values lower than 0.05 indicate that differences between frequency distributions of observed and predicted values are statistically significant and the null hypothesis should be rejected.

While comparing the quality of fit between modelled and observed values it is recommended to test if residuals are correlated within certain sampling lags. This can be achieved by means of autocorrelation function defined as:

$$r(l) = \left[\frac{\sum_{i=1}^{n-l} (e_i - \mathrm{MBE})(e_{i+l} - \mathrm{MBE})}{\sum_{i=1}^{n} (e_i - \mathrm{MBE})^2} \right] \tag{6}$$

where l is a lag size defined as: $l = 0, 1, \ldots, l_{\max}$; e_i is a residual between observations and modelled values $e_i = y_i - x_i$. According to Box and Jenkins (1976) the number of samples n in a time series should exceed 50 and $l_{\max}$ should not be larger than $n/4$. If residuals e_i are randomly distributed then $r(l)$ is equal to 0 or it is enclosed within confidence limits $0 \pm \frac{z_{1-\alpha/2}}{\sqrt{n}}$, where z is the quantile function of the standard normal distribution and α is the significance level. Serial correlation in residuals may imply that the fitted model is misspecified or it fails to reconstruct some periodic fluctuations.

3.2 Smoothing over noise

Noise due to errors of measurement or uncertainties will impact the ability and effectiveness of a method to reconstruct values that may be missing from time series. As hinted earlier, the underlying idea is to process the data in such a way that typically high-frequency random variations considered as noise are filtered out while low-frequency changes are left unaffected. In the case of spectral methods, this is most easily implemented by decomposing the original series in terms of a power spectrum and reconstructing the signal using all frequencies lower than some given threshold. The sensitivity of the methods to the presence of noise will be documented in the tests below.

3.3 Designing artificial test cases

A large set of test cases was constructed to evaluate the performance of the approaches described above when either the number of missing observations or the level of noise in the data increases. The idea of these tests is to generate complete time series representing the "truth", altering them by imposing data gaps and adding noise, and then analyzing these modified series with the methods described earlier to assess to what extent they are capable of generating reasonable values to replace the missing ones. Three different "base" signals were considered: a single sine wave (Eq. (7), Fig. 20), a superposition of three sine waves (Eq. (8), Fig. 21) and an aperiodic signal (Eq. (9), Fig. 19), respectively. They intend to represent functions resembling typical geophysical signals of increasing complexity:

$$x_1(t) = 0.5\sin(t - \pi/2) + 0.5 \tag{7}$$

$$x_2(t) = 0.28\sin(t - \pi/2) + 0.19\sin(2t - \pi/2) + 0.16\sin(0.5t - \pi/4) + 0.6 \tag{8}$$

$$x_3(t) = 0.35\sin(t - \pi/2) + 0.15\sin(20\sqrt{t} - \pi/2) + 0.5 \tag{9}$$

While these equations represent continuous time series, we generated discrete base series by sampling these functions in such a way that the argument t of the sine functions is successively incremented by 10 degrees over a total of 10 full cycles (simulated years), thus creating time series of $n = 361$ data points. These series were then "degraded" by introducing variable amounts of gaps and adding different levels of noise as follows.

3.3.1 Gaps

The following three types of gaps (missing data) were considered:

- Uniformly distributed gaps (Fig. 20). For each predefined percentage of missing data, a random number generator $\mathcal{U}$ was used to iteratively select the location of the next data point to be removed from the series: $x_m = \mathcal{U}[0,1] \times n$. This situation might arise when the system of interest is occasionally unobservable, for instance due to the presence of clouds, when analysing satellite data.

- Seasonal gaps (Fig. 19). In this case, the desired percentage of missing data was imposed by removing, from

each cycle in the time series, the required number of points around the lowest data values. In reality, this case may occur at high latitude because a lack of solar irradiance (or the presence of snow) might systematically prevent the acquisition of usable observations during the winter.

– Prolonged gaps (Fig. 21). For this scenario, a single continuous period of missing data was imposed in the middle of the time series, with a total length set to correspond to the predefined percentage of gaps desired. This pattern would emerge if the observing instrument failed to operate correctly for some time, for instance.

The results described below only refer to the performance of the methods to generate reasonable values in the artificial data gaps within the period of 10 cycles: no attempt was made to extrapolate beyond either end of the original series.

3.3.2 Noise

The simulated noisy data value $x^*(t)$ corresponding to the time series value $x(t)$ were estimated as $x^*(t) = x(t)[1 + N(0,1)S]$, where $N(0,1)$ represents a normal (Gaussian) distribution of mean 0.0 and standard deviation 1.0, and where S is a scaling factor to create different noise levels. In this process, only those values of $N(0,1)$ falling within the range $[-1.0, 1.0]$ were considered.

3.4 Using actual time series

In addition to these artificial test cases, the approaches will also be evaluated against a few actual time series, modified by adding gaps as before. The following cases were considered:

1. Some 400 time series of the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), derived from an analysis of SeaWiFS data, generated as part of the CarboEurope project database, were downloaded from the JRC FAPAR web site (http://fapar.jrc.ec.europa.eu/). The SeaWiFS scanner instrument has been in operation since September 1997. Complete data coverage for large areas requires compositing the daily acquisitions over longer periods, and standard products are typically generated every 10 days or monthly. The processing steps required for generating these products are described in Gobron et al. (2006). These time series are typical of sequences that may exhibit gaps due to cloudiness or lack of sunlight during the winter.

2. Sunspots have been observed by astronomers for centuries, and include the well-known 11 year cycle, although other fluctuations considerably affect the number of observable spots at anyone time. For the purpose of this study, the monthly number of Sunspots for the period 1900–2009 were obtained from the National Geophysical Data Center (NGDC) of the US National Oceanic and Atmospheric Administration (NOAA) (http://www.ngdc.noaa.gov/). This series exhibits strong periodicities but also some degree of unpredictability.

3. The record of atmospheric carbon dioxide (CO_2) concentration, in parts per million per volume, obtained at Mauna Loa constitutes probably one of the most emblematic time series of our times, as it unequivocally shows how human consumption of fossil fuels modifies the composition of our atmosphere. The values used here were downloaded from the Carbon Dioxide Information Analysis Center (CDIAC) at the Oak Ridge National Laboratory (ORNL) (http://cdiac.ornl.gov/). The main advantage of this series in the current context is that it exhibits a strong trend (itself somewhat variable in time) as well as a clear seasonal signal.

4. The Dow Jones Index (DJI) is clearly not a geophysical time series, but it offers the distinct advantage of being a non-periodic signal, with wild fluctuations and a strong overall trend. The mean weekly values of DJI were obtained from a public domain source (http://finance.yahoo.com/) for the period 1981 to 2009.

4 Numerical experiments and results

4.1 Numerical experiments with artificial time series

In the artificial time series experiments, the accuracy of the selected algorithms was estimated by calculating the statistics of the quality fit criteria (MAE, MBE, χ^2, p-value, $r(l)$) between series reconstructed by each of the methods described above and the original smooth dataset (without noise and gaps). Every experiment was repeated 10 times with the same set of parameters (data type, gap pattern, levels of noise, amount of gaps), but with different seed values for the pseudorandom number generator, thereby generating 10 different data sets with the same model parameters and different noise patterns. The results reported here exhibit the average of these 10 values, a compromise between available computational resources and the desire to establish reasonably stable results. This exercise yielded an extensive set of statistics computed as a function of five variables: gap-filling method, data type, gap pattern, amount of gaps, and noise level. As far as chi-squared test is concerned the frequencies of data values were derived from the histogram function with 0.05 bin size. Expected frequencies were obtained from the complete and noise-free time series generated according to Eqs. (7), (8), (9). The p-values reported correspond to one-tailed probability of obtaining a value of χ^2 or greater with 20 degrees of freedom.

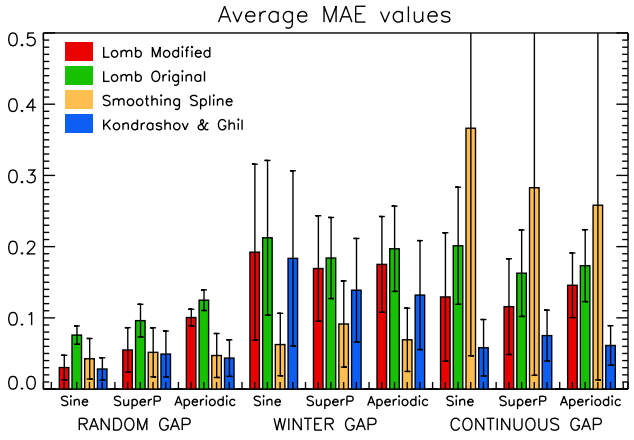

Fig. 1. Average values and standard deviations for MAE derived from an extensive set of the artificial time series. See text for details.

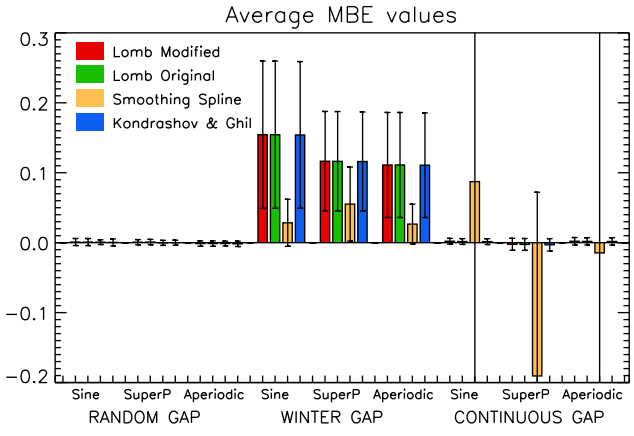

Fig. 2. Average values and standard deviations for MBE derived from an extensive set of the artificial time series. See text for details.

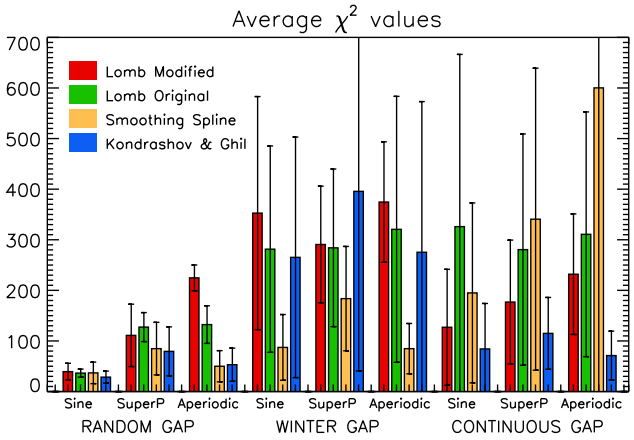

Fig. 3. Average values and standard deviations for χ^2 derived from an extensive set of the artificial time series. See text for details.

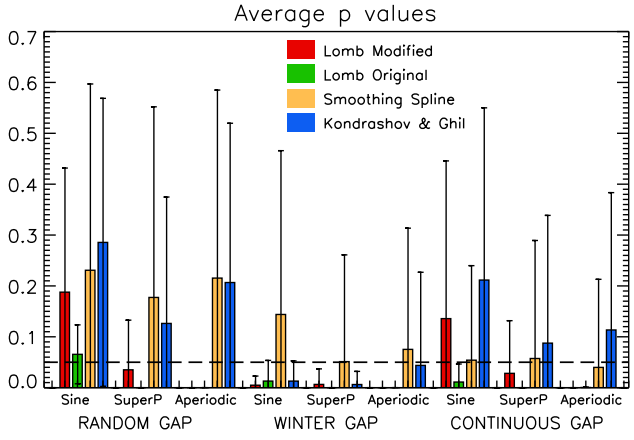

Fig. 4. Average values and standard deviations for p derived from an extensive set of the artificial time series. The dashed line indicates 0.05 probability level. See text for details.

4.1.1 Overall results classified by type of gap

Figures 1 to 4 summarize the results for all experiments with different amounts of data gaps and noise levels. Overall, the selected algorithms reconstructions reveal the best agreement with original datasets (smallest MAE, MBE and χ^2 values, with the highest relative probability of obtaining such results) in case of random gaps scenario. This gap pattern clearly have a less drastic effect on the signal reconstruction than seasonal or very long gaps. In other words, when missing values are sprinkled throughout the time series and do not excessively mask the underlying signal, the remaining data points still carry enough information to reconstruct a reasonably accurate version of the series.

Whenever missing values are clustered seasonally (as in winter gaps, for instance) the signal becomes severely corrupted because there is a deficit of information for some ranges of frequencies and the power spectrum cannot be reliably estimated. In this case, minor fluctuations in the data can induce the presence of spurious peaks when reconstructing the time series with the Lomb-Scargle method, and even the Kondrashov-Ghil algorithm is sensitive to this type of gaps (Figs. 12 and 19). The smoothing spline method is not affected by this problem since it does not rely on the power spectrum of the signal.

On the other hand, in the case of the "continuous gap" scenario (Fig. 21), the smoothing spline algorithm is not able to take advantage of the power spectrum information recovered from the rest of the time series to fill the large gap. The polynomial function adopts various shapes, depending on the distribution of the few points immediately before and after the continuous gap. The other approaches perform much better in this case, as can be seen from the MAE statistics reported in Fig. 1.

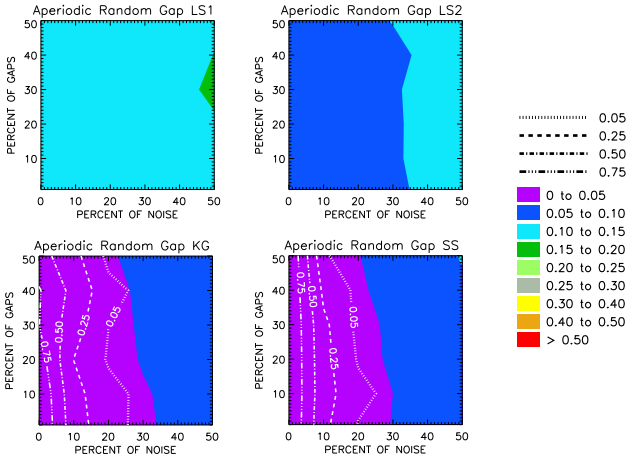

Fig. 5. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, aperiodic time series with random gap patterns: **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov and Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

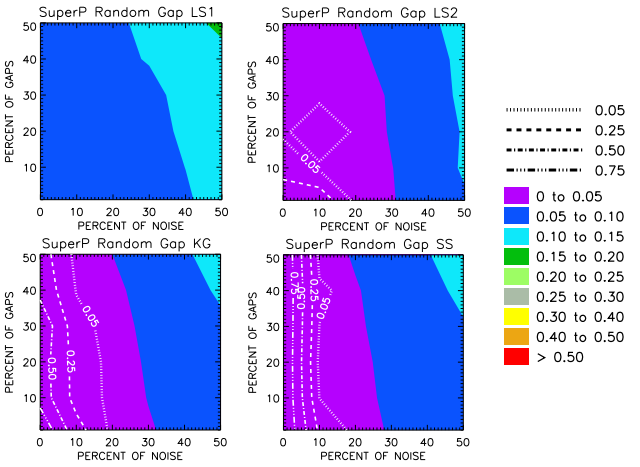

Fig. 6. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, superimposed time series with random gap patterns: **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov-Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

4.1.2 Overall results classified by type of data pattern

The underlying structure of the data also influences the performance of the time series reconstruction in the presence of gaps and noise. Our tests included both periodic (Figs. 20 and 21) and aperiodic times series (Fig. 19), and methods that essentially assume the existence of significant periodic fluctuations in the data (such as the Lomb-Scargle algorithm) naturally experience greater difficulties in regenerating reliable values for aperiodic time series (Hocke and Kämpfer, 2009). This effect is illustrated in Fig. 4 where in all aperiodic test cases Lomb-Scargle algorithm produced results that were significantly different from the original data and thus the p-values are equal to 0. The correlation of residuals is also stronger for this method (Fig. 22). By contrast, the data pattern does not seem to substantially influence the MBE values (Fig. 2).

Interestingly, statistics revealed some of the poorest results in the case of strictly periodic dataset composed of a superposition of several sine waves. This might be due to the presence of small second maxima (Fig. 21) which is hard to reconstruct when noise level and gaps amount are high. Therefore, the quality of the reconstruction typically can be affected by the shape of a particular time series.

4.1.3 Detailed results classified by type of gap

The performance of each gap-filling and noise-reducing technique was investigated against different distributions of data gaps and noise. The graphical representation of the results yields a set of contour plots (Figs. 5 to 10) depicting the distribution of MAE and p-values for the indicated method as a

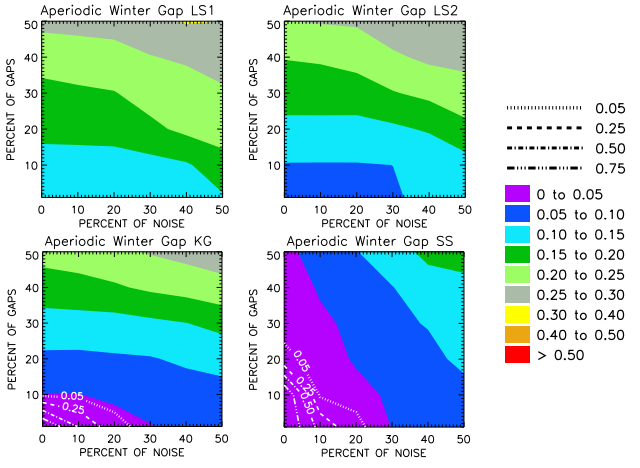

Fig. 7. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, aperiodic time series with winter gap pattern. **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov and Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

function of gap and noise percentages. Each plot was generated by linearly interpolating a 6×6 grid (hence the sometimes jagged appearance) composed of data points which represent an average of 10 independent experiments. Both axes have 10 % interval increments. The color scheme corresponds to fixed ranges of MAE values, as shown in the legends. The interpretation of these graphs should consider both the absolute values and the orientation of isolines. Vertical patterns (Fig. 5) indicate that the reconstruction of a

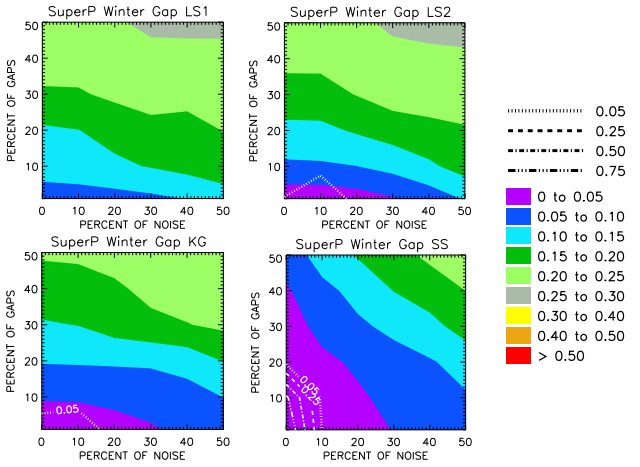

Fig. 8. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, superimposed time series with winter gap pattern. **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov and Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

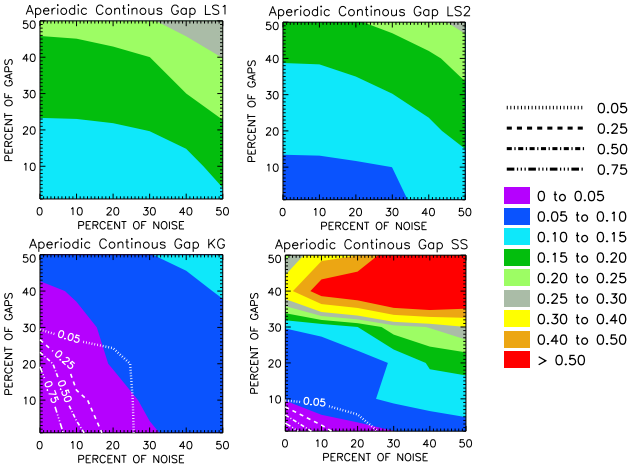

Fig. 9. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, aperiodic time series with continuous gap pattern. **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov and Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

time series by a particular algorithm is more depended on the noise level than on the amount of data gaps. The reverse situation takes place when the isolines are horizontal (Fig. 7a, b, c). When the graph exhibits a diagonal pattern (Fig. 7d), noise level and data gaps both contribute to a degradation in the quality of the reconstructed time series.

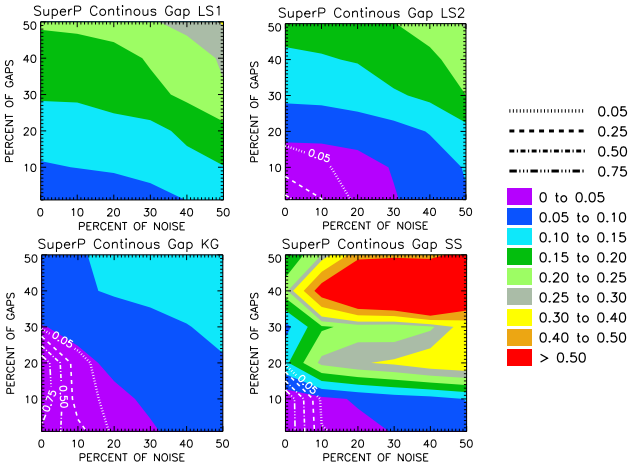

Fig. 10. The distribution of MAE and p values as a function of data gaps and noise level, acquired during the experiment with artificial, superimposed time series with continuous gap pattern. **(a)** original version of Lomb-Scargle algorithm (LS1), **(b)** modified version of Lomb-Scargle algorithm (LS2), **(c)** Kondrahov and Ghil algorithm (KG), **(d)** smoothing spline algorithm (SS).

– The distribution of MAE in the *random gap* scenario mainly depends on the level of noise (Figs. 5b, c, d and 6). The statistical significance of the differences in reconstruction p derived from the chi-squared test follows the same vertical pattern. All p-values lower than 0.05 indicate that the fitted model is significantly different from the original complete and noise-free dataset. Even when many data points are missing, most of selected algorithms perform well in reconstructing the original datasets provided the noise level remains low. It can also be seen that selecting a Kaiser-Bessel window instead of a Hamming window in the Lomb-Scargle algorithm significantly improves the results (Figs. 5a, b and 6a, b). However this technique is inferior to others when the oscillations are not periodic resulting in p-values equal to 0 for all combinations of noise and gaps levels (Fig. 5). Amongst all random gap scenarios, the Kondrashov-Ghil algorithm yielded the smallest MAE with relatively high confidence levels across all types of artificial time series (Figs. 1, 5c and 6c), though the differences in datasets reconstructions (and MAE values) between this method and the smoothing spline technique are very small (Figs. 1, 5c, d and 6c, d). The autocorrelation of residuals (Fig. 22) for these two methods is significant only for small lag size (5–6 data points) although slightly better results are obtained by the latter algorithm. Nevertheless, all selected techniques in this particular gap scenario provide residuals that are truly unbiased (Fig. 2).

– In the *winter gap* scenario, the MAE of the time series reconstruction by means of Kondrashov-Ghil and Lomb-Scargle algorithms is relatively high due to the appearance of spurious, intermittent peaks which occur in the reconstructed time series where the gaps used to be (Figs. 7 and 8). Analogically to the *winter gap* scenario one example of the *summer gap* scenario was generated in order to verify if these peaks appear and if they are negatively directed towards the original data. Figure 19 depicts both cases and it could easily be seen that in the case of *winter gap* scenario Kondrashov-Ghil and Lomb-Scargle will overestimate reconstructed values (Fig. 2) whereas in *summer gap* scenario they will underestimate them. The systematic presence and duration of data gaps in the input time series strongly influence the performance of these two algorithms (horizontal MAE isolines). However, the relative contributions of data gap percentages and noise levels in the total MAE and probability p values are more balanced (diagonal pattern, Figs. 7d and 8d) in the case of smoothing spline algorithm. The latter approach is unquestionably the most appropriate method for gap-filling and smoothing of datasets with *winter* and *summer gap* patterns, as it does not introduce spurious, intermittent features. Also the autocorrelation of residuals derived from the fitted model is weaker (Fig. 22).

– In the *continuous gap* scenario, the distribution of MAE depends more on the size of the gap than on the level of noise (Figs. 9 and 10). The smoothing spline method gives extremely odd results (Figs. 9d and 10d) because it fits a polynomial function only to the few existing data points at both ends of a large gap. Thus even averaging the results over 10 experiments does not yield stable results or a smooth contour plot. In this case, the best fit of reconstructed series to the original dataset is obtained by the Kondrashov-Ghil algorithm (Figs. 9c and 10c). However, the autocorrelation of residuals for this method is relatively significant (Fig. 22).

4.2 Numerical experiments with actual time series

Time series acquired directly from an instrument or retrieved from observational data automatically include the noise inherent to the instrument and retrieval methods, so no additional noise was added in these experiments. Similarly, since FAPAR time series naturally contain gaps, no further manipulation of these series took place (neither noise nor data gaps were introduced), and the goal was to compare the methods described above when applied to the same time series. The computation of the MAE and MBE values had to be modified, though, to include only those points for which data were available initially. New values can be generated by the algorithms within the gaps, but one must assume that the accuracy of the latter is similar to that during the times when data are available.

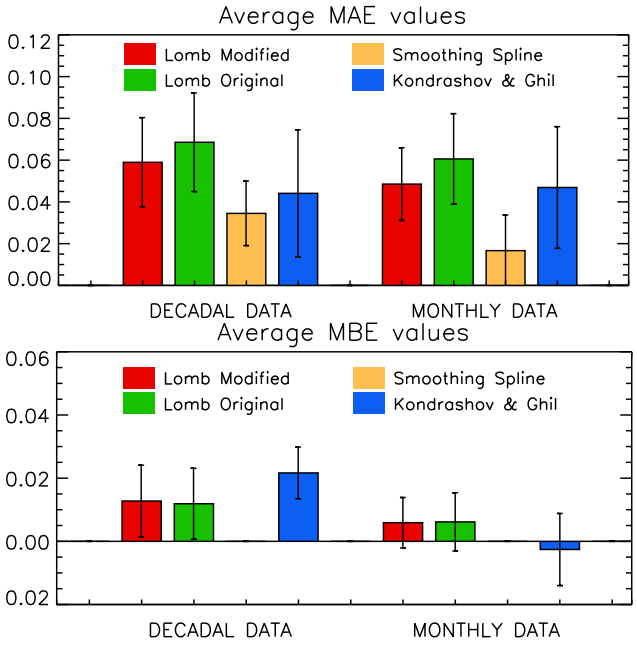

Fig. 11. Average values and standard deviations for MAE and MBE acquired during the experiment with FAPAR dataset. See text for details.

In the case of the atmospheric CO_2, sun spots and Dow Jones time series, the original records were complete so the modifications included only the artificial introduction of random, uniformly distributed gaps, and the MAE and MBE were computed as done previously with artificial time series.

In evaluating these experiments, it is important to note that we have compared the four gap-filling algorithms as such, without any manual adjustment. It may be feasible to "tune" each method to yield better results by tweaking individual parameters, but the results would then depend on each particular record and on the amount of time and energy spent in tuning. Our aim is to evaluate the performance of generally applicable methods that could be applied automatically to a large number of time series, without any human intervention.

4.2.1 Results from the SeaWiFS FAPAR time series experiment

The outcome of the experiment with FAPAR time series at monthly and decadal time resolutions is presented in Fig. 11. The patchiness of FAPAR datasets reflects the combination of gap patterns previously tested in the experiment with artificial time series. Missing values originate mainly from the temporary appearance of clouds or snow, though some are

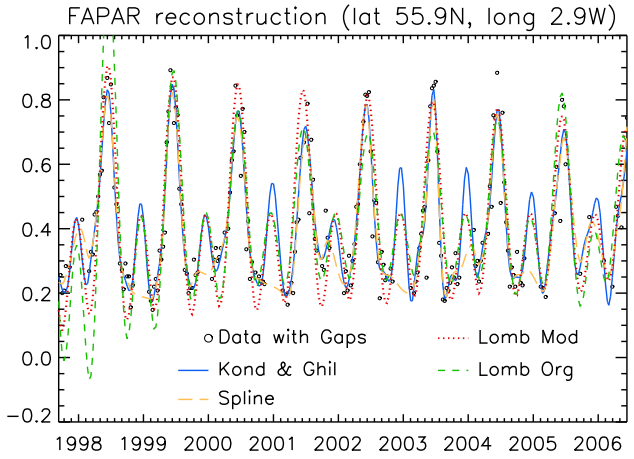

Fig. 12. Example of reconstruction of a FAPAR time series using the indicated methods. The dataset was derived from the SeaWiFS sensor as a decadal mean of FAPAR for a single pixel located around East Saltoun, UK (55.9° N, 2.9° W). Neither artificial gaps nor noise were introduced in this case, because the original time series contains these distortions. Some of the approaches generate spurious peaks in the winter due to lack of constraint. See text for details.

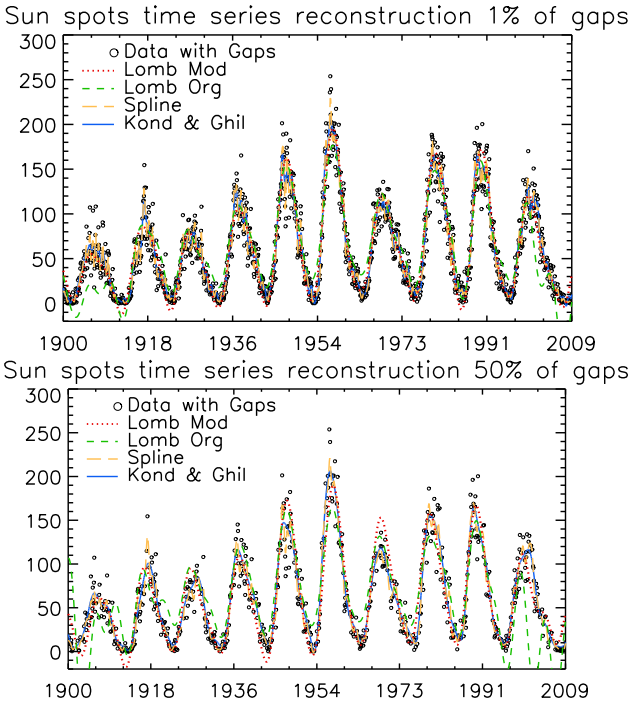

Fig. 13. Example of the reconstruction of sun spots time series using the indicated methods. The proportion of missing points is equal to **(a)** 1 % and **(b)** 50 %, no noise was introduced. See text for details.

related to the lack of solar illumination at high latitudes during the winter period (Fig. 12).

The reconstruction based on the Lomb-Scargle algorithm generates the highest MAE values because of the presence of aperiodic components in the original signals. These phase- and amplitude-modulated oscillations derive from natural and human-induced constrains such as the variability of incoming solar irradiance at the surface, a time-evolving supply of water and nutrients in the soil, the strong dependency of biochemical growth and development processes on ambient temperatures, agricultural practices, etc. (e.g., Verstraete et al. (2008)). The Kondrashov and Ghil method turns out to be more accurate than both Lomb-Scargle algorithms, but the best fit to the original data is produced by the smoothing spline algorithm. As opposed to other techniques it provides unbiased results with extremely small MBE values. While data availability at higher frequencies may generally improve the ability of a spectral method to reconstruct the signal, using temporally-averaged values tend to decrease the noise level. This can be seen in Fig. 11, which shows that reconstruction results based on monthly data tend to be marginally better than those based on decadal data, without altering the ranking of the methods. It may also be easier to fit a smaller set of points (fewer constraints), especially for the smoothing spline method, where number of nodes is a crucial variable.

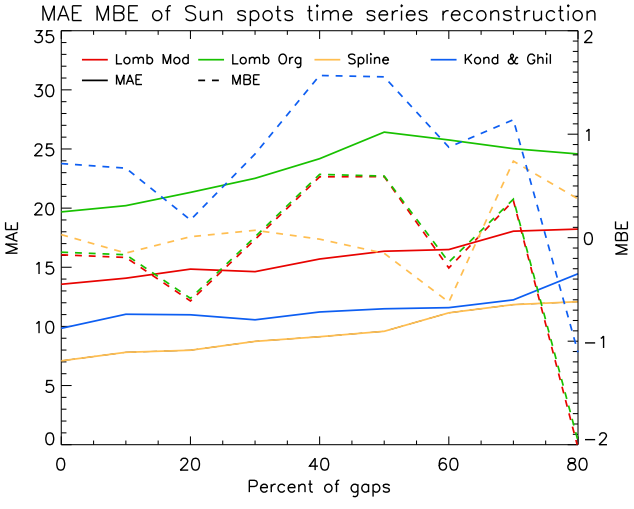

Fig. 14. Distribution of MAE and MBE values as a function of data gaps quantity, acquired during the experiment with sun spots time series. See text for details.

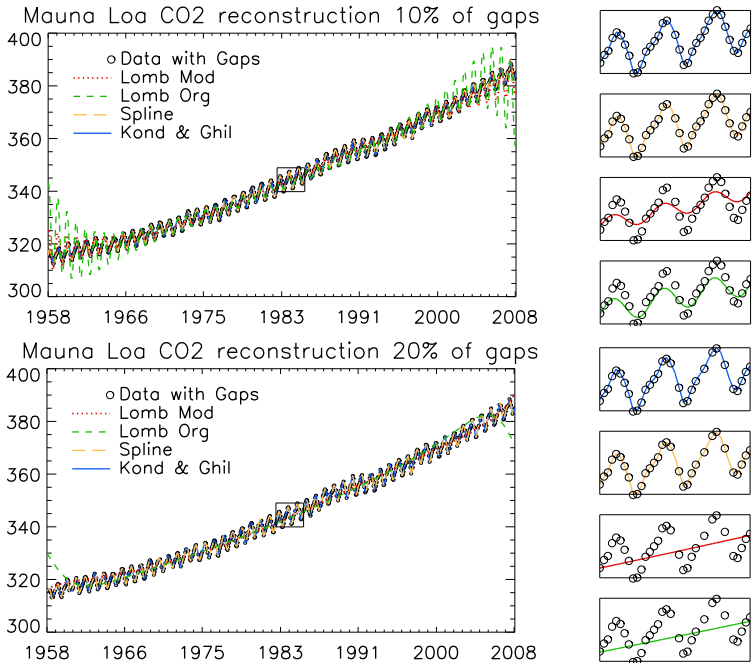

Fig. 15. Example of the reconstruction of the CO_2 record, expressed in ppm, from Mauna Loa station acquired by means of the indicated methods. The data fragment in the rectangle has been enlarged on the right. The fraction of missing points is equal to **(a)** 10 % and **(b)** 20 %, no noise was introduced. See text for details.

4.2.2 Results from the Sunspots time series experiment

The results of reconstructing the historical record of monthly sunspot numbers for the period 1900–2009 where some 50 % of the data have been artificially removed can be seen in Fig. 13. Clearly, all methods do a fair job in detecting and properly representing the large 11-yr cycle present in this time series, though the performance of these algorithms is actually quite variable. The Kondrashov-Ghil and the smoothing spline approaches exhibit similar MAE and MBE over all experiments, while the Lomb-Scargle algorithm does not appear to do so well, especially near either end of the record. However, it should be noted that this latter method smooths the reconstructed curve (Fig. 14) more extensively than the other techniques, which generates higher MAE values.

4.2.3 Results from the Mauna Loa CO_2 time series experiment

The famous record of monthly concentration of atmospheric CO_2 measured by Keeling et al. (1996) at Mauna Loa since 1958 constitutes another emblematic and very useful time series to test these algorithms, because it includes a powerful trend as well as a clear seasonal signal, both of which are somewhat variable in time. This results in significant power at very low frequency and smaller peaks to represent the seasonal fluctuations and their variations. The time series for the period 1958–2008 was artificially manipulated by introducing gaps, as explained earlier, to evaluate the capacity of the gap-filling methods to reconstruct a reasonable approximation of the original record.

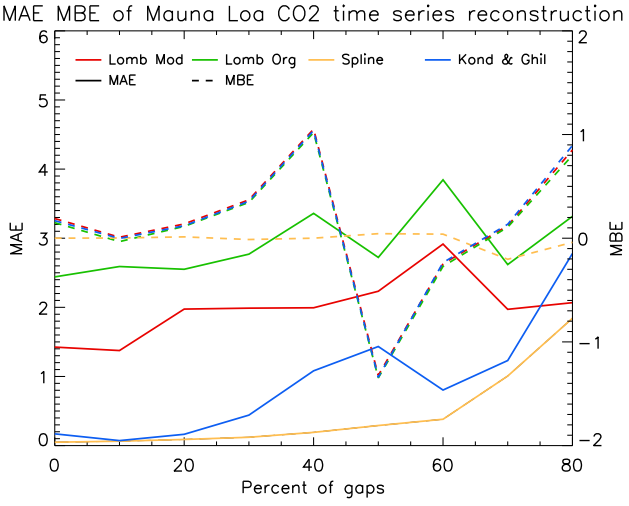

Fig. 16. Distribution of MAE and MBE values as a function of data gaps quantity, acquired during the experiment with the time series of CO_2 record from Mauna Loa station. See text for details.

In their original analysis, Hocke and Kämpfer (2009) proposed two approaches to select the relevant spectral components to use with the Lomb-Scargle algorithm (a fixed threshold, set as a fraction of the highest peak in the power

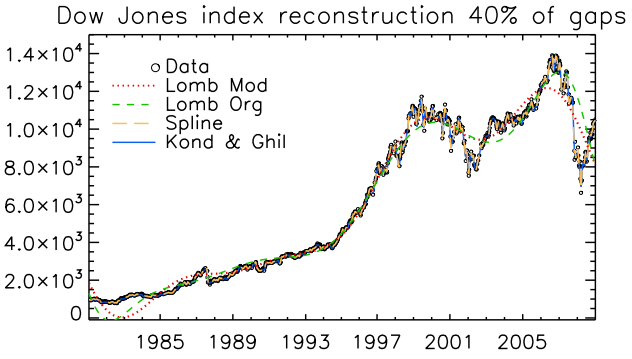

Fig. 17. Example of the reconstruction of Dow Jones time series acquired by means of the selected methods. The amount of missing points is equal to 40 %, no noise was introduced. See text for details.

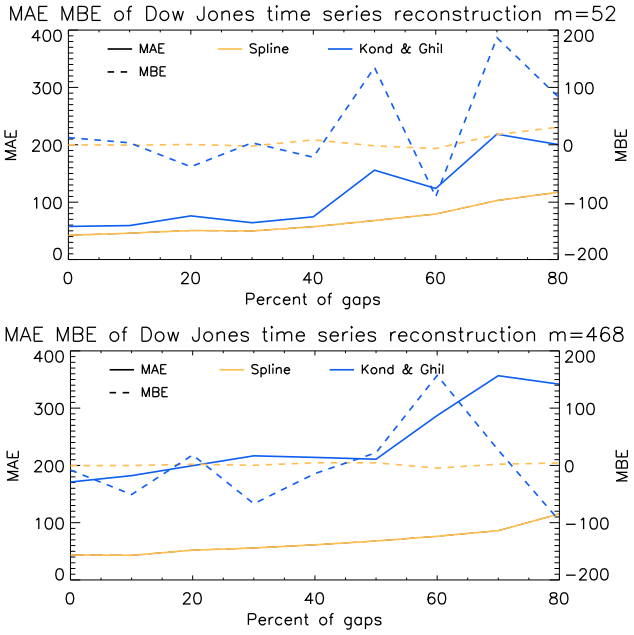

Fig. 18. Distribution of MAE and MBE values as a function of data gaps, acquired during the experiment with the time series of Dow Jones index. The SSA window size is 468 points (bottom panel) and 52 points (top panel).

spectrum or a statistical confidence analysis). Both approaches were tested in this experiment, and it turns out that, for small proportions of data gaps, the Lomb-Scargle method coupled with threshold set by statistical confidence was able to recover the annual cycle of the time series (Fig. 15, top panel). However, when the percentage of data gaps exceeds 10 %, the strength of the annual peak in the power spectrum decreases enough to become statistically insignificant and the Lomb-Scargle method only retrieves the main trend (Fig. 15, bottom panel). In both cases, the MAE associated with the Lomb-Scargle method remains significantly higher than the MAE for the Kondrashov-Ghil and smoothing spline methods, which perform almost equally well (Fig. 16).

In truth, Hocke and Kämpfer (2009) do recommend to remove any trend before applying the Lomb-Scargle approach, to process the residuals and then to add the trend back to the processed data. We have not implemented such a preprocessing step (which might be implemented in a variety of ways) in this exercise because it could introduce a variable and somewhat arbitrary bias in the comparison of gap-filling and smoothing algorithms.

4.2.4 Dow Jones index

The last test involves the long record of weekly values of the Dow Jones Index (DJI) for the period 1981 to 2009. This time series is very irregular due to the wide range of economical factors affecting this index. DJI is definitely not a periodic time series, so that the Lomb-Scargle method is not expected to be appropriate: In fact, it does capture the broad features, but not the smaller scale and somewhat arbitrary fluctuations. Figure 17 shows to what extent the four methods are capable to represent the overall time series.

The Kondrashov-Ghil method can account for such an unusual signal, though its performance depends strongly on the length of the window. If this parameter is large compared to the entire record, say 468 data points (equivalent to 9 years), the reconstruction is rather smooth but the MAE remains relatively high (see Fig. 18a). A shorter window, for instance of 52 data points (1 yr), yields results similar to the smoothing spline approach (Fig. 18b). As explained earlier, the length of the window controls the degree of smoothing of this algorithm and choosing the most appropriate value depends on the nature of the problem at hand (Golyandina et al., 2001). If the typical main periodicities were known or at least expected in previous experiments (e.g., an annual cycle), no such information can be assumed in this case. The choice of a window length is thus somewhat more arbitrary, and although this parameter can be optimized during the cross validation (CV) procedure, that step is computationally demanding and time consuming for such long time series (more than 1500 points). If the window length parameter is specified explicitly instead, that fact and the chosen value should be notified explicitly since it does significantly affect the results of the reconstruction.

5 Discussion

Gap-filling and smoothing unevenly sampled and noisy time series is a common and necessary process, especially in the analysis of geophysical signals. Various approaches to this problem have been proposed, but the choice of the most appropriate method for a particular dataset is non trivial. In this paper, three gap-filling techniques (one of them in two variants) were evaluated on the basis of experiments with

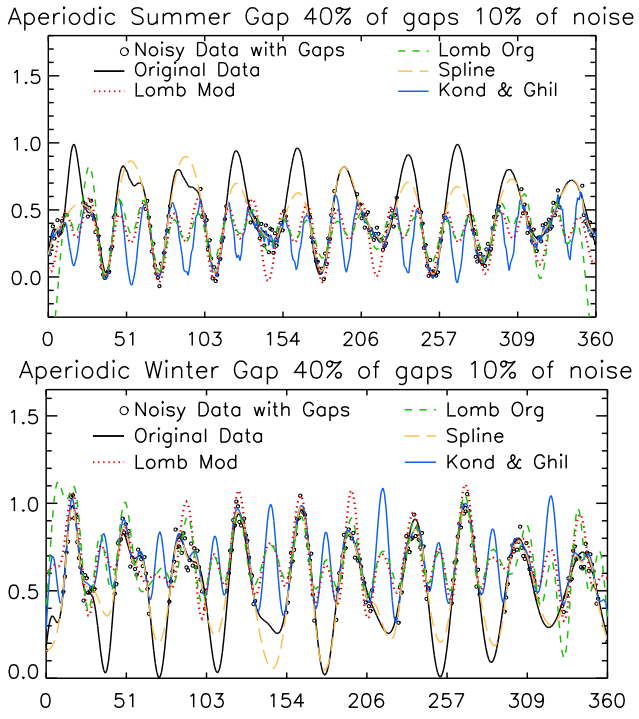

Fig. 19. Examples of reconstructions of an aperiodic time series with winter and summer gaps acquired by means of the indicated methods, to show the spurious peaks generated by the Kondrashov and Ghill as well as the Lomb-Scargle algorithms, but not the smoothing spline method.

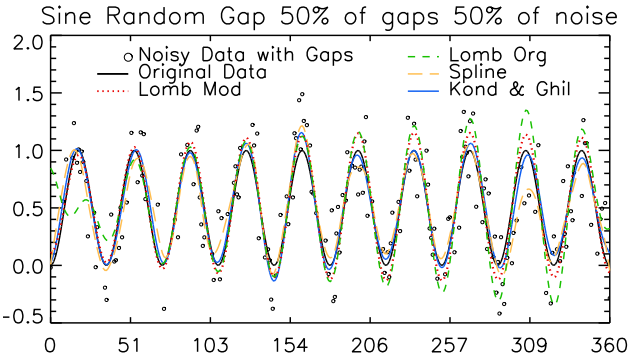

Fig. 20. Example of a reconstruction of a sine time series with random gaps acquired by means of the indicated methods. The amount of missing points is equal to 50 % and noise level is equal to 50 %. See text for details.

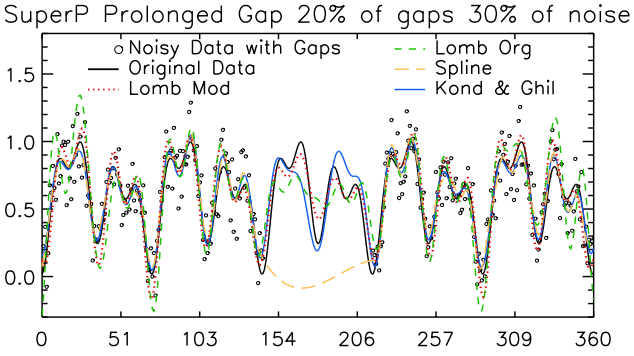

Fig. 21. Example of a reconstruction of a time series composed of superposition of trigonometric functions with a prolonged gap acquired by means of the indicated methods. The amount of missing points is equal to 20 % and noise level is equal to 30 %. See text for details.

artificial time series as well as measurement datasets. The strengths and limitations of these methods have been explored and are summarized below.

First and foremost, it must be realized that tests performed cannot fully represent the range and diversity of goals in any particular analysis. As mentioned earlier, a gap filling and smoothing algorithm that would exactly match each and every available data point would yield a null MAE but would also generate rather unrealistic and unusable values anywhere else. Thus, if one desires a smooth approximation to the original data, one must also accept larger values of the goodness of fit criteria.

All methods discussed above include a mechanism to adjust the degree of smoothness, be it the window length in the case of Kondrashov-Ghil, the threshold used to select the power spectrum peaks to be retained in the Lomb-Scargle approach, or the stiffness parameter of the smoothing spline. Objective algorithms have been proposed in each case to establish an "optimal" value of this parameter, though this step only formalizes a particular way of expressing the ultimate goal. The results presented in this paper thus highlight some of the strengths and weaknesses of these approaches in particular conditions, but cannot substitute a personal judgment that will also often involve other criteria, for instance the need to extrapolate the time series outside the range of available values or the need to achieve a minimum degree of smoothness in the reconstruction.

5.1 Lomb-Scargle

The Lomb-Scargle technique permits the estimation of the periodogram of a time series where the data points do not need to be equally spread in time. This is an extension of the classical Fourier approach, it leads to the natural decomposition of the original time series into a superposition of trigonometric functions, each associated with a certain "strength", quantified by the corresponding power in the periodogram or power spectrum. Assuming frequencies with limited power represent less relevant fluctuations, the bulk of the original signal can be reconstructed by summing only those components that have a sufficient power in the periodogram. Hocke and Kämpfer (2009) proposed the use of the Hamming window to enhance the retrieval of the spectral information in the time series, though the authors and our tests showed that the

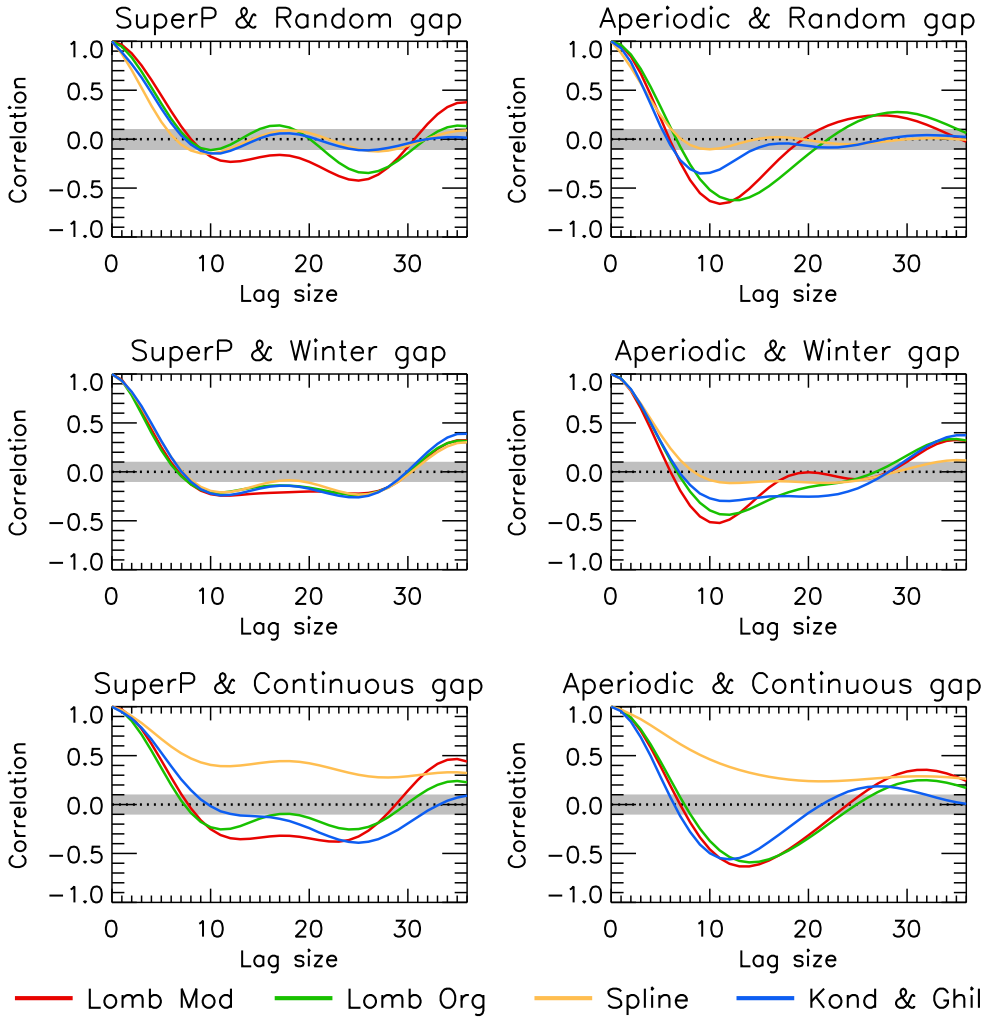

Fig. 22. Average autocorrelations of residuals between original and reconstructed time series derived from an extensive set of the artificial time series. Shaded areas indicate the 5 % confidence level that reported autocorrelations are not statistically significant. See text for details.

performance of the approach degrades near both ends of the record. The solution to this problem suggested by Hocke and Kämpfer involves modifying the spectral window and using only the middle part of the reconstructed data segment. This may not be satisfactory if most or all of the data record is required, but our investigation showed that this side effect can be largely controlled by using the Kaiser-Bessel window instead.

The strength of the Lomb-Scargle algorithm is also its weakness: it is particularly appropriate to process strongly periodic signals, but appears to be less apt than other methods to deal with aperiodic time series. In fact, the experiments with artificial time series described above showed that Lomb-Scargle algorithm, together with the Kaiser-Bessel window, is fully capable of regenerating credible values to fill gaps when the underlying function is periodic and the distribution of missing values is random. Aperiodic components in the signal or a distribution of gaps that interferes with the base frequencies of the signal are likely to cause less reliable results, up to the point of generating spurious, intermittent fluctuations in the reconstructed signal that were never present in the original data (e.g., in the "winter gap" and "summer gap" scenarios). This approach is also sensitive to the presence of a strong trend in the original data (as in the Mauna Loa CO_2 time series), which boosts the power spectrum at low frequencies and may therefore mask other frequencies that would be significant in the absence of that feature. This case is best handled by removing the trend from the signal first, processing the residuals, and then adding the trend back to the results.

On the positive side, it must be recalled that the capability of detecting significant frequencies in arbitrary time series is a very powerful tool, especially to project likely values of the record in the future.

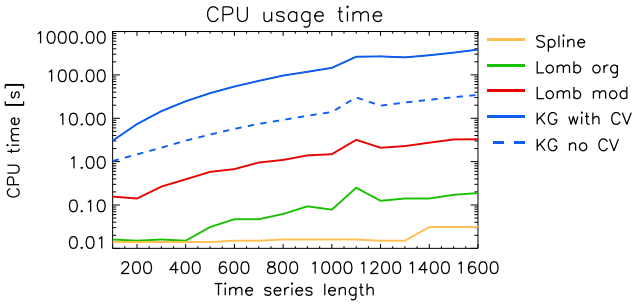

Fig. 23. The computational cost of selected methods expressed as CPU time in seconds as a function of time series length. Two implementations of the Kondrashov and Ghil method were tested: with and without the cross-validation procedure to optimize number of leading EOFs. It should be noted that for this method, values reported covary with different parameters such as: window length, number of leading EOF to be concerned, convergence test, number of runs required to optimize leading EOF (in the example presented this value was set to 10 runs), and a programming language. Nevertheless, overall relationship in computational cost should remain similar. The test was performed on a Compaq nx7400 with Intel Centrino Duo 2.00 GHz computer.

5.2 Kondrashov and Ghil

The Kondrashov and Ghil (2006) method, based on the Singular Spectrum Analysis (Golyandina et al., 2001), is a powerful and effective approach to fill gaps and smooth a univariate time series. Of all approaches tested here, this one generated the smallest MAE values in experiments with artificial time series involving either randomly distributed gaps or a long continuous period of missing data (Fig. 1). However, when the distribution of these missing data followed a seasonal patter (e.g., the "winter gap" scenario), it generated spurious, intermittent peaks in the reconstructed data record as did the Lomb-Scargle approach (see Fig. 19). As noted above, the simulated values remain reliable during those periods where a majority of the measurements are not missing, but neither the Kondrashov-Ghil nor the Lomb-Scargle approach can be recommended for approximating the reconstruction of the systematic seasonal gaps.

Experiments with real time series also demonstrated that the Kondrashov-Ghil approach is very flexible and effective in a wide range of applications. It handles datasets with strong linear trends as well as aperiodic components. This is not surprising, since decomposing a time series into a trend, a set of periodic components and other signals constitutes a primary objective of this approach. In principle, this technique should help suggest probable causes or explanatory factors for the observed variations and, as was the case for the Lomb-Scargle algorithm, the capacity to "learn" the primary modes of fluctuations from past records should provide some skill in predicting future values.

In general, the Kondrashov-Ghil method generated MAE values similar to slightly higher than those for the smoothing spline method in the various tests on actual time series. The statistics of the quality fit criteria from the artificial time series experiments revealed that this method is particularly suitable for reconstruction of long, continuous gaps. The main drawback of this method is its complexity and especially the large computational requirements (Fig. 23), which may quickly become prohibitive when processing very long or very many time series (Wang and Liang, 2008).

5.3 Smoothing spline

The smoothing spline algorithm delivers a piecewise cubic approximation of the noisy, unevenly sampled time series. The shape of the reconstructed spline depends on a smoothing parameter, optimized through the General Cross Validation (GCV) procedure.

Experiments with artificial datasets demonstrated that this method provides accurate reconstructions, comparable to the Kondrashov-Ghil algorithm, for all types of data in the random gap scenario. It performs much better than the latter or the Lomb-Scargle method in the winter gap scenario, precisely because it does not exploit any spectral information and thus does not introduce spurious peaks in the reconstructed time series. Moreover, for both gap scenarios mentioned the autocorrelation of residuals is weak for the fitted spline function. On the other hand, the smoothing spline algorithm is not able to accurately reconstruct reasonable missing values during long, continuous periods.

The smoothing spline method generated some of the smallest values of MAE in experiments involving real time series, even producing marginally better results than the Kondrashov and Ghil technique, and even for very small proportions of missing values (e.g., 1 %). As noted above, the main limitation of the smoothing spline gap filling algorithm is its inability to generate reasonable values when the missing values are clustered in one single long period. These results are quite understandable, since the smoothing spline method is essentially a "local" approximation, which takes advantage of neighboring observations to generate an estimate, but does not have any mechanism to "learn" the general properties of the whole time series and therefore guess adequate values in the absence of these neighbors. For the same reason, that method should have little or no predictive skill.

6 Conclusions

Four methods were evaluated in terms of their performance to fill gaps and filter noise in time series: two versions of the Lomb-Scargle algorithm, with different windowing schemes, the Kondrashov-Ghil approach and the smoothing spline method. The Mean Absolute Error (MAE), Mean Bias Error (MBE), chi-squared test and autocorrelation function were chosen as the goodness of fit criteria. The various tests

conducted showed that each method has its strengths and weaknesses.

The Lomb-Scargle approach, which is an extension of the classical Fourier analysis to the case where the observations or measurements in the original time series can occur at arbitrary times, works well as long as the underlying signal can be considered as a sum of periodic components. It therefore shares the same powerful theoretical basis as the classical spectral methods. However, when aperiodic fluctuations are introduced, additional frequencies should be called upon in the power spectrum, though the associated power may be too small to be statistically significant. Limiting the reconstruction to those frequencies associated with a statistically significant power then leads to a limited capacity to fit the original series. This method is of course particularly sensitive to the systematic removal of data points at specific frequencies (e.g., winter and summer gaps), in which case it can generate spurious values, though the values in other seasons remain credible. If a strong trend is present in the original data, it should be removed before starting the analysis proper. Lastly, it has been shown that the Kaiser-Bessel window appears to yield better results than the Hamming window, especially near either end of the reconstructed record.

The Kondrashov-Ghil method proved reliable and accurate in reconstructing a wide range of time series. As for the Lomb-Scargle method, it is not able to reconstruct periodically missing fragments of the dataset, producing false, intermittent peaks in these systematic gaps. In all other experiments, it performed well, especially in the case of long, continuous gap in the time series. From the analytical point of view SSA (Singular Spectrum Analysis), the core of the Kondrashov-Ghil algorithm, benefits from a strong theoretical foundation and provides a wide range of tools to process time series. This method is thus valuable to estimate missing values and smooth time series, but also to investigate the nature of underlying physical processes controlling this time series. The main drawback of this approach is its considerable computational cost, especially for very long or numerous time series, when it can quickly become prohibitive.

The smoothing spline gap filling method provided generally satisfying results, especially when the values of neighboring data points provide sufficient information for a local solution to guess the missing values (random or seasonal gaps). The dataset reconstruction results were accurate for most experiments with real and synthetic time series. The main limitation of this method is associated with continuous data gap scenario, where the algorithm is not able to utilize the spectral information retrieved from whole time series to fill a prolonged data gap. On the other hand, the computational cost of this approach is very limited, which is a definite advantage in operational environments.

The choice of a particular approach to estimate the values of missing observations in a time series thus depends very much on the underlying nature of the signal, on the type and distribution of the data gaps, and on the expectations of the investigator in terms of staying close to the existing data (low MAE) or requiring a smoother representation of the broad features of the time series.

Supplement related to this article is available online at: http://www.atmos-chem-phys.net/11/7905/2011/acp-11-7905-2011-supplement.pdf.

Acknowledgements. The authors thank K. Hocke and S. Wunderle, both at the University of Bern, Switzerland, for useful comments on an earlier draft of this manuscript, as well as the journal Editor and the Reviewers who provided very constructive remarks. The first author gratefully acknowledges the Joint Research Centre of the European Commission for financial support during an extended visit (October 2009 to September 2010). Some of the figures in an enhanced resolution are provided in the supplement.

Edited by: A. J. G. Baumgaertner

References

Box, G. E. P. and Jenkins, G. M.: Time Series Analysis Forecasting and Control, Holden-Day, Oakland California, 1976.

Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, Phys. D, 20, 217–236, 1986.

Broomhead, D. S., Jones, R., and King, G. P.: Topological dimension and local coordinates from time series data, J. Phys. A, 20, 563–569, 1987.

Craven, P. and Wahba, G.: Smoothing noisy data with spline functions estimating the correct degree of smoothing by the method of generalized cross validation, Numerische Mathematik, 31, 377–403, 1979.

Fourier, J.: Theorie Analytique de la Chaleur, Firmin Didot (reissued by Cambridge University Press, 2009; ISBN 978-1-108-00180-9), 1822.

Ghil, M. Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced Spectral Methods for Climatic Time Series, Rev. Geophys., 40, 1–41, 2002.

Gobron, N., Pinty, B., Aussedat, O., Chen, J. M., Cohen, W. B., Fensholt, R., Gond, V., Lavergne, T., Mélin, F., Privette, J. L., Sandholt, I., Taberner, M., Turner, D.P., Verstraete, M. M., and Widlowski, J.-L.: Evaluation of Fraction of Absorbed Photosynthetically Active Radiation Products for Different Canopy Radiation Transfer Regimes: Methodology and Results Using Joint Research Center Products Derived from SeaWiFS Against Ground-Based Estimations, J. Geophys. Res.-Atmos., 111, D13110, doi:10.1029/2005JD006511, 2006.

Golyandina, N., Nekrutkin V., and Zhigljavsky A.: Analysis of Time Series Structure: SSA and Related Techniques, Chapman and Hall, Washington DC, USA, 2001.

Harris, F. J.: On the use of windows for Harmonic Analysis with the Discrete Fourier Transform, Proceedings of the IEEE, 66, 51–83, 1978.

Hocke, K. and Kämpfer, N.: Gap filling and noise reduction of unevenly sampled data by means of the Lomb-Scargle periodogram, Atmos. Chem. Phys., 9, 4197–4206, doi:10.5194/acp-9-4197-2009, 2009.

Hutchinson, M. F. and de Hoog, F. R.: Smoothing Noisy Data with Spline Functions, Numerische Mathematik, 47, 99–106, 1985.

Hutchinson, M. F.: Interpolation of Rainfall Data with Thin Plate Smoothing Splines – Part I: Two Dimensional Smoothing of Data with Short Range Correlation, J. Geogr. Info. Dec. Anal., 2, 139–151, 1998.

Keeling, C. D., Chin, J. F. S., and Whorf, T. P.: Increased activity of northern vegetation inferred from atmospheric CO_2 measurements, Nature, 382(6587), 146–49, 1996.

Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, Nonlinear Proc. Geophys., 13, 151–159, 2006.

Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, Astrophys. Space Sci., 39, 447–462, 1976.

Mahecha, M. D., Fürst, L. M., Gobron, N. and Lange, H.: Identifying Multiple Spatiotemporal Patterns: A Refined View on Terrestrial Photosynthetic Activity, Patt. Recog. Lett., 31, 2309–2317, 2010.

Moffat, A., Papale, D., Reichstein, M., Hollinger, D. Y., Richardson, A. D., Barr, A. G., Beckstein, C., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., and Stauch, V. J.: Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, Agr. For. Meteorol., 147, 209–232, 2007.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipes in Fortran, Cambridge University Press, Cambridge, USA, 2 edn., 1992.

Reinsch, C. M.: Smoothing by spline functions, Numerische Mathematik, 10, 177–183, 1967.

Vautard, R., Yiou, P., and Ghil, M.: Singular-spectrum analysis: A toolkit for short, noisy chaotic signals, Physica D, 58, 95–126, 1992.

Verstraete, M. M., Gobron, N., Aussedat, O., Robustelli, M., Pinty, B., Widlowski, J.-L., and Taberner, M.: An automatic procedure to identify key vegetation phenology events using the JRC-FAPAR products, Adv. Space Res., 41, 1773–1783, 2008.

Wahba, G.: Spline Models for Observational Data, SIAM, Philadelphia, 1990.

Wang, D. and Liang, S.: Singular Spectrum Analysis for Filling Gaps and Reducing Uncertainties of MODIS Land Products, International Geoscience and Remote Sensing Symposium (IGARSS), 5, 558–561, 2008.

Whittaker, E. T.: On a new method of graduation, Proc. Edinburgh Math. Soc., 41, 63–75, 1923.

Willmott, C. J. and Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res., 30, 79–82, 2005.