Hydrol. Earth Syst. Sci. Discuss., 10, 12793–12827, 2013 www.hydrol-earth-syst-sci-discuss.net/10/12793/2013/ doi:10.5194/hessd-10-12793-2013 © Author(s) 2013. CC Attribution 3.0 License.



This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Applications of time series analysis in geosciences: an overview of methods and sample applications

W. Gossel and R. Laehne

Martin Luther University Halle, Institute for Geosciences and Geography, Dept. Hydro- and Environmental Geology, Halle, Germany

Received: 23 August 2013 - Accepted: 20 September 2013 - Published: 28 October 2013

Correspondence to: W. Gossel (wolfgang.gossel@geo.uni-halle.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.





Abstract

5

Time series analysis methods are compared based on four geoscientific datasets. New methods such as wavelet analysis, STFT and period scanning bridge the gap between high resolution analysis of periodicities and non-equidistant data sets. The sample studies include not only time series but also spatial data. The application of variograms as an addition to or instead of autocorrelation opens new research possibilities for

as an addition to or instead of autocorrelation opens new research possibilities for storage parameters.

1 Introduction

Time series analysis is an agglomeration of methods used for the a posteriori analysis of data recorded in dependency of each other. The methods identify trends, periodicities, and autocorrelative relations. The questions focus not only on time domain data but can also be used for spatial datasets. The variety of applications is therefore very high and reaches from astronomy, meteorology, geography and biology to geology, and in the social sciences and economics, from elections to stock values to name only some applications.

There are two main reasons for the application of time series analysis:

- 1. Finding out a regular or steady behaviour for the identification and quantification of processes.
- 2. The results of the time series analysis are used for the prediction of values.
- ²⁰ This means that the position of this method has a more general character in the applied geosciences than other analytical methods, for e.g. parameter identification, and has less importance for prediction than numerical modelling. It is used for both aspects and additionally can be used for the identification of processes.

The investigation of trends has become extremely important in the frame of anthropogeneous impacts on the environment: trends of atmospheric or meteorological data





influence the daily weather forecast (SEOS, 2013; DWD, 2013), prognoses for climate change (Schotterer et al., 2007; Ghil et al., 2007), and trend calculations are an essential part of the EU water framework directive (BMU, 2011; Aguilar et al., 2007). As a statistical method, trend analysis will not explain the reasons for trends but identification is supported. Trend analysis as well as correlations may mislead to dependencies

⁵ tion is supported. Irend analysis as well as correlations may mislead to dependencies that are not explainable in process identification and modelling. This has to be tested carefully.

Cycles of events or developments have been regarded in geology for a long time: they include developments of orogenesis as well as cycles of ice ages, sedimentary sequences and their results in lithology and biological facies development. Also, growth

- cycles of reefs and corals, the warves, eruptions of volcanoes and earthquakes have been investigated for cyclic behaviour (Denlinger and Hoblitt, 1999; Bufe and Varnes, 1993). Period analysis, as it is carried out in most applications, is a harmonic analysis as proposed by Fourier (1822). In general this means a reduction to the most important
- frequencies. The general aspect of this theory as shown by Dirichlet (1829), is not applicable in the analysis of measurement datasets because the time series data from observations are not necessarily continuous functions. The difficulties aforementioned for trend analysis in a process oriented interpretation are also given for period analysis, but due to the higher level of the analysis method the chances for misinterpretation are not applicable and period analysis.

²⁰ not as high as in trend analysis and correlation analysis.

10

Autocorrelation describes a dependency on former events. The connection to the other two methods is obvious: a trend implies a strong relation to the predecessor measurement and a periodicity should also be detected by an autocorrelation after one period. Autocorrelation is also investigated indirectly by geostatistical methods:

²⁵ a variogram shows autocorrelation in detail and can therefore also be tested for cyclic events. Autocorrelation analysis leads in general to more spatial aspects of the proposed methods. Trends, periods and autocorrelation are generally applicable to spatial data as well as to time series data but only autocorrelation analysis has been further developed for spatial data.





Time series analysis is commonly applied to regularly measured data sets (trend analysis, Fast Fourier Transformation and autocorrelation). The spatial application in geostatistical methods shows the necessity for an extension to irregularly spaced data sets. Not only autocorrelation but also trend analysis and period analysis can be methodically substituted for irregular spaced data sets:

- 1. Time dependent correlation instead of trend analysis,
- 2. Lomb's method, the STFT or the method of period scanning explained in this paper in detail, instead of Fast Fourier Transformation, and
- 3. Variogram analysis instead of autocorrelation.

5

The reasons for this methodical substitution are manifold: sometimes the data cannot be sampled in equidistant time steps, sometimes accidents, malfunctions or wrong adjustments of instruments lead to non-equidistant recordings. The completion of the recordings by interpolation methods is misleading because the original data set is changed. Numerous applications of the time dependent correlation and the period scanning for hydrograph analysis have been presented by Gossel (1999).

Additionally, sometimes an index system is applied to work with nominal or ordinal data. This can be explained on a geological data set: In the field a profile (in a quarry or a borehole) gives lithological information or some geophysical information. Geologists derive via expert knowledge the stratigraphy and therefore the time dimension. In this

- ²⁰ way a time series analysis can be carried out and used for parameter estimation especially if the lithological data are converted to hydrogeological or engineering geological data, or reservoir data. An application of parameter identification and input for a numerical groundwater flow model is presented in Gossel (1999). In this way a double conversion leads to parameters of models. For the analysis of hydrographs the advan-
- tage can be that parameters can be estimated based on long term observation at only one point instead of different observation points at one time. An example for the field of groundwater observation is the calculation of hydraulic conductivities based on the





propagation of tidal water level fluctuations at coastlines. The calculation is based on the heat propagation described by Fourier (1822).

Another question that is directly connected to time series analysis, especially for dynamic models, is the input of boundary conditions. An application in a numerical groundwater model is given in Gossel et al. (2009) and Gossel (2011) where a ten year hydrograph is elongated into the past. The disadvantage is also clear: process oriented corrections of the results of the time series analysis cannot be applied.

Time series analysis is presented in this paper with some applications from the geosciences: time series of meteorological data from Berlin (Chowanietz and Gossel,

- 10 1997) and the hydrograph of a groundwater observation well in Halle (Saale) (Germany) are focused upon in detail. Their spatial applicability is tested at an example of glacial channels in Lower Saxony (Germany). An overview of the analyzed data sets is given in Table 1. The data sets are evenly sampled but to investigate the capabilities of the methods some of them are resampled irregularly. Additionally, the data set of the Berlin climatic water balance is cut at half of the total length to test the prediction
- possibilities.

This procedure shows that the applications do not play a central role in this investigation: some of the results have already been published, others are easily predictable. Here, the focus will be set on the applicability, the comparability and the resolution of the methods on real data sets and not on predefined functions.

2 Methods

25

The methods used for the time series analysis of the proposed data sets are very diverse as there has been some interest in this topic for the last 20 yr and computer evolution has made routine calculations quick and easy to complete. Therefore the combination of trend analysis, FFT and autocorrelation is compared to the combination of time dependent trend correlation, Lombs method, STFT, wavelets analysis, period scanning and variogram analysis as shown for the applications in Table 1.





The methods used here have to be formulated in mathematical equations to have an exact procedure for the comparison. The software tools for the methods are freely available and they are mentioned according to the methods.

2.1 Trend analysis

5 The basic equation for the trend analysis is given by

$$X_{\rm Trend}(t) = \sum_{n=0}^{m} k_n t^n$$

with:

 $X_{\text{Trend}}(t)$ = estimated value at time t; t = time;

10 k = parameter;

 $n = 0, 1, 2, 3, \ldots, m.$

Commonly only the first member of the sum is used to calculate a linear trend which results in

 $X_{\rm Trend}(t) = k_0 + k_1 t$

¹⁵ In general this is the same as for a regression with *t* as the independent variable and *X* as the dependent variable. If the time series is not equidistant, the trend can therefore also be calculated as a time dependent correlation and regression. The correlation coefficient is in this case a measure for the time dependency of the data set. It is calculated via Eq. (3)

$$r_{XY} = \frac{\operatorname{Cov}(X, Y)}{\sigma(X)\sigma(Y)}$$

with:

X = independent data set (x axis), in time series analysis: time t;



(1)

(2)

(3)

Y = dependent data set (*y* axis), in time series analysis: sample values; r_{XY} = correlation coefficient; Cov(X, Y) = Covariance $(X, Y) = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$;

$$\sigma(X) = \text{Standard deviation } (X) = \sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2};$$

5 $\sigma(Y)$ = Standard deviation $(Y) = \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}$; *n* = number of samples.

The parameters of the regression equation can be calculated according to Eqs. (4) and (5)

$$k_1 = r_{XY} \frac{\sigma(X)}{\sigma(Y)}$$

10 and

$$k_0 = \overline{x} - k_1 \overline{y}$$

The correlation coefficient is also used for period scanning but without testing the time dependency of the data set. The significance of the regression is derived from the correlation coefficient and the number of samples. The correlation coefficient is between -1 and 1 with both values indicating the highest significance and a value of 0 has the lowest significance. The significance level in hydrology is normally set to the 95 % confidence interval. The link between correlation coefficient and significance is either derived from the normal distribution or can be found in tables as in Siegle (2009).

20 2.2 Period analysis

The second step in time series analysis is the identification of periods. The development here is of high importance because it is – similar to trend analysis – used in a wide range of scientific disciplines.



(4)

(5)



2.2.1 Analysis of equidistant time series

Fast Fourier Transformation

For the identification of periods in most cases the FFT is used (Fourier 1822); this is based on a harmonic analysis and therefore an equidistant sampling is necessary. To

- ⁵ find the most important (and significant) period(s) in the data set, the amplitudes of the periods are visualized in a periodogram. On the *x* axis of a periodogram the harmonic number, period or frequency is given, the *y* axis represents the intensity (power) or amplitude of a wave. The phase shift of an optimal wave can also be calculated with this method.
- ¹⁰ The periodical part of a time series can be represented by waves with Eq. (6).

$$X_{\text{Period}}(t) = \overline{x} + \sum_{m=1}^{n/2} \left(\alpha_m \cos \frac{2\pi m}{n} t + \beta_m \sin \frac{2\pi m}{n} t \right)$$
(6)

The parameters α and β in Eq. (6) can be calculated from the time series as follows (Eqs. 7 and 8).

$$\alpha_m = \frac{2}{n} \sum_{t=1}^n X(t) \cos \frac{2\pi m t}{n}$$

15 and

$$\beta_m = \frac{2}{n} \sum_{t=1}^n X(t) \sin \frac{2\pi m t}{n}$$

The intensity or power is calculated based on these parameters by Eq. (9).

 $I_m = \alpha_m^2 + \beta_m^2$



(7)

(8)

(9)

The amplitude is just the square root of the intensity as shown in Eq. (10).

$$A_m = \sqrt{\alpha_m^2 + \beta_m^2}$$

The optimal phase shift for the best correlation can be calculated by Eq. (11).

$$\varphi_m = \tan^{-1} \frac{\alpha_m}{\beta_m} \tag{1}$$

5 Wavelets

The wavelet analysis is similar to the FFT as it identifies relevant frequencies by their power. It also requires evenly spaced data sets but it works with windows within the data set. These windows are shifted over the full time range (continuous wavelet transformation). With the wavelet technique the windows (of predefined time length) are tested for a wide range of frequencies simultaneously so that a periodogram is not any

- more suitable for visualization: on the *x* axis of a continuous wavelet transformation diagram the time is given as usual, but on the *y* axis the frequency is shown so that the power has to be shown in grey scale or colours. Wavelets do not have the assumption of stationary frequencies as they test for every time window of predefined length the frequencies that are possible for their power. An application of wavelet analysis is given
- by Nakken (1999) for rainfall-runoff patterns.

2.2.2 Analysis of non-equidistant time series

The time series may be sampled non equidistant. In this case three methods can be applied:

- 1. Lomb's method
 - 2. Short Time Fourier Transform (STFT)
 - 3. Period scanning

(10)

1)



Lomb's method

The Least-Squares Frequency Analysis (Lomb, 1976) uses the method of least squares fitting to find periodicities in unequally spaced observations. By applying different frequencies, a normalized periodogram with the power as a frequency dependent function

5 is calculated. Afterwards, the calculated power can be tested on different significance levels.

STFT

The Short-Time Fourier Transform works (as the wavelet technique) with windows of the total time series. Within these windows the FFT as described above is used for the analysis. This allows the investigation of unevenly sampled data, which are evenly sampled in discrete windows, but it has the disadvantage that the whole data set is not analyzed in one step. The windowing technique runs into the problem of decision for long regarded time window or high frequency because within such a time window stationarity of the waves is assumed.

15 Period scanning

Period scanning does not work like the FFT, because it calculates just a simple correlation between the data set and a synthetic cosine signal. This method was introduced by Gossel (1999), and focuses on computational power to optimize the amplitudes, phase shifts and to find the highest value for the correlation coefficient as given in Eq. (3). The method has the advantage of being completely independent of equidistant compliant

method has the advantage of being completely independent of equidistant sampling. It has the disadvantage of testing a lot of frequencies for the whole data set – which makes it a stationary method – and the frequencies are definitely not independent from each other. The correlation coefficients identify the most relevant periods accurately, which may be not possible with the FFT or the STFT as will be discussed later.





2.3 Autocorrelation analysis

The dependency of a sample on its predecessor(s) is tested by autocorrelation methods. The importance of this analysis is a bit neglected in the sciences compared to trend analysis and period analysis. In the earth sciences it is useful for the identification and guantification of system storages.

2.3.1 Autocorrelation method

5

10

Autocorrelation works with the correlation coefficient as described in Eq. (3). The data set is compared with itself with a time shift of 1, 2, 3, ..., n so that the most important condition for this procedure is again (as in the FFT) the constant time lag between the measurements.

2.3.2 Autocorrelative variogram analysis

For non-equidistant time series the autocorrelation analysis can be solved via an analogy to spatial data analysis. Spatial data are normally not equidistant and therefore another kind of diagram is used to find out the spatial correlation of data; the variogram. On the *x* axis the distance and on the *y* axis the semivariance of data is shown (for details see Deutsch and Journel, 1998). The visually derived continuous function is used in geostatistics for spatial interpolation. For time series analysis a similar diagram to this variogram can be used as a substitute for the autocorrelogram in case of non-equidistant data. Two approaches are possible: If the gaps or the time shifts in the time series are not so big, an adjusted correlation may be calculated that defines a small time difference (less than the average time lag) to correlate the time series with itself – this is not easy to implement. The alternate approach would be to calculate a variogram – with lag distance of the average lag distance – and to convert semivariance





values to correlation coefficients. This is quite easy, as shown by Eqs. (12)-(14).

$$r(h) = \frac{\operatorname{Cov}(h)}{\operatorname{Var}}$$

and

 $\gamma(h) = \text{Var} - \text{Cov}(h)$

5 gives

$$r(h) = \frac{\operatorname{Var} - \gamma(h)}{\operatorname{Var}}$$
(14)

with:

r = correlation coefficient;

Cov = covariance;

• Var = total variance of the whole data set;

 γ = semivariance;

 $h = \log$.

These equations allow the conversion of a variogram to an (auto)correlogram under the conditions of second-order stationarity.

15 2.4 Tools

The big data sets require computer methods for their analysis. Of course, standard spreadsheet calculations are used for basic statistics but for the FFT they were improved by self-developed scripts. The period scanning was carried out with the Open-Source tool tsa (Gossel, 2013) which is based on Gossel (1999). The tool "PAST",

²⁰ which is developed and maintained by Hammer and Harper (2005), was applied for the remaining analyses.



(12)

(13)



3 Applications

The range of applications of time series analysis in the sciences is as wide as the observations of time series. From this big pool only a few applications from the earth sciences – especially hydrology and hydrogeology – are picked out:

- 5 1. Climatic data from Berlin (Germany)
 - 2. A groundwater hydrograph of the University Campus in Halle (Germany)

Additionally the applicability to spatial data is linked to time series and tested via profile data and horizontally oriented data:

- 3. A sedimentary sequence of the lower Triassic
- 4. The base of the Quaternary in a cross section in Lower Saxony

All applications are used to test the capabilities for the analysis of non-equidistant time series. The prognostic capabilities are tested by cutting half of the time/spatial series and trying to predict the other half. Additionally the time series of the climatic data of Berlin is made non-equidistant by taking out one value every year, and from the groundwater hydrograph irregular gaps of one week are inserted as shown in Fig. 2.

A very short introduction outlines the questions connected to the analysis of the data sets.

3.1 Climatic data from Berlin, Germany

The data set from Berlin comprises monthly evapotranspiration, precipitation, and cli-²⁰ matic water balance data over 160 yr (1851–2010). The data set of Chowanietz and Gossel (1997) is extended by the last decade and the investigated time series is that of the climatic water balance (see Fig. 1). The objective here is to find out the trend in the whole data set (and its significance), and to find periods other than the expected yearly one. From this big data set the end of the hydrological year (October) is picked ²⁵ out in a second step to test the methods for this kind of gap.





3.2 Groundwater hydrograph of the University Campus in Halle (Saale), Germany

The groundwater levels at an observation well on the University Campus in Halle (Saale) are recorded automatically every hour. The averages for every day have been taken for the method testing. The recording started in May 2004 and was updated until 2009. The data were already published in Gossel et al. (2011). The complete hydrograph is shown in Fig. 2. The one week gaps to test the results of the period scanning are distributed irregularly over the whole time period of measurements, so that in total 10% of the values are taken out.

3.3 Sedimentary sequence of the Lower Triassic

In the Lower Buntsandstein (Bernburg Formation) a profile about 20 km northwest of Halle (Germany) was investigated very intensely. It is described lithologically and with a gamma-ray log in Hauschke and Szurlies (1998). The gamma-ray log was digitized and the "time series analysis" was carried out with the height instead of time as the x axis. The digitized log is shown in Fig. 3.

3.4 Quaternary base in Lower Saxony

15

The Quaternary base in Lower Saxony was cut into the Tertiary by channels of the first glacial. A cross section was taken from the interpolated base as visualized by Gossel et al. (2012). The cross section and an overview map are shown in Fig. 4. The periodicity can already be estimated visually. In this case the distance from the line start was taken as *x* axis.



4 Results

The results are presented according to the applications. The "prognostic" calculation for the second half of the time series is also presented for all applications.

4.1 Climatic data Berlin

- The trend in the precipitation data is not significant and would have been without relevance with a slope of only 0.0003, but some periods are significant. The periodograms in Fig. 5 outline the periods of 4, 6, 12, 36, and 48 months with the highest correlation coefficients, and most of the periods are also found by the FFT and the Lomb method. The results for the STFT and wavelet analysis are visualized in Fig. 6. The latter methods show the yearly and the six months frequencies, too. The periodogram of
- the reduced time series shows very similar patterns, so that a stationary behaviour can be assumed, see Fig. 7. The predictability, as via trend and period analysis, is nevertheless restricted as can be assumed from the comparison of measured and simulated data.

15 4.2 Groundwater hydrograph of the University Campus

The trend in the water levels is significant but has no relevance due to a very small slope value. Several periods are highly significant. The overlay of measured and simulated data show quite good approximation based on the ten most important periods (see Fig. 2). The difference between the results of period scanning and FFT, and the Lomb
²⁰ method are negligible. The gaps in the hydrograph have minor influence on the results. For this case study, autocorrelation and variogram were also compared and Fig. 8 shows the good link via Eq. (14).





4.3 Sedimentary sequence

The gamma log of a sedimentary sequence of the Lower Buntsandstein has a significant trend but this trend has a slope of 0.0001 so it is no more than an average. The period analysis is very interesting because it reveals periods of 0.2 m, 3.5 m, 9.8 m and

more, see Figs. 9 and 10. The 9.8 m period can be referred to as the Milancovich cycles due to the sedimentary conditions. The results of the STFT analysis are quite difficult to interpret in this series. The long periods are dominating but frequencies higher than 0.1 have a tendency to higher powers. In certain parts of the profile distinct frequencies have higher powers than in other parts. The stationarity of this series has to be
 evaluated critically.

4.4 Base of the Quaternary

The Quaternary base as a spatial distribution needs reduction to a cross section to treat it like a time series. The trend is significant and rises from west to east with a gradient of nearly 0.02. The period analysis via period scanning has a peak at 8.5 km, which is also identified by the FFT and Lomb's method (see Fig. 11). STFT and wavelet analysis have to be regarded with high differentiation (see Fig. 12). The wavelet analysis shows the band quite impressively but with the STFT it is clear that this band is only that obvious in certain parts of the "time" series. The stationarity of this series is quite doubtful.

20 5 Discussion

15

The discussion follows the methods and compares the advantages and disadvantages of the methods used here.





5.1 Trend analysis

Trend analysis is fast, reliable and easy to carry out. The disadvantage is the missing applicability to non-equidistant data sets. The substitution by time dependent correlation and regression gives more information and is as quick and easy as the trend analysis. Although the trend should be subtracted from the time series before analyzing the periods it should be regarded critically for its correlation coefficient and its significance.

5.2 Time dependent correlation

The advantages of time dependent correlation and regression are obvious: the correlation coefficient expresses the significance of the time dependency, and the regression conveys the equation of the trend. It is as fast as trend analysis and can be carried out with every spreadsheet tool. There is no need for trend analysis and it can be substituted in general by the time dependent correlation.

5.3 FFT

5

The FFT shows a high dependency in the applications on the investigated data set. The transparency and elegance suffers at this point and additionally the investigated data set has to be equidistant. The results also show that the method is not able to identify long periods (low frequencies) adequately because of the focus on harmonic analyses. On the other hand this method is fast and is also applicable to calculation by hand.

By using the algorithms of Fisher (1929) or Nowroozi (1967) it is possible to investigate the reliability of the calculated magnitudes of the amplitudes, which are frequency dependent. Random fluctuations in the time series can be separated from reliable frequencies with a physical meaning.





5.4 Wavelets

The wavelet analysis is a very powerful tool although it depends on equidistant data. The windowing technique shows not only the "intensity" for the frequencies but also the time frame for which they have been detected. In the applications investigated here, it resolved the periods well and very fast. The identification of the periods is quite

⁵ here, it resolved the periods well and very fast. The identification of the periods is quite intuitive although the background theory is complicated. The precision and resolution of the results depends highly on the data set because some periods are not detected adequately due to the restrictions of the FFT to harmonic values.

5.5 Lomb's method

¹⁰ This method has the advantage of non-equidistant sampling analysis. Another advantage is the calculation of significance levels, to decide which frequency is reliable. But the calculated results of this method depend on the chosen dataset. So the choice of which part of the time series should be investigated has a big influence on the calculated result.

15 5.6 STFT

The STFT combines the advantages of the wavelet analysis and the non-equidistant sampling. In the examples above the method evolved to be robust against outliers and especially gaps in the series. The interpretation is similar to wavelet analysis (which is a bit easier as it directly shows the period instead of the frequency in the STFT) but is differentiated higher and is more sensitive to short periods. This sounds good but the method has one small disadvantage: It is impossible to identify high frequencies (short periods) in time frames of low sampling rates. But this is a minor disadvantage compared to the information density. The non-stationary analysis is also a big advantage.





5.7 Period scanning

This method has several advantages, but it is quite slow and it is based on the assumption of stationarity in the dataset. With high end computers a multithreading approach will result in fast analyses. The biggest advantage is the calculation of correlation co-

⁵ efficients that can be tested for significance instead of intensities or amplitudes. The identification of frequencies with a very high resolution of the user's choice allows for a detailed analysis. Even the measurements in periods with big gaps are used for the analysis results so that the data are used on the highest available level.

5.8 Autocorrelation

Autocorrelation reveals good results for equidistant data. It is absolutely necessary to subtract the trend and the most important periods from the dataset before analyzing for autocorrelation. Some datasets can nevertheless be interpreted adequately due to dominating periodicities.

5.9 Autocorrelative variogram analysis

¹⁵ Autocorrelation with a variogram shows very interesting results: the method is generally feasible for this purpose and it can be interpreted easily in terms of the background mathematics. The problem arises in the interpretation topic: in hydrogeology a non-quantitative relation to storages can be assumed which has to be identified by the slope of the graph in the first lags. The qualitative result is only meaningful in comparison with other autocorrelative variograms.

6 Conclusions

The methods used and discussed here all have their own advantages and disadvantages. Only a few methods are no longer helpful and can/should be substituted by





other – mostly newer – methods. A comparative application of the methods within the datasets of a project should be most advantageous and is not as complicated as it sounds for a new user. The methods are well implemented, documented and in general quite easy to handle.

⁵ In general the methods of time series analysis are used for the identification of processes. This was always obvious and it will be the most important factor for the application of the methods.

The application of the methods does not normally lead to direct parameter identification for modelling purposes and can by now only support the calibration process. The perspectives are good (by an increasing number of applications) to get a step forward with this approach.

An interesting aspect is the comparison of time series data and spatial data. The examples show the general applicability of methods for time series analysis to spatial data. The results are convincing but this may also be due to the selection of datasets. In general the link between spatial and time dependent data is underesti-

- ¹⁵ datasets. In general the link between spatial and time dependent data is underestimated: for groundwater recharge the spatial variability is in most regional sized groundwater catchments as high in time as it is in spatial dimensions. This aspect should and will be investigated in the future, e.g. for runoff in surface water catchments and depth dependent datasets in hydrogeology.
- 20 Another question indirectly connected to time series analysis is the relation to fractals. This topic was not discussed here but the comparison of spatial and time dependent data may be linked – especially for modelling aspects – in the fractal dimensions.

References

10

25

Aguilar, J. B., Orban, P., Dassargues, A., and Brouyere, S.: Identification of groundwater quality

trends in a chalk aquifer threatened by intensive agriculture in Belgium, Hydrogeol. J., 15, 1615–1627, 2007.





- Bufe, C. G. and Varnes, D. J.: Predictive modeling of the seismic cycle of the Greater San Francisco Bay Region, J. Geophys. Res., 98, 9871–9883, 1993.
- Chowanietz, U. and Gossel, W.: Konzeption einer ressourcenschonenden Wasserbewirtschaftung für die Region Berlin, Teil 2: Klima, Boden und Grundwasser, BUND Berlin, available at:
- http://www.wolfgang-gossel.de/htm/download.htm (last access: 23 August 2013), 1997.
 Denlinger, R. P. and Hoblitt R. P.: Cyclic eruptive behavior of silicic volcanoes, Geology, 27, 459–462, 1999
 - Deutsch, C. V. and Journel, A. G.: GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press, Oxford, 1998
- ¹⁰ Deutscher Wetterdienst (DWD): Lange Reihen und Trends, available at: http://www.dwd.de/bvbw/appmanager/bvbw/ (last access: 23 August 2013), 2013.
- Dirichlet, P. G. L.: Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, J. Reine Angew. Math., 4, 157–169, 1829
 Federal Ministry for the Environment, Nature Conservation and Nuclear
 15 Safety (BMU): The European Water Framework Directive and its Imple-
- mentation in Germany, available at: http://www.bmu.de/en/detailview/artikel/ the-european-water-framework-directive-and-its-implementation-in-germany-1/ (last access: 23 August 2013), 2011.

Fisher, R. A.: Test of significance in harmonic analysis, Proc. R. Soc. A, 125, 54–59, 1929.

²⁰ Fourier, J. B. J.: Theorie analytique de la chaleur, Didot, Paris, 1822.

30

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, 2007.

Gossel, W.: Hydrogeologie und Grundwasserhaushalt ausgewählter anthropogen wenig bee-

- ²⁵ influßter Grundwassereinzugsgebiete in Berlin, Berliner geowiss. Abh. A, 199, Selbstverlag der FU Berlin, Berlin, 1999.
 - Gossel, W.: Interfaces in Coupling of Hydrogeological Modelling Systems, Shaker-Verlag, Aachen, 2011.

Gossel, W.: Description TSA, available at: http://www.mo2geo.org/download/description_tsa. pdf (last access: 23 August 2013), 2013.

Gossel, W., Stollberg, R., and Wycisk, P.: Regionales Langzeitmodell zur Simulation von Grundwasserströmung und Stofftransport im Gebiet der Unteren Mulde/Fuhne, Grundwasser, 14, 47–60, 2009.





Gossel, W., Lähne, R., Rienäcker, J., Krauss, G., and Wycisk, P.: Auswertung von Grundwasserstandsdaten auf dem Weinberg-Campus der Martin-Luther-Universität Halle-Wittenberg, Hallesches Jahrb. Geowiss., 32/33, 129–141, 2011.

Gossel, W., Chudy, T., and Falkenhagen, M.: Interpolation based on isolines: line-geometrybased inverse distance weighted interpolation (L-IDW) with sample applications from the geosciences, Z. dt. Ges. Geowiss., 163, 493–505, 2012.

Hammer, O. and Harper, D. A. T.: Palaeontological Data Analysis, Blackwell, Oxford, 2005.

10

- Hauschke, N. and Szurlies, M.: Excursion B, stratigraphy, sedimentary environment and cyclicity of Buntsandstein and Lower Muschelkalk in the Central Germanic Basin (North of the Harz Mountains), Hallesches Jahrb. Geowiss., 6, 86–98, 1998.
- Lomb, N. R.: Least-squares frequency analysis of unequally spaced data, Astroph. Space Sci., 39, 447–462, 1976.

Nakken, M.: Wavelet analysis of rainfall-runoff variability isolating climatic from anthropogenic patterns, Environ. Model. Softw., 14, 283–295, 1999.

- Nowroozi, A. A.: Table for Fisher's test of significance in harmonic analysis, Geophys. J. R. Astr. Soc., 12, 517–520, 1967.
 - Schotterer, U., Stichler, W., Leuenberger, M., Nyfeler, P., Bürki, H., Kozel, R., and Schürich, M.: Zeitreihenanalyse in der Klimabeobachtung am Beispiel von Deuterium und Sauerstoff-18 im Niederschlag und in alpinen Eisbohrkernen, NABO-Fachtagung Umweltbeobachtung – ein
- Zeitreihenproblem, Bern, available at: http://www.bafu.admin.ch/bodenschutz/10161/10188/ index.html?lang=de (last access: 23 August 2013), 2007.
 - Science Education through Earth Observation for High Schools (SEOS): What is a time series and why are they important?, available at: http://lms.seos-project.eu/learning_modules/timeseries/timeseries-c01-p03.html (last access: 23 August 2013), 2013.
- Siegle, D.: Critical values of the Pearson Product-Moment Correlation Coefficient, available at: http://www.gifted.uconn.edu/siegle/research/correlation/corrchrt.htm (last access: 23 August 2013), 2009.



Discussion

Paper

Discussion Paper

Discussion Paper

Discussion Paper

10, 12793–12827, 2013

Time series analysis with sample applications in geosciences

W. Gossel and R. Laehne

Title Page

Abstract

Introduction

References

Figures

 Table 1. Overview of data sets analyzed with time series analysis methods.

Application	Analysis method	Data span	Data resolution	e/ne
Climatic data Berlin	Trend analysis Time dependent correlation FFT Lomb Wavelets STFT Period scanning Autocorrelation	Jan 1851– Dec 2010	Monthly data Non-equidistant: Oct deleted annually	e e/ne e/ne e/ne e/ne e/ne e
Hydrodraph Campus	Trend analysis Time dependent correlation FFT Lomb Wavelets STFT Period scanning Autocorrelation Variogram analysis	29 Feb 2004– 31 Oct 2009	Daily data Non-equidistant: Irregular removal of 7 days, 10 % of the data	e e/ne e e/ne e/ne e/ne e/ne
Gamma log sedimentary sequence	Trend analysis Time dependent correlation FFT Lomb Wavelets STFT Period scanning Variogram analysis	50 m profile in 2 quarries	0.1 m	e e e e e e e
Quaternary base in cross section	Trend analysis Time dependent correlation FFT Lomb Wavelets STFT Period scanning Autocorrelation Variogram analysis		50 m	e e e e e e e e e

e/ne = equidistant/non-equidistant.





Fig. 1. Monthly precipitation, evapotranspiration, and climatic water balance of Berlin 1851–2010. The gaps for testing the capabilities of the methods are set every October.





Fig. 2. Daily water levels at the Heide Campus of the Martin-Luther University Halle (Saale), Germany, 29 February 2004–31 October 2009. The irregularly distributed gaps are marked in blue. The simulated groundwater hydrograph is derived from the results of the period scanning and the FFT and recognizes the 10 most significant periods resulting from period scanning for the simulation.







Fig. 3. Digitized profile of the gamma-ray log of the Lower Buntsandstein and location map.





Fig. 4. Cross section of the Quaternary base in the investigation area in Lower Saxony. A map overview of the structures is given in the inserted picture.







Interactive Discussion

12820





Full Screen / Esc

Printer-friendly Version



Fig. 7. Results of the period scanning for the equidistant time series and the reduced time series of the monthly Berlin climatic water balance. The correlation coefficient for the significance level of 95 % for the equidistant time series is 0.037 and for the reduced time series it is 0.053.







Fig. 8. Comparison of autocorrelation and variogram. The inserted x-y-plot shows a very good correlation of 0.95 ($R^2 = 0.898$).





Fig. 9. Periodogram of the sedimentary sequence of the Lower Buntsandstein with period scanning, FFT and Lomb.



Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Back

Discussion Paper



Fig. 10. Results of the wavelet analysis (left) and STFT (right) for the sedimentary sequence of the Lower Buntsandstein.





Fig. 11. Periodogram of the Quaternary base in Lower Saxony in a cross section (FFT, Lomb and period scanning).







