

# Characterizing ecosystem-atmosphere interactions from short to interannual time scales

M. D. Mahecha<sup>1</sup>, M. Reichstein<sup>1</sup>, H. Lange<sup>2</sup>, N. Carvalhais<sup>3</sup>, C. Bernhofer<sup>4</sup>, T. Grünwald<sup>4</sup>, D. Papale<sup>5</sup>, and G. Seufert<sup>6</sup>

<sup>1</sup>Max-Planck-Institut für Biogeochemie, P.O. Box 10 01 64, 07701 Jena, Germany

<sup>2</sup>Norsk Institutt for Skog og Landskap, P.O. Box 115, 1431 Aas, Norway

<sup>3</sup>Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

<sup>4</sup>Institut für Hydrologie und Meteorologie, Technische Universität Dresden, 01062 Dresden, Germany

<sup>5</sup>DISAFRI, Università degli Studi della Tuscia, Via Camillo de Lellis, snc – 01100, Viterbo, Italy

<sup>6</sup>Climate Change Unit, European Commission – Joint Research Centre, Via E. Fermi 1, 21020 Ispra, Italy

Received: 3 April 2007 – Published in Biogeosciences Discuss.: 2 May 2007

Revised: 27 August 2007 – Accepted: 11 September 2007 – Published: 13 September 2007

**Abstract.** Characterizing ecosystem-atmosphere interactions in terms of carbon and water exchange on different time scales is considered a major challenge in terrestrial biogeochemical cycle research. The respective time series currently comprise an observation period of up to one decade. In this study, we explored whether the observation period is already sufficient to detect cross-relationships between the variables beyond the annual cycle, as they are expected from comparable studies in climatology.

We investigated the potential of Singular System Analysis (SSA) to extract arbitrary kinds of oscillatory patterns. The method is completely data adaptive and performs an effective signal to noise separation.

We found that most observations (Net Ecosystem Exchange, *NEE*, Gross Primary Productivity, *GPP*, Ecosystem Respiration, *R<sub>eco</sub>*, Vapor Pressure Deficit, *VPD*, Latent Heat, *LE*, Sensible Heat, *H*, Wind Speed, *u*, and Precipitation, *P*) were influenced significantly by low-frequency components (interannual variability). Furthermore, we extracted a set of nontrivial relationships and found clear seasonal hysteresis effects except for the interrelation of *NEE* with Global Radiation (*R<sub>g</sub>*).

SSA provides a new tool for the investigation of these phenomena explicitly on different time scales. Furthermore, we showed that SSA has great potential for eddy covariance data processing, since it can be applied as a novel gap filling approach relying on the temporal correlation structure of the time series structure only.

## 1 Introduction

Eddy covariance measurements at tower sites provide time series of CO<sub>2</sub>, H<sub>2</sub>O, and energy fluxes which can be used to characterize the temporal development of ecosystem-atmosphere interactions (Aubinet et al., 2000; Baldocchi, 2003). The data are collected globally within a variety of international projects (e.g., CarboEuropeIP, Fluxnet-Canada, Ameriflux). Currently, many sites are close to producing continuous flux data records that are more than a decade in length, permitting ecosystem fluxes to be investigated on a variety of time scales beyond seasonal and annual cycles (Baldocchi, 2003; Saigusa et al., 2005; Wilson and Baldocchi, 2000). The observed fluxes can be regarded as ecophysiological responses to meteorological and climatological conditions but also to other types of external and intrinsic ecosystem modifications (Baldocchi, 2003), and thus the observed flux variability can be attributed at least in part to the variability of the driving variables (Law et al., 2002; Richardson et al., 2007).

The time-scale dependencies of cross-relationships between the variables are quite well understood for the short term (from hours to seasonal patterns) in a variety of micro-meteorological, ecophysiological, and statistical aspects (Baldocchi, 2003). Moreover, it is well known that some of the driving variables (e.g., temperature and precipitation) depict climate-induced low-frequency oscillations and trends (Ghil and Vautard, 1991; Plaut and Vautard, 1994; Paluš and Novotná, 2006). A self-evident hypothesis is whether long-term temporal structures are detectable in existing flux data, given the current time series length. Consequently, much effort has been made since the initial eddy covariance measurement setups to investigate ecosystem flux variability beyond the annual cycle (e.g., Goulden et al., 1996), which is crucial for assessing ecosystems under

Correspondence to: M. D. Mahecha  
(miguel.mahecha@bgc-jena.mpg.de)

changing environmental conditions (Dunn et al., 2007). In this context, three main questions arise: (a) Can we identify and provide accurate descriptions of flux variability on different temporal scales? (b) Is it possible to detect and describe the statistical properties of “interannual variability”? (c) are (possibly nonlinear) trends identifiable in the time series? All of these are tied to the problem of identifying and separating the relevant time scales of the observed time series.

Related research fields have already provided a variety of approaches for the investigation of time series on multiple time scales including low-frequency oscillations and trends (von Storch and Zwiers, 1999). One method for extracting signals from time series is “Singular System Analysis” (SSA), an approach originating in systems dynamics (Broomhead and King, 1986). The method is highly superior to the well-known Fourier analysis, since it is fully phase- and amplitude-modulated (Allen and Smith, 1996) and suitable for analyzing short and nonstationary signals (Yiou et al., 2000). SSA had already been successfully applied to problems in the fields of hydrometeorology (Shun and Duffy, 1999), hydrology (Lange and Bernhardt, 2004), climatology (Plaut and Vautard, 1995; Ghil et al., 2002), and oceanography (Jevrejeva et al., 2006). This study aims at exploring the potential of SSA within the context of eddy covariance flux data. The idea behind SSA is that each observed time series is a set of (linearly) superimposed sub-signals (Golyandina et al., 2001). In other words, we investigated whether SSA could provide an option for the extraction of components of ecosystem fluxes corresponding to different time scales. Partitioning a time series into sub-signals is thought to separate long-term signals from the annual cycle and high-frequency components (Yiou et al., 2000; Ghil et al., 2002).

Our goal with this study is to provide a brief methodological introduction to SSA, including test statistics and the derived gap-filling strategy. The results and discussion focus on the variance allocation of different time scales in a set of fundamental observations. For characterizing ecosystem-atmosphere interactions, particular attention is given to net ecosystem exchange  $NEE$ , gross primary productivity  $GPP$ , ecosystem respiration  $R_{eco}$ , temperature  $T$ , global radiation  $R_g$ , precipitation  $P$ , vapor pressure deficit  $VPD$ , latent heat  $LE$ , sensible heat  $H$ , and wind speed  $u$ . Their temporal behavior and cross-relationships are explored on a range from intra- to interannual time scales. Finally, the methodological innovations for and limitations on future data adaptive ecosystem assessments and forecasts are highlighted.

## 2 Methods

### 2.1 Singular system analysis

The goal of SSA is to identify sub-signals of a given time series  $X(t)$ ,  $t=1, \dots, N$  and to project them to the correspond-

ing temporal scales. The time series (centered to zero mean) is subjected to SSA, which can be described as a two-step procedure consisting of a signal decomposition and a signal reconstruction (Golyandina et al., 2001). The decomposition aims at finding relevant orthogonal functions, which enables the partial or, if required, entire reconstruction of the time series.

The analysis first needs the a priori definition of an embedding dimension, which is a window of length  $P$ . Sliding the window along the time series leads to a trajectory matrix consisting of the sequence of  $K=N-P+1$  time-lagged vectors of the original series. The  $P$  dimensional vectors of the trajectory matrix  $\mathbf{Z}$  are set up as described in Eq. (1), (Golyandina et al., 2001).

$$\mathbf{Z}_i = (X(i), \dots, X(i+P-1))^T \quad 1 \leq i \leq K \quad (1)$$

Based on the trajectory matrix  $\mathbf{Z}$  a  $P \times P$  covariance matrix  $\mathbf{C} = \{c_{i,j}\}$  is built, which according to Vautard and Ghil (1989) can be estimated directly from the data in form of a Toeplitz matrix; see Eq. (2).

$$c_{i,j} = \frac{1}{N-|i-j|} \sum_{t=1}^{N-|i-j|} X(t)X(t+|i-j|) \quad (2)$$

The entries of the resulting  $P \times P$  matrix represent the captured covariance and depend on the lag  $|i-j|$  only, where  $i, j=1, \dots, P$ . Based on this lag-covariance representation, one can determine the orthonormal basis by solving Eq. (3).

$$\mathbf{E}^T \mathbf{C} \mathbf{E} = \Lambda \quad (3)$$

In this equation,  $\mathbf{E}$  is a  $P \times P$  matrix containing the eigenvectors  $E_i$ , also called empirical orthogonal functions (EOFs) of  $\mathbf{C}$ . The matrix  $\Lambda$  contains the respective eigenvalues in the diagonal, sorted by convention in descending order  $diag(\Lambda) = (\lambda_1, \dots, \lambda_P)$ , where  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_P$ . It can be shown that due to the properties of covariance matrix  $\mathbf{C}$  – preserving symmetry and being real valued and positive semidefinite – all eigenvectors and eigenvalues are real valued, where the latter are nonnegative scalars. The eigenvalues are proportional to the fraction of explained variance corresponding to each EOF. In analogy to the well known Principal Component Analysis, the decomposition allows the construction of principal components (PCs) as generated time series representing the extracted orthogonal modes (Eq. 4). This is why SSA is often also called a “PCA in the time domain.”

$$A^\kappa(t) = \sum_{j=1}^P X(t+j-1)E^\kappa(j), \quad 1 \leq \kappa \leq P \quad (4)$$

As it can be seen in Eq. (4), the principal components are obtained by simply projecting the time series onto the EOFs. This projection constructs a set of  $P$  time series of length  $K$ .

The last step in SSA is the reconstruction of the time series through the principal components  $A^\kappa(t)$ , see Eq. (5). The

original signal can be fully or partially reconstructed. This is a selective step, and the analyst has to decide which  $A^{\kappa}(t)$  are combined so that one obtains an interpretable combination of principal components. This enables signal-noise separation and the reconstruction of specifically selected frequency components, as illustrated by Eq. (5).

$$R^{\kappa}(t) = \frac{1}{M_t} \sum_{\kappa \in \mathcal{K}} \sum_{j=L_t}^{U_t} A^{\kappa}(t-j+1)E^{\kappa}(j) \quad (5)$$

In this reconstruction procedure,  $\kappa$  is an index set determining the selection of modes used for the reconstruction,  $M_t$  is a normalization factor, and the corresponding extension for the series boundaries are given by  $L_t$  and  $U_t$  (definitions for the boundary terms are given in Tab. 1; a comprehensive derivation can be found in Ghil et al., 2002).

The selective time series reconstruction creates the opportunity of depicting the behavior of the series explicitly on different temporal scales. The time scale of variation corresponding to an EOF or PC can be found by analyzing their respective power spectra. The individual modes usually have a very simple spectrum, being dominated by a single dominant frequency only. Vautard et al. (1992) pointed out that the summation of the power spectra of the PCs preserves the fundamental features of the power spectrum of the original series. However, the “embedding dimension”  $P$  sets some limits: The lowest frequency recovered by the individual modes has a period  $\leq P$  (Ghil et al., 2002). Periodicities of length  $P$  either correspond to oscillations with period  $\leq P$  or are induced by (possibly nonlinear) trends (Yiou et al., 2000).

As the objective of this study is to also explore explicitly long-range structures in the data, we searched for the greatest reliable value of the embedding dimension. The choice of  $P$  is a trade-off. On the one hand, maximizing the information content of the analysis requires a large embedding window  $P$ . On the other hand, it is crucial for optimizing the statistical confidence of the decomposition to use a high number of channels  $K$ . This balance is expressed by the ratio  $\frac{N}{P}$ , which was minimized here because the investigated time series are quite short ( $N=8.5$  yr) for the purpose of finding modes of interannual variability. We used the lowest ratio reported in the literature  $\frac{N}{P}=2.5$  (Lange and Bernhardt, 2004), which is equivalent to  $P=3.4$  yr throughout the entire analysis.

## 2.2 Signal selection and separability

The finest temporal resolution for the selective time series reconstruction can be determined by reconstructing the series individually for each mode. However, a typical phenomenon in the decomposed representation of the time series is the appearance of two principal components of almost identical structure and period length but with opposite parity (phase shift  $\pi/2$ ). This can be explained by the fact that the representation of periodic modes requires at least two linear PCs (Hsieh and Wu, 2001). Thus the first step is to reconstruct

**Table 1.** The values of the normalization factor  $M_t$  and of the lower  $L_t$  and upper  $U_t$  bounds of summation.

Temporal locations	$M_t$	$L_t$	$U_t$
for $1 \leq t \leq P-1$	$t^{-1}$	1	$t$
for $P \leq t \leq K$	$P^{-1}$	1	$P$
for $K+1 \leq t \leq N$	$(N-t+1)^{-1}$	$t-N+P$	$P$

the analyzed time series based on identified corresponding oscillatory modes of equal dominant frequency. Whether two components set up as a quasi oscillatory pair has to be checked manually for modes of similar dominant frequency. Golyandina et al. (2001) described this heuristic procedure as “Caterpillar SSA,” which is also beneficial for detecting unexpected oscillations. This is facilitated by reordering the eigenvalues according to the dominant frequency of the associated EOF (Allen and Smith, 1996). After normalizing the eigenvalues, this illustrates the variance allocation on the identified dominant frequencies. This can be regarded as a “discretized power spectrum” (Shun and Duffy, 1999), which is commonly called the “eigenspectrum” of a given series.

However, the eigenspectrum does not overcome the critical point of finding a heuristic approach to signal selection for the subsequent interpretations. An useful alternative is to test the null hypothesis that the SSA output is compatible with a red noise assumption. The assumption of “red noise” seems the most appropriate null hypothesis in geosciences, since many records (e.g., air and sea surface temperature, river runoff, climate indices) usually depict “red-dened” spectral properties (Ghil et al., 2002). In fact, this can also be shown for *NEE* time series from the eddy covariance measurements by investigating their specific autocorrelation, which is nontrivial (Richardson et al., 2007b). This can be regarded as statistical validation for the “red noise” null hypothesis. Several Monte Carlo SSA (MCSSA) approaches were developed for testing whether the eigenspectrum is compatible with an equivalent spectrum corresponding to a set of surrogate data generated through an autoregressive process of the first order (*AR*(1); for different test variants see, e.g., Allen and Smith, 1996; Paluš and Novotná, 1998). Here, we followed an approach introduced by Shun and Duffy (1999) that in terms of computation was much more effective and straightforward, and in which the analytic expression for the red noise spectrum is fit to the eigenspectrum:

$$\phi(f) = \frac{a}{b + (2\pi f)^2} \quad (6)$$

In this equation,  $a$  and  $b$  are process parameters, whereas  $f$  represents the observed frequencies. The model is first fit to the overall eigenspectrum and a 95% confidence interval is calculated. The fit is then repeated for the nonsignificant

fraction of the eigenspectrum only. This repetitive model fitting is stopped if no additional significant modes are identified. This approach is not very common, therefore we ran several tests for the present study where the standard MC-SSA proposed by Allen and Smith (1996) based on more than 300 iterations was shown to produce similar results than the above model given by Shun and Duffy (1999).

### 2.3 SSA for time series with missing values

The classic SSA variant presented is applicable to equidistant time series without missing values. Kondrashov and Ghil (2006) introduced an iterative SSA gap-filling strategy, which allows (in the sense presented above) a time series reconstruction feasible for fragmented time series. This can be interpreted as an opportunity for applying SSA to fragmented records or as a new tool for gap filling. The method consists of a two-loop gap filling, which for the sake of simplicity is described here in form of a “how to” recipe:

1. The first step is to remove the time series mean, where the mean is estimated from the present data only.
2. An inner-loop iteration starts as SSA of the zero-padded time series. The leading (highest eigenvalue) reconstructed component (RC) is used to fill the values in the gaps. This allows a new estimate of the time series mean, which is used for recentering where the padded values are set to their reconstructions. This procedure is carried out based on the computed and recomputed RCs until a convergence criterion is met. Since the original publications did not specify the type and value of convergence, we used the correlation coefficient between the subsequently filled time series and stopped the iteration when  $r^2 > 0.98$ .
3. After the first inner-loop iteration meets the convergence criterion, the method switches to an outer-loop iteration. This is the natural extension of the described procedure above, achieved by simply adding a second (third, etc.) newly reconstructed component to inner-loop iteration.

### 2.4 Eddy covariance data processing

The eddy covariance data were collected based on a 20Hz sampling frequency and subsequently aggregated to obtain the half-hourly fluxes following the EUROFLUX methodology (Aubinet et al., 2000). The CO<sub>2</sub> data were corrected for canopy storage and  $u_*$  filtered to avoid measurements of insufficient turbulence (Papale et al., 2006). A fundamental problem of eddy covariance measurements is the intricate measurement setup, which leads to gaps of varying length, from hours to months (Richardson and Hollinger, 2007).

The primary gap filling of the half-hourly data is possible through a variety of methods which do not lead to fundamental differences (for a broad statistical comparison see Moffat

et al., 2007). All gap-filling methods developed (or adapted) for processing eddy covariance data make use of empirical knowledge on cross-relationships between the different variables. The gap filling at this level works very well when gaps are rare and their distribution is random throughout the time series. However, these gap-filling algorithms fail in the presence of large gaps, where no secondary information, e.g., on meteorological conditions, is available. The simultaneous occurrences of instationarities in related meteorological variables (e.g., during spring green-up) and large gaps induce additional high ranges of uncertainties when conventional gap-filling algorithms are applied (Richardson and Hollinger, 2007). These findings provide some motivation for adopting different strategies to fill gaps based on the long-range autocorrelation structure of the time series.

The obvious approach would be to use SSA for gap filling and further analyzing these half-hourly data. However, due to the high computational memory demand of SSA, this was not possible. In general, the SSA bottleneck constructing the lag-covariance matrix limits the application to time series with several thousands of observations. Furthermore, the goal of this study was to extract modes of variability from intermediate to low frequencies. Hence, we adhered to a two-step gap-filling strategy: we first used the method provided by Reichstein et al. (2005, Appendix A), which is a locally data-adaptive look-up table for filling missing values on a half-hourly basis. The data were subsequently averaged to obtain a daily sampling frequency. At this scale, the data received quality flags indicating the fraction of filled (half-hourly) data per day. An aggregated daily value was considered missing if the amount of original observations fell below 90%. These values were considered highly uncertain and re-estimated in a second gap-filling step by the univariate SSA strategy introduced. The SSA and all related analyses were performed on these preprocessed daily flux data. The same two-step gap-filling and aggregation procedures were applied to all analyzed time series. The complete procedure including data preprocessing and SSA is summarized in Fig. A1 as a flowchart-pseudocode.

### 2.5 Site description

The eddy covariance data were measured at the Anchor Station Tharandt (50°57'49" N, 13°34'01" E, 380 m a.s.l.), which is ca. 25 km SW of Dresden, Germany, and corresponds to a suboceanic/subcontinental climate. Long-term meteorological records indicate a mean annual air temperature of 7.8°C and a mean annual precipitation sum of 823 mm. The dominant wind direction is SW. The area has been episodically affected by summer droughts. The Anchor Station Tharandt has been a spruce stand (72% *Picea abies* L. (Karst.)) since 1887, interspersed with further coniferous evergreen (15%) and deciduous species (13%). A detailed description of the environmental conditions of the site, including descriptions

of the EC data recording and meteorological data, is provided by Grünwald and Bernhofer (2007).

### 3 Results and discussion

#### 3.1 Significant frequencies and variance allocation

The eigenspectra for the analyzed ecosystem variables ( $NEE$ ,  $GPP$ ,  $R_{eco}$ ,  $VPD$ ,  $P$ ,  $T$ ,  $R_g$ ,  $LE$ ,  $H$ , and  $u$ ; Fig. 1) reveal a set of nontrivial patterns; a variety of significant frequencies was found in the time series. The identified components are summarized in Tab. 2.

Analyzing the eigenspectra from the low- to the high-frequency domain shows that several variables are significantly driven by interannual variability (Tab. 2). The derived carbon and energy fluxes,  $GPP$ ,  $R_{eco}$ ,  $LE$ , and  $H$  depict significant modes at the very edge of their eigenspectra, indicating the presence of even longer-term structures than effectively describable by the applied embedding procedure. These nonextractable subsignals from minimum frequency EOFs can be either oscillations or trends, which cannot be determined at the given time series length. In the following, these patterns will be summarized as edge cycles. It is important to note that similar components are also contained in most of the other variables, but not on a significant level, so that in the overall view we expect the signals to contain more low-frequency modes.

Apart from the identified edge cycles, most time series depict oscillations with periods of around 1.4, 1.7, 1.8, and 2.3 yr., except  $T$ . The question remains why  $T$  does not depict low-frequency modes within the observation period when even  $R_{eco}$  (estimated from  $NEE$ ,  $T$ , and a reference respiratory component) does so. It is a well-known phenomenon in the northern hemisphere that temperature time series that integrate large geographical areas depict a set of oscillatory patterns and trends beyond the annual scale, (e.g., Ghil and Vautard, 1991). Quasi biennial oscillations in the climate system, for example, are visible in temperature records and can also be traced back to the NAO index (Paluš and Novotná, 2006). The absence of similar observations could be either an artifact of the very short record or a local climatic property. Paluš and Novotná (1998) already showed that components of variability of space-replicated time series are highly dependent on their geographical location.

With respect to the annual cycle, it is obvious that it explains most of the variance in all fluxes, except for precipitation  $P$ , which exhibits very peculiar behavior in the overall eigenspectrum. The  $P$ -eigenspectrum indicates that one can assume that we are dealing with an almost entirely noise-dominated signal (97.5% of the signal is compatible with the red noise model), which is consistent with previous findings in the literature (Shun and Duffy, 1999; Tessier et al., 1996) if the temporal resolution is not too small (e.g., daily). On very small time scales, however, rainfall also exhibits rather

intriguing multiscaling behavior. Peters et al. (2002), for example, uncovered a characteristic power law decay of rain events as a function of their size when inferring precipitation at 1min intervals. The aggregation of the data to daily precipitation integrals in this study eliminates such effects and provides an explanation for why the precipitation eigenspectrum appears much closer to noise than the other variables.

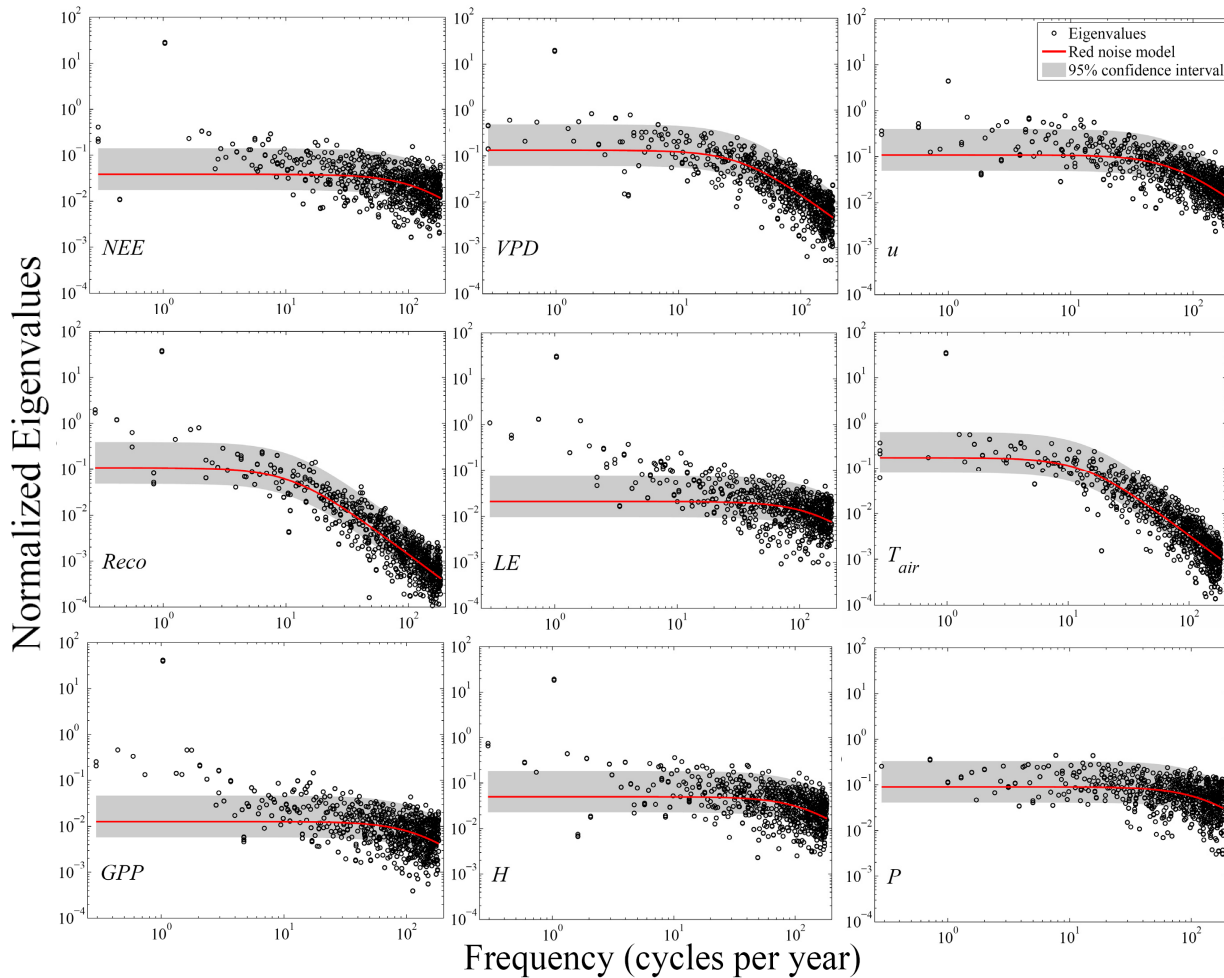
The highest frequency components detectable by SSA in the present case lay within the range of intraannual variability. Noticeably, a semiannual component was found to be common to most ecosystem variables where the exceptions are again  $T$  and  $P$ . One has to take into account the relative importance of the different subsignals, as for some variables the significant high frequency components contribute a considerable fraction of variance to the entire time series (e.g.,  $NEE$ : 6.5%,  $GPP$ : 5.1%,  $VPD$ : 5.7%,  $LE$ : 12.5%,  $H$ : 9.8%, and  $u$ : 14%, see Table 2). In the overall view, however, the vast majority of high-frequency components is not considered to be significant (see Fig. 1). The reconstruction of the time series from subsignals within the range of the red noise assumption is predominantly a high-frequency signal. This reconstruction could be of interest for further investigations focusing on short-term ecosystem-atmosphere relationships.

The results of the eigenspectra rely on the decomposition of the time series lag covariance structure as described in Eq. (3). Yet only the subsequent time series reconstruction provides insight into the effective behavior of the observation on the respective time scale. Figure 2 presents an example for the reconstruction of  $NEE$ . The figure also shows clearly that it is not discernable whether the interannual variability is a sinusoidal oscillation or a (possibly nonlinear) trend. Furthermore, the semiannual component depicts a strong amplitude modulation, indicating that its importance in the time series structure is also time-dependent. It can also be seen how a substantial part of the time series variance (38.1%) is compatible with the red noise model according to the test arrangement. As indicated above, for the most part this signal is composed of high-frequency components. Finally, the bottom panel is an example for how the fundamental time evolution of  $NEE$  was reconstructed based on very few significant modes.

#### 3.2 Scale-dependent $NEE$ - $T$ relationships

As highlighted in Table 2 and Fig. 1, it was possible to reveal that the different variables are dominated by a variety of subsignals, which can be in part attributed to comparable frequency classes. Thus it is possible to assess the cross-relationships of the data explicitly on the identified time scale rather than establishing empirical relationships from the entire observation space, as is the standard approach in most classic regression approaches.

Figure 3 shows the relationship between  $NEE$  and  $T$  where different components are plotted in contrast with each other. The relationship between the pure annual components of



**Fig. 1.** The eigenspectra of the observed time series for *NEE*, *Reco*, *GPP*, *VPD*, *LE*, *H*, *u*, *T*, and *P*. The dots represent the dominant frequencies of the time series at the respective normalized eigenvalue. The latter is equivalent to the amount of allocated variance. The red lines show the fitted red noise model. The gray areas are the corresponding 95% significance intervals.

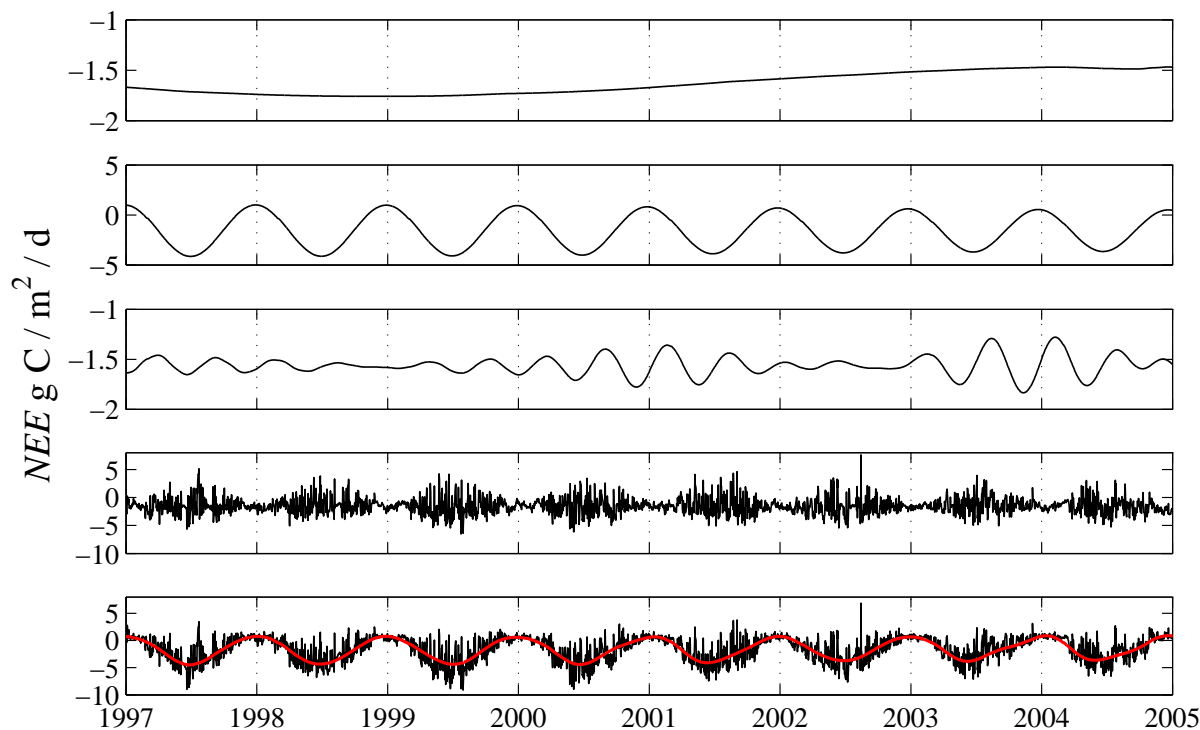
both variables depicts a clear hysteresis effect (Fig. 3, left panel). Adding the significant interannual component to the annual cycles has a remarkable effect on the observed hysteresis (Fig. 3, center panel). Although the interannual component accounts only for a small part of the data variability (Table 2), the figure shows that this is sufficient to induce a trend in the summer fluxes toward decreased carbon uptake. On the contrary, the observed loss of carbon during the winter days found by the relationship of the annual components is neutralized by the interannual component (a more accurate description of the year-to-year anomalies is given in Sect. 3.4 and Fig. 5). The total allocation of carbon which is due to the interannual components is rather small. Despite this, we now have to assert that the contribution of the interannual variability could be essential for assessing the ecosystem under climate change conditions, given the observed forcing of

this component to the entire flux observations. Our findings corroborate the assumption that interannual variability is a major qualitative contribution required to describe the temporal behavior of the net carbon balance in terrestrial ecosystems. These empirical findings are of general importance for modeling purposes, since ecosystem models encounter limitations in the presence of interannual variability (Fujita et al., 2003; Hanson et al., 2004; Richardson et al., 2007; Siqueira et al., 2006).

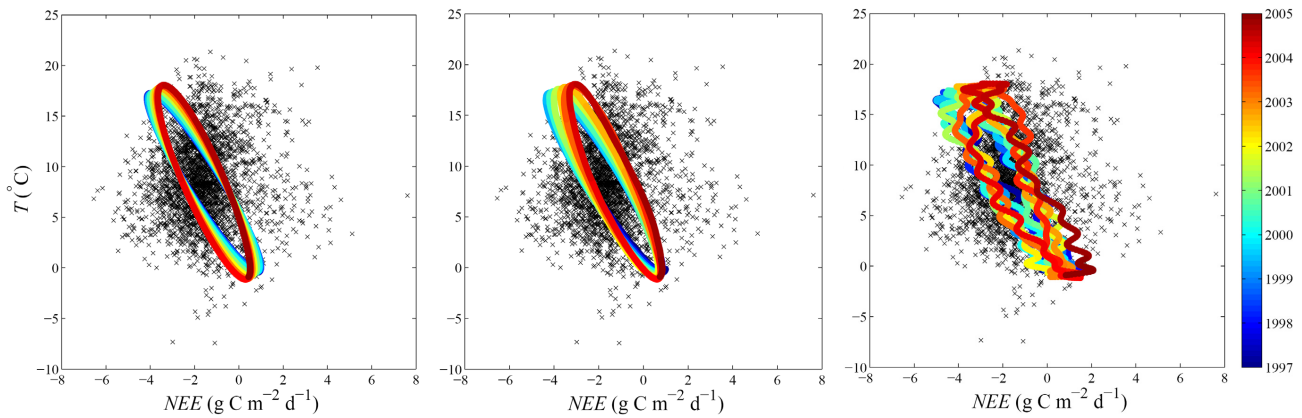
It has been shown that the seasonal and interannual components are not the only significant part of the time series. The intraannual component also plays a crucial role in the flux data, and their contribution induces complex variability (Fig. 3, right panel). These results provide a strong argument for understanding cross-relationships in flux data as a temporal multilayer problem. In particular the components

**Table 2.** Significant periods identified in the flux time series. The periods are given in years (rounded values). Modes at the edge of the maximal observed frequency domain (in this case, 3.4 years) could be either “real” trends or edge cycles. The values of the variables at the respective period indicate the amount of explained variance in percentages. Note that each value is the sum of explained variance for all significant components of a similar period, e.g., the annual components are always set up by two EOFs.

Period $p$ (years)	<i>NEE</i>	<i>GPP</i>	<i>R<sub>eco</sub></i>	<i>T</i>	<i>R<sub>g</sub></i>	<i>VPD</i>	<i>P</i>	<i>LE</i>	<i>H</i>	<i>u</i>
$\geq 3.4$	0.8	0.5	3.7	–	–	–	–	2.2	1.4	–
2.4	–	–	2.4	–	–	0.6	–	–	–	–
2.3	–	0.5	–	–	–	–	–	1.1	–	–
1.8	–	–	0.6	–	–	–	–	–	–	1.4
1.7	–	0.3	–	–	–	–	–	–	0.6	–
1.4	–	0.1	–	–	0.2	0.5	0.7	2.6	–	–
1	55.6	80.7	74.5	69.2	60.6	38.9	–	60.4	37.5	8.8
0.8	–	0.3	0.4	–	–	–	–	0.5	0.9	–
0.7	–	0.1	–	–	–	1.1	–	–	–	0.7
0.6	0.2	0.9	1.5	–	0.5	–	–	2.4	–	–
0.5	0.7	0.4	1.6	–	0.2	1.7	–	0.7	0.7	–
0.4	0.6	0.2	–	–	–	–	–	0.9	–	0.9
0.3	1.4	0.6	–	–	0.5	1.3	–	1.5	1.1	1.2
0.2	0.8	0.3	–	–	0.7	1.6	–	0.8	0.5	3.1
0.1	1.7	1.8	–	–	1.4	–	1.8	4.2	5.2	8.1
<0.1	1.1	0.5	–	–	0.8	–	–	1.5	1.4	–
red noise	38.1	12.9	15.3	30.8	35.1	54.3	97.5	21.3	50.6	75.0



**Fig. 2.** The partly reconstructed NEE flux time series. From top to bottom: the interannual variability; the annual cycle; the semiannual oscillation; the time series reconstruction from the not significant components; the original signal (black line) and the reconstructed signal based on the significant modes (red line) shown above.



**Fig. 3.** The cross-relationship of the partly reconstructed  $NEE$  and  $T$  time series. From left to right: The first panel shows the pure annual components of both variables where the color code permits tracing the temporal evolution. The black crosses show the respective reconstructions from the non-significant components. The center panel time series reconstruction is based on the annual and interannual components. The right panel shows a reconstruction, in which the significant intraannual components were also taken into consideration.

that contain most of the high-frequency variability and were integrated here into the “red noise” component (Fig. 1) do not contain intuitive cross-relationships between  $NEE$  and  $T$  (Fig. 3). However, further investigation is needed to determine whether short-term responses of  $NEE$  to fluctuations in the environment are extractable by means of SSA.

### 3.3 Hysteretic ecosystem-atmosphere relationships

There was evidence that similar hysteretic patterns as found between the seasonal  $NEE$  and  $T$  components can be found in the relationships of  $NEE$  versus  $GPP$ ,  $VPD$ ,  $u$ ,  $LE$ , or  $H$ . Figure 4 shows how these variables are linked when inferred from a joint reconstruction of their annual and interannual components. These hysteresis loops are particularly well pronounced in terms of  $NEE$ - $VPD$ ,  $NEE$ - $u$ , and  $NEE$ - $H$  dependencies. Less clear patterns were found in the  $NEE$ - $LE$  space and the carbon cycle relationship  $NEE$ - $GPP$ .

The question that arises here is whether the hysteresis loops shown are indications for nonlinear cross-relationships in the observation space or if they might instead be due to linear, yet time-lagged cross-relationship in ecosystem-atmosphere variables. Other open questions are whether non-linearity is induced at certain time scales, e.g., by the interannual variability, which was shown to be responsible for transitions of the hysteresis loops in time. All these questions need further investigations and require the application of information-theory-based test statistics (e.g., Paluš, 1996; Diks and Mudelsee, 2000), which is beyond the scope of this paper.

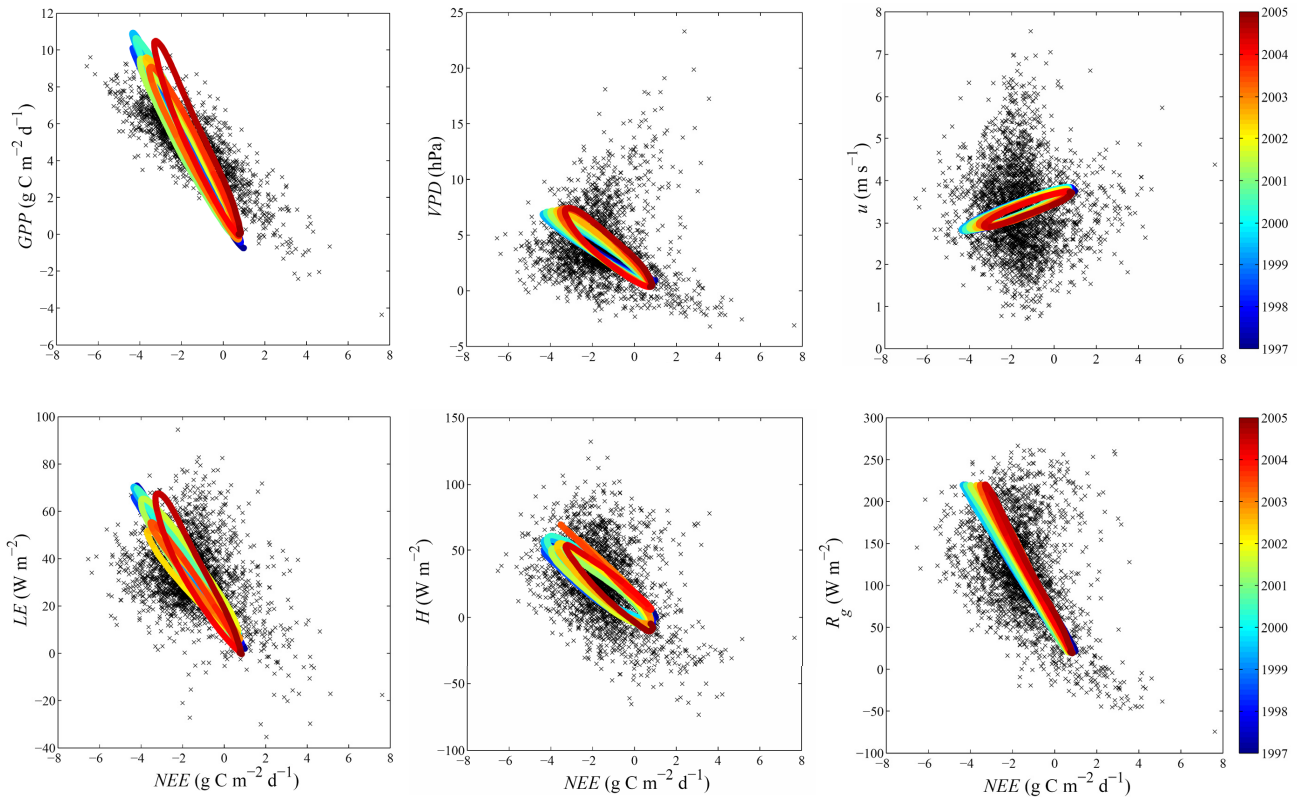
Above all, it proved to be the case that a sole dependency on annual time scales, which can apparently be well captured by linear regression, was found when relating  $NEE$  to  $R_g$ . Having observed that cross-relationships between

RCs originating from different variables can exhibit a wide range of dependencies, this occurrence of an apparently non-lagged linear relation is by no means trivial. Whether this phenomenon is unique for the investigated site and related to the dominance of evergreen coniferous species (implying low variations of the leaf area index) or a more general phenomenon requires further investigation. This observation is in accordance with previous findings that showed that  $R_g$  provides a good basis for modeling terrestrial carbon fluxes.

In the overall view, most cross-relationships between the ecosystem variables contain more or less pronounced hysteresis effects (Fig. 4). There is a sound explanation in particular for the  $NEE$ - $T$  relationship: Owing to a  $R_g$ - $T$  hysteresis (not shown), the  $NEE$ - $T$  dependence is an imparted effect. From an ecological point of view, this might appear trivial and can be understood as the time-varying ecosystem heat capacity in the course of the seasonal cycle. Indeed, the existence of hysteretic behavior of ecosystem-atmosphere interactions is well known, e.g., Nakai et al. (2003) showed hysteretic  $NEE$ - $T$  relationships on short-time scales. With respect to the hysteresis on seasonal time scales, our results are in line with findings from Richardson et al. (2006), who showed similar effects in the relationship between soil temperature and the sensitivity parameter of respiration to temperature ( $Q_{10}$ ).

As far as the magnitude of impact is concerned, the seasonal hysteresis might be of fundamental importance for further ecosystem modeling tasks. Apart from the “obvious”  $NEE$ - $T$  example, the causes for the unexpected but equally well-pronounced hysteresis effects, e.g., in the  $NEE$ - $VPD$  relationships, remain unclear. This could lead to further questions, for example regarding how to distinguish between the simple coexistence of seasonal cycles and their mutual dependence. We presume that the extracted hysteresis loops





**Fig. 4.** The cross-relationships of the major variables with  $NEE$ . Black crosses show the respective red noise components. The color-coded reconstructed components comprise the annual and interannual components of the respective variables, where the color code permits tracing the temporal evolution.

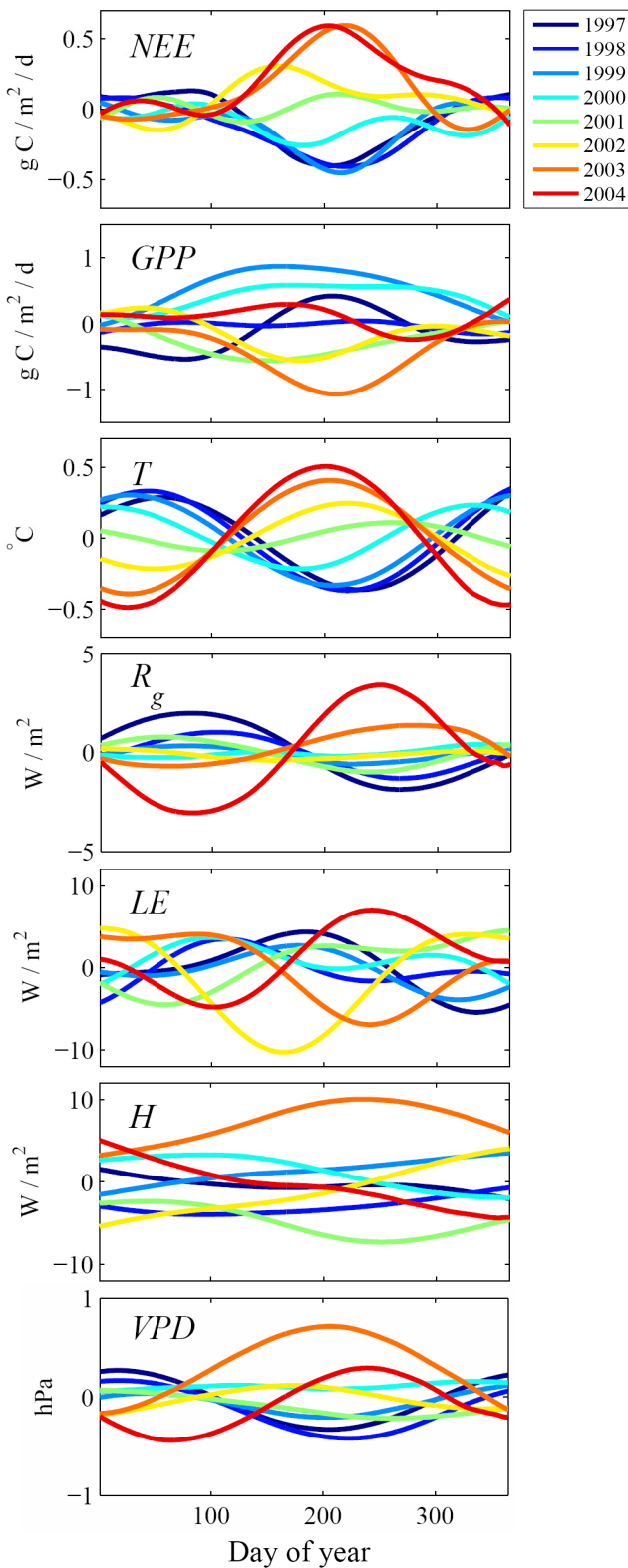
lead to more general problems in ecosystem theory. Additional research on hysteretic ecosystem-atmosphere exchanges could, for example, contribute to a better understanding of ecosystem memory. This is of relevance for the quantification of lag effects in ecosystem responses to any type of recurrences on a variety of time scales, including summer droughts or winter anomalies (Seneviratne et al., 2006). An example is given here by the  $NEE-H$  relationship (Fig. 4, lower central panel), where a clear deviation of the hysteretic behavior occurred precisely during the summer heat wave in 2003 (cf. Ciais et al., 2005; Granier et al., 2007; Reichstein et al., 2007, for a comprehensive discussion of this climate anomaly).

Aside from the relationships of the annual-interannual RCs previously discussed, Fig. 4 also shows the scatters of the red noise components for these variables. As discussed above (shown in Fig. 1 and Table 2), the red noise fraction comprises most of the high-frequency components of the time series. When contrasting the cross-relationships of the non-significant and the annual-interannual components, we observed that there might be different types of relation-

ships acting on different time scales. This phenomenon is highlighted, for example, in the  $NEE-GPP$  space (Fig. 4, upper left panel). In this particular case, simple linear regression analysis would lead to systematically varying parameters, depending on the time scale concerned. It is interesting to note that the red noise  $NEE-R_g$  relationship appears similar to known light-use saturation effects. Overall, these results comprise intuitive examples for how the analysis of an entire signal might obscure the inherent cross-relationship on a seasonal time scale.

### 3.4 Deviations from mean annual-interannual components

One main advantage of SSA is that the partly reconstructed variables can be further analyzed in analogy to existing approaches that conventionally take the whole data series into account. Figure 5 supplies such a simple application example: Based on the annual-interannual components of  $NEE$ ,  $GPP$ ,  $T$ ,  $R_g$ ,  $LE$ ,  $H$ , and  $VPD$ , we estimated a mean seasonal cycle for each of them. In this respect, the capacity of SSA to extract phase- and amplitude-modulated signals allows the



**Fig. 5.** The deviations of the estimated mean reference component of *NEE*, *GPP*, *T*, *R<sub>g</sub>*, *LE*, *H*, and *VPD* based on the joint reconstruction of their respective annual and interannual components

calculation and visualization of the annual deviations of the RCs. This enables us, e.g., to investigate the effect of the summer drought that affected the site in 2003, which is ongoing with deviations in the investigated time series. A positive deviation of *NEE*, for example, turned out to be equally pronounced in the year 2004. By contrast, *GPP* seems to be affected only during 2003, depicting decreased productivity. The *T* record shows a smooth variation over the years, indicating a systematic amplitude shift, which can also be seen in *R<sub>g</sub>*. A more complicated pattern is found in the behavior of *LE*, which is apparently not trivially related to the other variables. From our results, the drought seem to have significant bearing on the deviations of *H* and *VPD*. For an interpretation of these results, it is crucial to understand these deviations again in the context of the range of variation of the entire signal and in terms of cross-relationships. We have to distinguish, for example, between the *VPD*-deviations occurring on quantitatively very small scales (compared to the range of the seasonal-interannual variability; Fig. 4) and the deviations of *H*. The latter is the reason for the well-pronounced deviation form the *NEE*-*H* hysteresis during 2003 (Fig. 4, lower central panel).

Without going into a detailed discussion, this example illustrates one of the strengths and pitfalls of SSA: on the one hand, even simplistic approaches become powerful when the time series are explored explicitly on the relevant temporal scales only. On the other hand, events like the summer drought of 2003 apparently influence the whole annual cycle of some flux time series, but they are not detectable in the temperature cycle. This is suspicious, since especially a summer heat wave was responsible for the anomaly in ecosystem productivity (Ciais et al., 2005). One reason might be that SSA allocates this effect on shorter time scales, of which periodicity matches the length of the heat wave. The open question requiring further investigation is which time scales need to be incorporated to achieve a sound representation of time-localized anomalies.

### 3.5 Comparison of SSA with wavelet analysis

Singular System Analysis is an entirely data-driven method, and the achieved time series decomposition and projection to different time scales can be seen as an “empirical variance partitioning.” This is in contrast to methods using certain functional classes (and usually parametric) as a fixed basis. Of course, fundamental (quasi) oscillatory RCs as the annual cycle can also be identified by a linear combination of weighted sines and cosines as it is commonly applied in terms of Fourier analysis. However, for phase- and amplitude-varying signals (e.g., as it was shown here very clearly for the semiannual cycle of *NEE*; Fig. 2), two EOFs are fully sufficient, whereas a Fourier type analysis would require a large number of coefficients, in particular in the event of sudden changes. Being able to decompose a signal into (possibly) nonharmonic or aperiodic subsignals is considered

one of SSA's major strengths (Yiou et al., 2000). A further advantage is the relative robustness of SSA to instationarities (not restricted to trends) of the signal mean and variance. This allows the analysis of nonstationary time series (Vautard et al., 1992) and is a major reason why using SSA can be specifically advocated for environmental records such as ecosystem-atmosphere fluxes.

Several recent studies investigating the temporal behavior of eddy covariance (and related) time series on varying time scales are based on wavelet analysis (e.g., Katul et al., 2001; Braswell et al., 2005; Stoy et al., 2005; Richardson et al., 2007). Indeed, wavelet analysis shares some important features with SSA, especially the ability to uncover the impact of particular time scales on variation (Torrence and Compo, 1998). Wavelets use a (parametric) time-localized shifting-window approach to extract the phase and amplitude information of a time series. This leads to complex representations in the Fourier domain, but enables *localizing* the importance of a given frequency scale in time: this is why wavelets can be seen as “mathematical microscopes” (Yiou et al., 2000) in time. The opposite is the case for SSA, which is a *global* approach for the retrieval of the variation of sub-signals, as the shifting-window technique is solely exploited to achieve statistical reliability. Wavelets and SSA can be applied as complementary approaches. Jevrejeva et al. (2006), for example, used SSA for extracting nonlinear trends and varying periodicities from sea-level records. In a subsequent step, they used wavelet analysis as an independent tool for confirming their results, gaining additional information from both algorithms.

Yiou et al. (2000) found that SSA and wavelet analysis share fundamental features: they showed, e.g., that the EOFs can be interpreted in a way similar to mother wavelets, and they provide further details (see also Ghil et al., 2002). From this perspective, a multiscale SSA (MS-SSA) was developed, extending the classical variant in analogy to wavelet analysis. The MS-SSA can be understood as the consecutive application of SSA to sliding windows. Roughly put, this is equivalent to applying SSA to each subvector in Eq. (1), where each subvector  $Z_i$  is treated as an independent time series  $X(t)$  in SSA. Obviously, a local embedding dimension  $P^{Z_i}$  needs to be defined. With this they localized transitions in the SSA reconstructions on varying time scales in terms of wavelet analysis. When searching for low frequency components in the available eddy covariance records, one is already operating at the very limit with classic SSA. Thus at the moment, MS-SSA is not a feasible approach to the characterization of ecosystem-atmosphere interactions.

### 3.6 Limitations of SSA and potential extensions

The major problem is that SSA requires the heuristic setting of the embedding dimension  $P$  (or  $P^{Z_i}$ ). Due to the limited series length, we did not have much space for varying the parameter in this study and therefore opted for a ratio of  $\frac{N}{P} = 2.5$ . Additional tests showed that the quality of signal separability suffered by a further decrease of this ratio. Most extracted RCs were not well separable from each other, indicating a high degree of superimposition of signals (for an accurate definition of signal separability cf. Golyandina et al., 2001). Increasing the ratio  $\frac{N}{P}$  also did not lead to a better signal separability and was not justified here. We found that the identified frequencies and their respective variance allocations remain fairly stable over a certain range of embedding dimensions. Major shifts in the variance allocations in the frequency domain occur mainly in the low-frequency modes ( $f < 1\text{yr}^{-1}$ ) due to the decrease of correlation strength in the trajectory space. A less heuristic strategy could be to optimize the embedding dimension by means of more sophisticated procedures (e.g. Cao, 1997), which were developed for an optimal application of Takens' embedding theorem (Takens, 1981).

Apart from these technical aspects of SSA performance, one has to consider the limitation of the method to linear feature extraction. Recall that SSA is called a “PCA in the time domain,” which holds true for the technical implementation of the decomposition as well as for the essential PCA outcome, which is to achieve a dimensionality reduction. The main limitation of the presented SSA variant is its linear representation of the embedding space, which is inherited from PCA. The question whether this is a suitable approach for dealing with real-world data is not trivial. E.g., it is well known that PCA is a suboptimal solution for dimensionality reduction in nonlinear systems and could be replaced by nonlinear dimensionality reduction (Kramer, 1991). In accordance with these findings, Hsieh and Wu (2001); Hsieh and Hamilton (2001) and Hsieh (2004) showed that the SSA results can be nonlinearly generalized at different levels, e.g., at the decomposition step or by generalizing the RCs with artificial neural networks. On the one hand, these features are principally challenging where the main advantage is that oscillatory modes of arbitrary shape can be represented by one single nonlinear component (Hsieh, 2004). On the other hand, such generalizations are fundamentally questionable; the critical point is that the neural network approaches mentioned rely exclusively on previously linearly generated features, since the direct nonlinear feature extraction always leads to a parameter space that exceeded the number of observations. In this context, new developments of data exploration (e.g. Coifman et al., 2005; Tenenbaum et al., 2000) could provide alternative solutions for the next generation of SSA.

### 3.7 Pattern extraction with missing data

In the context of eddy covariance measurements, missing data and rejected values, e.g., due to unfavorable conditions, are unavoidable (Foken and Wichura, 1996). In terms of SSA applications this is especially problematic where more than 10% of data points were rejected and the method operates at the very edge of a reasonable gap-filling procedure (Kondrashov and Ghil, 2006).

Here, we applied SSA to a gappy time series, and the overall results did not reveal fundamental problems. The SSA gap-filling strategy has the advantage that it interprets only available temporal correlation structures. This leads to a further analogy to wavelet analysis; Katul et al. (2001) showed that missing data do not necessarily limit the investigation of eddy covariance data in the frequency domain. They affirm being able to “remove the effect of missing values on spectral and co-spectral calculations.” All the same, this did not provide a tool for filling large gaps in time series where SSA has some potential. This potential needs further exploration, especially considering the multivariate SSA gap-filling strategy proposed by Kondrashov and Ghil (2006). The latter would make use of cross-correlations between ancillary variables where available, and could switch to the univariate gap filling presented above in the absence of secondary information.

The present study could not provide an in-depth quality analysis comparing the SSA-based gap filling with established techniques as they are currently used for eddy covariance data (Moffat et al., 2007). The introduction of SSA to the analysis of eddy covariance data only hints at how available gap-filling methods could be improved by incorporating the information lying on the temporal correlation structure. It has to be pointed out that the SSA gap-filling method is itself not in a fully mature stage of development (cf. the current state of the discussion: Kondrashov and Ghil, 2007; Schneider, 2007) and alternative approaches were recently formulated (Golyandina and Osipov, 2007). Despite these details, SSA gap filling is already a considerable step forward toward dealing with cases where, e.g., due to power outages none of the target variables or standard meteorological observations are available (see Richardson and Hollinger, 2007, for an in depth sensitivity analysis of data uncertainty in the presence of long gaps).

### 3.8 Implications for future flux assessment and outlook

We have shown that SSA provides a new way for the partitioning of ecosystem flux variability into different time scales. There is considerable research interest in such data-adaptive tools for assessing ecosystem-atmosphere carbon, water, and energy fluxes. The application of advanced time series techniques could reveal further unexpected patterns that were previously hidden in the raw data. There are many possible fields of application for SSA and related methods

in addition to the presented “variance partitioning,” e.g. advanced flux-partitioning methods that separate the contributions of  $GPP$  and  $R_{eco}$  to the net carbon flux. Comparisons between different flux-partitioning methods showed that there are essential differences between the different approaches, but also that all known methods make comparable assumptions on underlying physiological kinetics (Stoy et al., 2006). The question is whether these models could be refined and validated on different time scales, as discovered by applying SSA. With this we aim at stimulating further discussions on how future flux-partitioning algorithms could work, taking the different behavior of the variables on different time scales into account.

A further perspective relevant to the analysis of eddy covariance data is presented by Golyandina and Stepanov (2005). They showed how SSA could be expanded to a statistical time-series forecast. This problem is closely related to the problem of gap-filling, where we showed the usefulness of SSA.

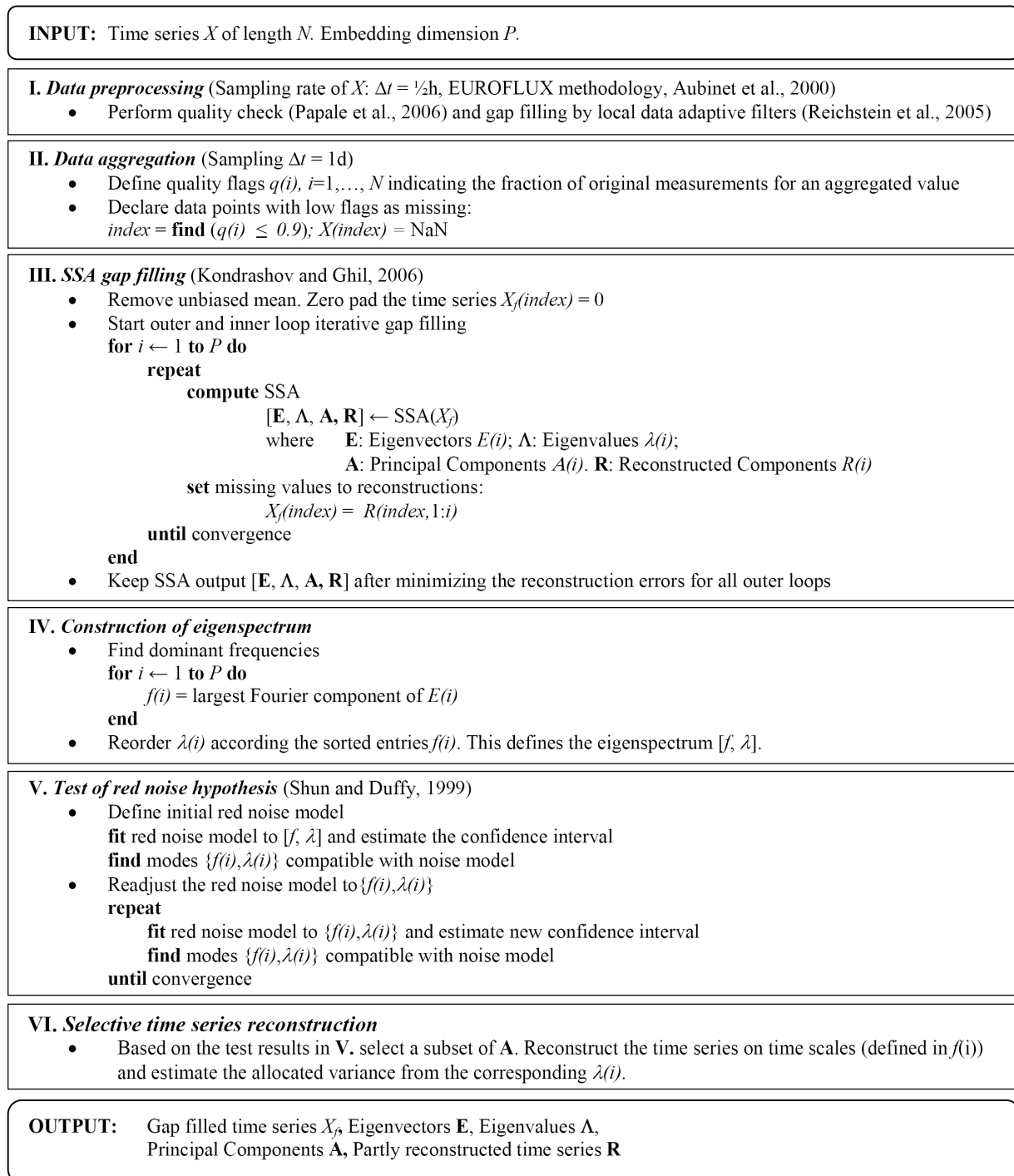
This study highlighted a series of relevant applications of SSA inferring ecosystem-atmosphere interactions from varying time scales. Extending the observation period or realizing multi-site analysis in future investigations is expected to provide major insight into cross-relationships between ecosystem fluxes and both related meteorological and derived eco-physiological time series. This could improve our general understanding of ecosystem-atmosphere relationships.

## 4 Conclusions

Ecosystem-atmosphere interactions inferred from eddy covariance data now comprise a time window that allows investigating flux behavior beyond annual cycles. This study showed that by applying advanced time-series analysis methods, one can extract a wide range of significant modes ranging from the high-frequency domain to long-term components. One major innovation of the approach presented is that it provides an integrated methodological framework comprising data analysis, including signal to noise enhancement and gap-filling.

SSA proved to be able to separate the inherent temporal scales of a time series. Here, we showed that this “variance partitioning” could serve as the general basis of further data-adaptive or process-oriented modeling tasks. Based on these methodological advances, we were able to demonstrate that the eddy covariance observations are dominated to a nonnegligible extent by interannual variability. Overall, the investigation showed that the relationships between water, carbon, and energy fluxes are strongly determined by seasonal hysteresis and vary fundamentally between different time-scales.

## Appendix A



**Fig. A1.** The complete data processing and analysis of the study summarized as as a flowchart-pseudocode.

**Acknowledgements.** We gratefully acknowledge M. Paluš, A. D. Richardson, S. I. Seneviratne, D. Kondrashov, and two anonymous referees for their comments on the manuscript. The continuous help of the technical staff in Tharandt, U. Eichelmann, H. Prasse, H. Hebertanz and U. Postel, is greatly acknowledged. This work is part of the integrated project “Carboeurope” (GOCE-CT2003-505572) of the European Union (EU). This research was funded in part by the Marie Curie European Reintegration Grant “GLUES” (MC MERG-CT-2005-031077). M. D. Mahecha and M. Reichstein would like to thank the Max Planck Society for supporting the “Biogeochemical Model-Data Integration Group” as an Independent Junior Research Unit. N. Carvalhais thanks the Portuguese Foundation for Science and Technology and the EU through the Operational Program “Science, Technology, and Innovation”, (PhD grant SFRH/BD/6517/2001).

Edited by: T. Laurila

## References

- Allen, M. R. and Smith, L. A.: Monte Carlo SSA: Detecting irregular oscillations in the presence of coloured noise, *J. Climate*, 9, 3373–3404, 1996.
- Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology, *Adv. Ecol. Res.*, 30, 113–175, 2000.
- Baldocchi, D.: Assessing the eddy covariance technique for evaluating carbon exchange rates of ecosystems: past, present and future, *Global Change Biol.*, 9, 479–492, 2003.
- Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, *Physica D*, 20, 217–236, 1986.
- Braswell, B. H., Sacks, W. J., Linder, E. and Schimel, D. S.: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, *Global Change Biol.*, 11, 335–355, 2005.
- Cao, L.: Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D*, 110, 43–50, 1997.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogee, J., Allard, V., Buchmann, N., Aubinet, M., Bernhofer, C., Carrara, A., Chevalier, F., Noblet, N. D., Friend, A., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, 437, 529–533, 2005.
- Coifman, R. R., Lafon, S., Lee, A. B., Maggioni, M., Nadler, B., Warner, F., and Zucker, S. W.: Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps, *Proceedings of the National Academy of Sciences*, 102, 7426–7431, 2005.
- Diks, C. and Mudelsee, M.: Redundancies in the Earth’s climatological time series, *Phys. Lett. A*, 275, 407–414, 2000.
- Dunn, A. L., Barford, C. C., Wofsy, S. C., Goulden, M. L., and Daube, B. C.: A long-term record of carbon exchange in a boreal black spruce forest: means, responses to interannual variability, and decadal trends, *Global Change Biol.*, 13, 577–590, 2007.
- Falge, E., Baldocchi, D., Olson, R. J., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grünwald, T., Hollinger, D., Jensen, N.-O., Katul, G., Keronen, P., Kowalski, A. nad Ta Lai, C., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, Ü., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and S., W.: Gap filling strategies for long term energy flux data sets, *Agric. Forest Meteorol.*, 107, 71–77, 2001.
- Foken, T. and Wichura, B.: Tools for quality assessment of surface-based flux measurements, *Agric. Forest Meteorol.*, 78, 83–105, 1996.
- Fujita, D., Ishizawa, M., Maksyutov, S., Thornton, P. E., Saeki, T., Nakazawa, T.: Inter-annual variability of the atmospheric carbon dioxide concentrations as simulated with global terrestrial biosphere models and an atmospheric transport model, *Tellus B*, 55, 530–546, 2003.
- Ghil, M. and Vautard, R.: Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, 350, 324–327, 1991.
- 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, 1–25, 2002.
- Golyandina, N., Osipov, E.: The “Caterpillar”-SSA method for analysis of time series with missing values, *J. Stat. Plan. Infer.*, 137, 2642–2653, 2007.
- Golyandina, N. and Stepanov, D.: SSA-based approaches to analysis and forecast of multidimensional time series, in: *Proceedings of the 5th St. Petersburg Workshop on Simulation*, pp. 293–298, St. Petersburg State University, St. Petersburg, 2005.
- Golyandina, N., Nekrutkin, V., and Zhigljavsky, A.: *Analysis of Time Series Structure: SSA and related techniques*, no. 90 in *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Boca Raton, 2001.
- Goulden, M. L., Munger, J. W., Fan, S. M., Daube, B. C., and Wofsy, S. C.: Exchange of Carbon Dioxide by a Deciduous Forest: Response to interannual Climate Variability, *Science*, 271, 1576–1578, 1996.
- Granier, A., Reichstein, M., Bréda, N., Janssens, I. A., Falge, E., Ciais, P., Grünwald, T., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Facini, O., Grassi, G., Heinesch, B., Ilvesniemi, H., Keronen, P., Knohl, A., Köstner, B., Lagergren, F., Lindroth, A., Longdoz, B., Loustau, D., Mateus, J., Montagnani, L., Nyst, C., Moors, E., Papale, D., Peiffer, M., Pilegaard, K., Pita, G., Pumpanen, J., Rambal, S., Rebmann, C., Rodrigues, A., Seufert, G., Tenhunen, J., Vesala, T., and Wang, Q.: Evidence for soil water control on carbon and water dynamics in European forests during the extremely dry year: 2003, *Agric. Forest Meteorol.*, 143, 123–145, 2007.
- Grünwald, T. and Bernhofer, C.: A decade of carbon, water and energy flux measurement of an old spruce forest at the Anchor Station Tharandt, *Tellus B*, 59, 387–396, 2007.
- Hanson, P. J., Amthor, J. S., Wullschleger, S. D., Wilson, K. B., Grant, R. E., Hartley, A., Hui, D., Hunt, E. R., Johnson, D. W., Kimball, J. S., King, A. W., Luo, Y., McNulty, S. G., Sun, G., Thornton, P. E., Wang, S., Williams, M. Baldocchi, D. D.

- and Cushman, R. M.: Oak forest carbon and water simulations: Model intercomparisons and evaluations against independent data, *Ecol. Monogr.*, 74, 443–489, 2004.
- Hsieh, W.: Nonlinear multivariate and time series analysis by neural network methods, *Rev. Geophys.*, 42, 1–25, 2004.
- Hsieh, W. and Hamilton, K.: Nonlinear singular spectrum analysis of the tropical stratospheric wind, *Q. J. R. Meteorol. Soc.*, 129, 2367–2382, 2001.
- Hsieh, W. and Wu, A.: Nonlinear multichannel singular spectrum analysis of the tropical Pacific climate variability using a neural network approach, *J. Geophys. Res.*, 107(C7), 3076, doi:10.1029/2001JC000957, 2001.
- Jevrejeva, S., Grinsted, A., Moore, J. C., and Holgate, S.: Nonlinear trends and multiyear cycles in sea level records, *J. Geophys. Res.*, 111, C09012, doi:10.1029/2005JC003229, 2006.
- Katul, G., Lai, C.-T., Schäfer, K., Vidakovic, B., J., A., Ellsworth, D., and Oren, R.: Multiscale analysis of vegetation surface fluxes: from seconds to years, *Adv. Water Resour.*, 24, 1119–1132, 2001.
- Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing data in geophysical data sets, *Nonlin. Processes Geophys.*, 13, 151–159, 2006, <http://www.nonlin-processes-geophys.net/13/151/2006/>.
- Kondrashov, D. and Ghil, M.: Reply to T. Schneider's comment on "Spatio-temporal filling of missing points in geophysical data sets", *Nonlin. Processes Geophys.*, 14, 3–4, 2007, <http://www.nonlin-processes-geophys.net/14/3/2007/>.
- Kramer, M. A.: Nonlinear principal component analysis using autoassociative neural networks, *AIChE Journal*, 37, 233–243, 1991.
- Lange, H. and Bernhardt, K.: Long-term components and regional synchronization of river runoffs, in: *Hydrology: Science and Practice for the 21st Century*, edited by Butler, A., British Hydrological Society, London, 165–170, 2004.
- Law, B. E., Falge, E., Gu, L., Baldocchi, D., Bakwin, P., Berbigier, P. Davis, K. J., Dolman, H., Falk, M., Fuentes, J., Goldstein, A. H., Granier, A., Grelle, A., Hollinger, D., Janssens, I., Jarvis, P., Jensen, N. O. Katul, G., Malhi, Y., Matteucci, G., Monson, R., Munger, J., Oechel, W., Olson, R., Pilegaard, K., Paw, U. K. T., Thorgeirsson, H., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Carbon dioxide and water vapor exchange of terrestrial vegetation in response to environment, *Agric. Forest Meteorol.*, 113, 97–120, 2002.
- Moffat, A., Papale, D., Reichstein, M., Barr, A. G., Braswell, B. H., Churkina, G., Desai, A. R., Falge, E., Gove, J. H., Heimann, M., Hollinger, D. Y. Hui, D., Jarvis, A. J., Kattge, J., Noormets, A., Richardson, A. D. Stauch, V. J.: Gap filling methods intercomparison, *Agric. Forest Meteorol.*, in press, 2007.
- Nakai, Y., Kitamura, K., and Abe, S.: Year-long carbon dioxide exchange above a broadleaf deciduous forest in Sapporo, Northern Japan, *Tellus B*, 55, 305–312, 2003.
- Paluš, M.: Detecting nonlinearity in multivariate time series, *Phys. Lett. A*, 213, 138–147, 1996.
- Paluš, M. and Novotná, D.: Enhanced Monte Carlo SSA and the case of climate oscillations, *Phys. Lett. A*, 248, 191–202, 1998.
- Paluš, M. and Novotná, D.: Quasi-biennial oscillations extracted from the monthly NAO index and temperature records are phase-synchronized, *Nonlin. Processes Geophys.*, 13, 287–296, 2006, <http://www.nonlin-processes-geophys.net/13/287/2006/>.
- Papale, D. and Valentini, R.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, *Global Change Biol.*, 9, 525–535, 2003.
- Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, 3, 571–583, 2006, <http://www.biogeosciences.net/3/571/2006/>.
- Peters, O., Hertlein, C., and Christensen, K.: A Complexity View of Rainfall, *Physical Review Letters*, 88, 018 701–1–018 701–4, 2002.
- Plaut, G. and Vautard, R.: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, 51, 210–236, 1994.
- Plaut, G. and Vautard, R.: interannual and interdecadal variability in 335 years of Central England temperatures, *Science*, 268, 710–713, 1995.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Valentini, R., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havrnkov, K., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Rambal, S., Rotenberg, E., Sanz, M., Seufert, G., Vaccari, F., Vesala, T., and Yakir, D.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biol.*, 11, 1–16, 2005.
- Reichstein, M., Ciais, P., Papale, D., Valentini, R., Running, S., Viovy, N., Cramer, W., Granier, A., Ogé, J., Allard, V., Aubinet, M., Bernhofer, C., Buchmann, N., Carrara, A., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Pilegaard, K., Rambal, S., Schaphoff, S., Seufert, G., Soussana, J.-F., Sanz, M.-J., Schulze, E. D., Vesala, T., and Heimann, M.: A combined eddy covariance, remote sensing and modeling view on the 2003 European summer heatwave, *Global Change Biol.*, 13, 634–651, 2007.
- Richardson, A. D., Braswell, B. H., Hollinger, D. Y., Burman, P., Davidson, E. A., Evans, R. S., Flanagan, L. B., Munger, J. W., Savage, K., Urbanski, S. P., and Wofsy, S. C.: Comparing simple respiration models for eddy flux and dynamic chamber data, *Agric. Forest Meteorol.*, 141, 219–234, 2006.
- Richardson, A. D., Hollinger, D. Y., Aber, J. D., Ollinger, S. V., and Braswell, B. H.: Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biol.*, 13, 788–803, 2007.
- Richardson, A. D. and Hollinger, D. Y.: A method to estimate the additional uncertainty in gap-filled NEE resulting from long gaps in the CO<sub>2</sub> flux record, *Agric. Forest Meteorol.*, in press, doi:10.1016/j.agrformet.2007.06.004, 2007a.
- Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO<sub>2</sub> flux measurement uncertainty inferred from model residuals, *Agric. Forest Meteorol.*, in press, 2007b.
- Saigusa, N., Yamamoto, S., Murayama, S., and Kondo, H.: Inter-annual variability of carbon budget components in an AsiaFlux forest site estimated by long-term flux measurements, *Agric. For-*

- est Meteorol., 134, 4–16, 2005.
- Seneviratne, S. I., Koster, R. D., Guo, Z., Dirmeyer, P. A., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Mocko, D., Oleson, K. W., and Verseghy, D.: Soil Moisture Memory in AGCM Simulations: Analysis of Global Land-Atmosphere Coupling Experiment (GLACE) Data. *J. Hydrometeorol.*, 7, 1090–1112, 2006.
- Schneider, T.: Comment on “Spatio-temporal filling of missing points in geophysical data sets” by D. Kondrashov and M. Ghil. *Nonlin. Processes Geophys.*, 13, 151–159, 2006, *Nonlin. Processes Geophys.*, 14, 1–2, 2007, <http://www.nonlin-processes-geophys.net/14/1/2007/>.
- Shun, T. and Duffy, C.: Low-frequency oscillations in precipitation, temperature, and runoff on a west facing mountain front: A hydrogeologic interpretation, *Water Resour. Res.*, 35, 191–201, 1999.
- Siqueira, M. B., G. G. Katul, D. A. Sampson, Stoy, P. C., Juang, J.-Y., McCarthy, H. R. and Oren, R.: Multi-scale model inter-comparisons of CO<sub>2</sub> and H<sub>2</sub>O exchange rates in a maturing southeastern U.S. pine forest, *Global Change Biol.*, 12, 1189–1207, 2006.
- Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., McCarthy, H. R., Kim, H.-S., Oishi, A. C., and Oren, R.: Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis, *Tree Physiol.*, 25, 887–902, 2005.
- Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J.-Y., Novick, K. A., Joshua, M., Uebelherr, and Oren, R.: An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration, *Agric. Forest Meteorol.*, 141, 2–18, 2006.
- Takens, F.: Detecting strange attractors in turbulence, in: *Lecture Notes in Dynamical Systems and Turbulence* edited by Rand, D. A., and Young, L.-S., *Lecture Notes in Mathematics*, 898, 366–381, Springer, 1981.
- Tenenbaum, J. B., de Silva, V., and Langford, J. C.: A global geometric framework for nonlinear dimensionality reduction, *Science*, 290, 2319–2323, 2000.
- Tessier, Y., Lovejoy, S., Hubert, P., Schertzer, D., and Pecknold, S.: Multifractal analysis and modeling of rainfall and river flows and scaling, causal transfer functions, *J. Geophys. Res.*, 101(D21), 26 427–26 440, 1996.
- Torrence, C., and Compo, G. P.: *A Practical Guide to Wavelet Analysis*, *B. Am. Meteorol. Soc.* 61–79, 1998.
- Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395–424, 1989.
- Vautard R., Yiou P., and Ghil M.: Singular spectrum analysis: a toolkit for short noisy chaotic signals, *Physica D*, 58, 95–126, 1992.
- von Storch, H. and Zwiers, F.: *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 1999.
- Wilson, K. and Baldocchi, D.: Seasonal and interannual variability of energy fluxes temperate deciduous forest in North America, *Agric. Forest Meteorol.*, 100, 1–18, 2000.
- Yiou P., Sornette, D., and Ghil M.: Data-adaptive wavelets and multi-scale singular-spectrum analysis, *Physica D*, 142, 254–290, 2000.