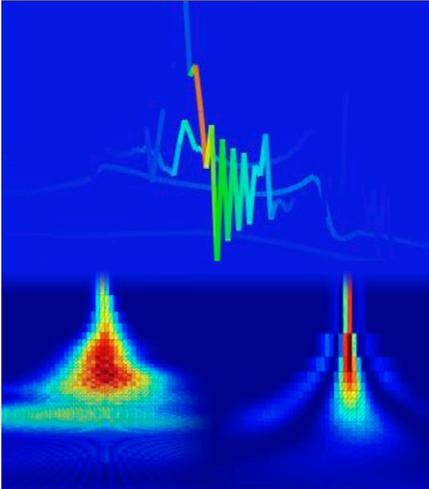


J. Rudi
 R. Pabel
 G. Jager
 R. Koch
 A. Kunoth*
 H. Bogen



For analyzing hydrological time series data we employ the technique recently developed called the Hilbert–Huang Transform (HHT). We provide a comparison of the Hilbert spectra with Fourier and wavelet spectra to demonstrate a better localization of the energy components which also exhibit strong seasonal components.

J. Rudi, R. Pabel, G. Jager, R. Koch, and A. Kunoth, Institut für Mathematik, Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany; and H. Bogen, Forschungszentrum Jülich GmbH, ICG-4 52425 Jülich, Germany. *Corresponding author (Kunoth@math.uni-paderborn.de).

Vadose Zone J. 9:925–942
 doi:10.2136/vzj2009.0163
 Received 15 Nov. 2009.
 Published online 7 Sept. 2010.

© Soil Science Society of America
 5585 Guilford Rd., Madison, WI 53711 USA.
 All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher.

Multiscale Analysis of Hydrologic Time Series Data using the Hilbert–Huang Transform

For the analysis of time series data from hydrology, we used a recently developed technique that is by now widely known as the Hilbert–Huang transform (HHT). Specifically, it is designed for nonlinear and nonstationary data. In contrast to data analysis techniques using the short-time, windowed Fourier transform or the continuous wavelet transform, the new technique is empirically adapted to the data in the following sense. First, an additive decomposition, called *empirical mode decomposition* (EMD), of the data into certain multiscale components is computed. Second, to each of these components, the Hilbert transform is applied. The resulting Hilbert spectrum of the modes provides a localized time–frequency spectrum and instantaneous (time-dependent) frequencies. In this study, we applied the HHT to hydrological time series data from the Upper Rur Catchment Area, mostly German territory, taken during a period of 20 yr. Our first observation was that a coarse approximation of the data can be derived by truncating the EMD representation. This can be used to better model patterns like seasonal structures. Moreover, the corresponding time–frequency energy spectrum applied to the complete EMD revealed seasonal events in a particular apparent way together with their energy. We compared the Hilbert spectra with Fourier spectrograms and wavelet spectra to demonstrate a better localization of the energy components, which also exhibit strong seasonal components. The Hilbert energy spectrum of the three measurement stations appear to be very similar, indicating little local variability in drainage.

Abbreviations: EMD, empirical mode decomposition; HHT, Hilbert–Huang transform; IMF, intrinsic mode function.

Given empirical data, the detection and parameterization of multiscale patterns and shapes in the measurements is an important task. Specifically, to study the effect of patterns on water and solute fluxes, temporal and spatial data have to be analyzed at various stages so that their parameterization can eventually be used in simulation flux models.

This study was part of a SFB/TR32 project (Transregional Collaborative Research Centre 32, www.tr32.de; verified 3 June 2010) for which the overall objective is to improve our knowledge about the mechanisms leading to spatial and temporal patterns in energy and matter fluxes of the soil–vegetation–atmosphere system. Part of the objectives is the determination, description, and analysis of patterns derived from different sources. For instance, the large spatial and temporal variability of soil moisture patterns is determined by factors like atmospheric forcing, topography, soil properties, and vegetation, which interact in a complex, nonlinear way (see, e.g., Grayson and Blöschl, 2001; Western et al., 2004). Thus, a very large number of continuous soil moisture measurements are necessary to adequately capture this variability. In the framework of the TR32, a dense soil moisture sensor network for monitoring soil water content changes at high spatial and temporal scales has been set up (see Bogen et al., 2007, 2009, 2010; Rosenbaum et al., 2010). Due to the fact that long-term soil moisture data from the SFB/TR32 study area, the catchment of the Rur River, were not available, we focus here on long-term runoff discharge measurements taken from three rivers with low management influence located in the southern part of the Rur catchment. Since runoff discharge and soil moisture time series data exhibit similar temporal patterns, this study can be considered as preliminary for the analysis of the soil moisture sensor network data.

For the detection of structures and patterns at different scales, the method of choice is to transform and thereby decompose the measurement data into multiscale components. Classical and widely established methods are the short-time Fourier transform or, more recently, continuous wavelet transforms. The potential of wavelets as an analysis and approximation tool has been

demonstrated in different areas (see, e.g., Castaño and Kunothe, 2006; Castaño Diez et al., 2009; DeVore and Kunothe, 2009).

An example of the analysis of hydrometeorologic data by means of wavelets was given in Bachner et al. (2006). In that study, the goal was the identification and filtering of dominant time scales in statistical indices of daily time series. Wavelet analysis provided a temporally varying power spectrum and does not require stationarity of the data. Kang and Lin (2007) performed wavelet analysis of hydrologic and water quality signals in an agricultural watershed using a weighted wavelet-Z transform. Schaeffli and Zehe (2009) proposed a method for rainfall-runoff model calibration and performance analysis by fitting a wavelet power spectrum estimated from measurement data. A review of spectral and wavelet methods together with detailed procedures for providing spatial scaling analyses of physical soil properties was provided in Si (2008).

Nevertheless, methods based on the short-time Fourier transform or the continuous-wavelet transform have the following disadvantages for time series data. First, they require data on uniform grids. Second, as demonstrated in Huang et al. (1998), if the data are nonlinear and nonstationary, these transforms often do not lead to physically meaningful results. The reasons are fundamental construction ingredients: they allow only certain linear combinations of predefined bases and for each component the same frequency is used across the whole time domain. Although wavelets have the advantage over the Fourier basis in that they allow a localized identification of the significant frequency components, the motivation for the new approach developed in Huang et al. (1998) was to find a better decomposition into components that are adapted to the specific data set.

In this study, we followed this approach to analyze nonlinear and nonstationary temporal data by applying the Hilbert-Huang transform (HHT) developed in Huang et al. (1998). In its original version, it was designed to handle nonlinear and nonstationary time series in one dimension. The principle of the HHT is as follows: First, the time series is iteratively decomposed into empirical adaptive nonlinear modes (intrinsic mode functions or IMFs) that exhibit nonlinear shapes and patterns and are physically meaningful (in a sense to be made precise below). This process is called empirical mode decomposition (EMD). Second, the Hilbert spectral analysis of the IMFs then provides a localized time-frequency spectrum and a possible extraction of instantaneous frequencies.

By now this method has been used for many physical long- and short-term univariate data (see, e.g., Huang and Shen, 2005, and references therein). A very recent study where Hilbert spectral analysis was applied to hydrologic data is Huang et al. (2009), where daily river flow fluctuations were analyzed.

The original scheme has been modified and refined by several researchers during the past decade (see, e.g., Chen et al., 2006; Deléclle et al., 2005; Flandrin and Gonçalvès, 2004; Flandrin et al., 2004). First,

extensions to derive an EMD of two-dimensional data was developed by Damerval et al. (2005) and Xu et al. (2006), and applications for adaptive image compression by Linderhed (2004). Koch (2008) compared several methods for computing the EMD for two-dimensional data with respect to their numerical performance. This was extended by Jager et al. (2010) with a focus on numerical efficiency, which is mandatory, in particular, for multisensorial data in several space dimensions. Moreover, the second step, the Hilbert transform, also needed to be generalized to extend the analytical signals to arbitrary dimensions. Based on ideas using Clifford algebra and Riesz transforms from Felsberg and Sommer (2000, 2001), a corresponding fast numerical scheme and results for two-dimensional data was developed by Koch (2008) and Pabel et al. (2010).

• The Hilbert-Huang Transform Empirical Mode Decompositions

The advantage of the HHT developed by Huang et al. (1998) over the Fourier decomposition or a wavelet decomposition is that it can be used to analyze nonlinear and nonstationary data across irregular time grids. The HHT works by iteratively decomposing the time series into a finite number of IMFs through an EMD process. After the data-driven additive decomposition is obtained, a second step, detailed below, is applied to each IMF component in the Hilbert transform, yielding a time-frequency distribution of the energy, the so-called *Hilbert spectrum*.

First we describe the EMD process as originally designed by Huang et al. (1998), with a synthetic example from Rilling et al. (2003). A set of real-valued time series data $\{(t_\ell, z_\ell)\}_{\ell \in \mathbb{Z}}$ is called a nonlinear and nonstationary data set if there exists a number $m \in \mathbb{Z}$ such that the common probability distribution of $z_\ell, \dots, z_{\ell+m}$ depends on the time index ℓ . All sorts of measurement data from physical processes usually satisfy this condition. It is convenient to describe the methods by considering, instead of discrete time series data, a continuous function $s: \mathbb{R} \rightarrow \mathbb{R}$. Note that a continuous function can always be generated from discrete data by (linear continuous) interpolation or by a least-squares approximation of the data.

The goal of the method is to decompose the function into finitely many components, which later allows the definition of instantaneous (time-dependent) amplitudes and frequencies. The specific feature of the decomposition considered here is that these components are adaptively derived from the input data. These data-driven components are determined in such a way that they satisfy the following properties. We say that a function $\text{imf}: \mathbb{R} \rightarrow \mathbb{R}$ is an IMF if (i) the number of local extrema and zero points of the IMF differ mostly by one; and (ii) at any point, the mean value of the cubic spline that interpolates all local maxima and of the cubic spline that interpolates all local minima is zero. A cubic spline is a function consisting of piecewise polynomials of the third degree joined together such that their second derivative is still a continuous function. Thus, an IMF represents a basic oscillation that is symmetrically localized around the ordinate axis. The

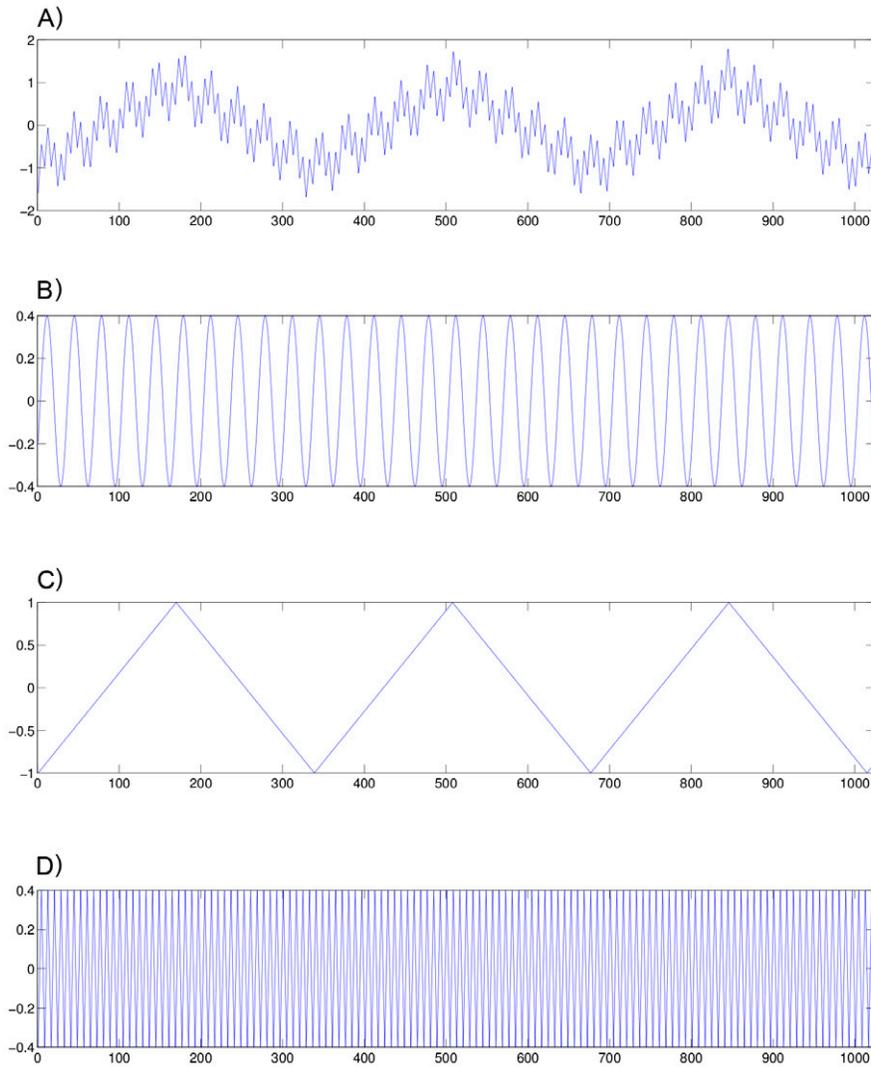


Fig. 1. (A) Synthetic function s , additively composed from (B) a sine wave and (C and D) two piecewise linear continuous functions with different periodicities.

collection of IMFs may be viewed as a data-adapted basis that is not known a priori since they additively decompose the data (viewed as a continuous function) in a unique way once the iteration parameters are fixed. These functions are the natural generalizations of Fourier components, with the difference that they have a variable amplitude and a variable frequency as a function of time, called the *instantaneous amplitude* and *instantaneous frequency*. In contrast, the Fourier basis and the wavelet basis are known a priori.

For illustration purposes, we describe the subsequent EMD process using the synthetic function displayed in Fig. 1. This function s is additively composed from a sine wave and two piecewise linear continuous functions with different periodicities. It will be decomposed into a finite number of IMFs and a monotone residual by the following iterative process.

The decomposition is based on two iterations: one inner loop, called *sifting*, which has generated a single IMF after its completion, and an

outer loop, which consists of the decomposition into the different IMFs. For computing a single IMF, an upper and lower envelope s_+, s_- of the local maxima or minima is computed, consisting of cubic splines, which contain in their convex hull the original signal s (see Fig. 2A).

For the difference between the original signal $s(t)$ and the mean $m_{1,1}(t) := 1/2[s_+(t) + s_-(t)]$, which is shown in Fig. 2B, that is, $h_{1,1}(t) := s(t) - m_{1,1}(t)$, one again computes the upper and lower envelopes, subtracts the mean $m_{1,2}(t)$ and repeats the process with the new signal $h_{1,2}(t) := h_{1,1}(t) - m_{1,2}(t)$ until a standard mean deviation criterion is met; for details and a discussion of appropriate termination criteria, see Huang et al. (1998) and Koch (2008). The result is called $\text{imf}_1(t)$. The next inner sifting process then starts with the residual $r_1(t) := s(t) - \text{imf}_1(t)$, yielding the second $\text{imf}_2(t)$. After termination of the last outer iteration, combining all components finally results in an additive decomposition:

$$s(t) = \sum_{j=1}^n \text{imf}_j(t) + r_n(t) \quad [1]$$

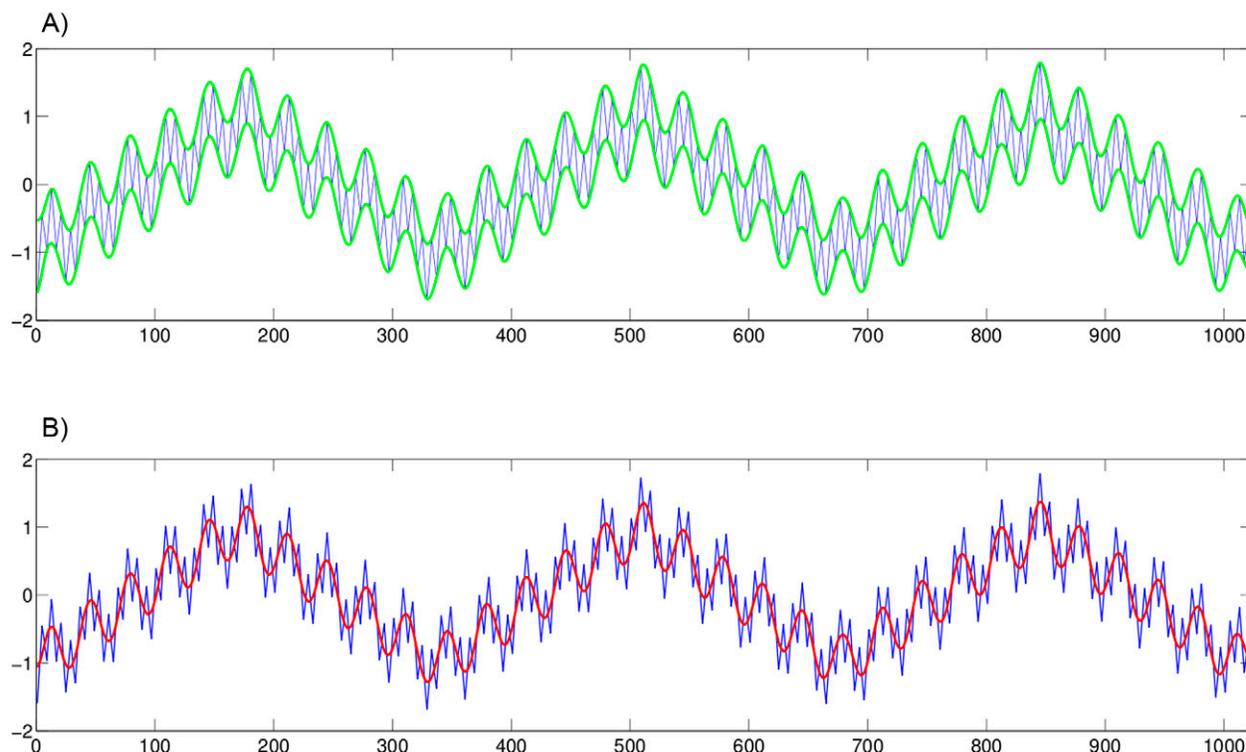


Fig. 2. Synthetic function s (blue) (A) between its upper envelope s_+ and its lower envelope s_- (both green), and (B) with its mean function $m_{1,1}$ (red).

where r_n is a constant or monotone residual. By this we mean that r_n is the remainder of the decomposition, which is a monotone function and, therefore, has at most one root (i.e., at most one zero). This residual may be viewed as a trend in the data. For the synthetic data from Fig. 1, the result of the iterative process, the EMD, is shown in Fig. 3.

Although synthetic, this example is very illustrative: it can be seen that the first three components are clearly recovered from the original data; according to the construction, the function imf_1 contains the fastest oscillations. The appearance of an additional function imf_4 (compared with the original signal s) is caused by interpolation errors and boundary effects. Note, however, the scale: the modulus of the amplitude is bounded by 0.04, which can be considered negligible in this example. Also, the monotone residual is on this order.

Note that the different IMFs are approximations of linear combinations of cubic B splines—they are approximations because the process is iterative. This can be seen from the results for the synthetic example by comparing the plots in Fig. 1 with the ones in Fig. 3. Each IMF is approximately a linear combination of cubic B splines for which, however, the expansion coefficients are determined by the data. This means that different data sets, even if sampled on the same uniform grid, yield different IMFs. Also, the frequencies may differ within an IMF. In that sense, the different IMFs are not the same as simply a linear combination of cubic B splines all stemming from the same resolution grid.

Finally, note that we did not claim that to not make use of a basis at all. In fact, for the iteration procedure, it is essential to have one. The main difference between the EMD decomposition Eq. [1] (or Eq. [11] below) and, e.g., the Fourier representation Eq. [12] below is that the Fourier representation has a fixed basis $\exp(i\omega t)$, and its expansion coefficients a_j can only attain a certain constant value for each j , whether this is needed locally or not. For a wavelet expansion, although the basis functions have local support, also the expansion coefficients can only take on a constant value. In contrast, the EMD decomposition or its HHT version (Eq. [11]) allows a time-dependent coefficient that is constructively adapted to the data.

◆ Hilbert Spectral Analysis

Once the data are decomposed into an EMD according to Eq. [1], the Hilbert transform can be applied to each IMF component and instantaneous frequencies can be computed by means of these. The main idea is to generate from a real-valued signal s a complex-valued extension (by means of the Hilbert transform), called the analytical representation of the signal or, shortly, the *analytical signal*. This technique has a long tradition in signal processing (see, e.g., Cohen, 1994; Flandrin, 1999; and, specifically, Hahn, 1995). The main idea is the observation that the negative frequency components of the Fourier transform of a real-valued function do not have to be taken into account, due to the Hermitian symmetry of such a spectrum (see Eq. [7] below). Then these negative frequency components can be discarded without losing information under the condition that a

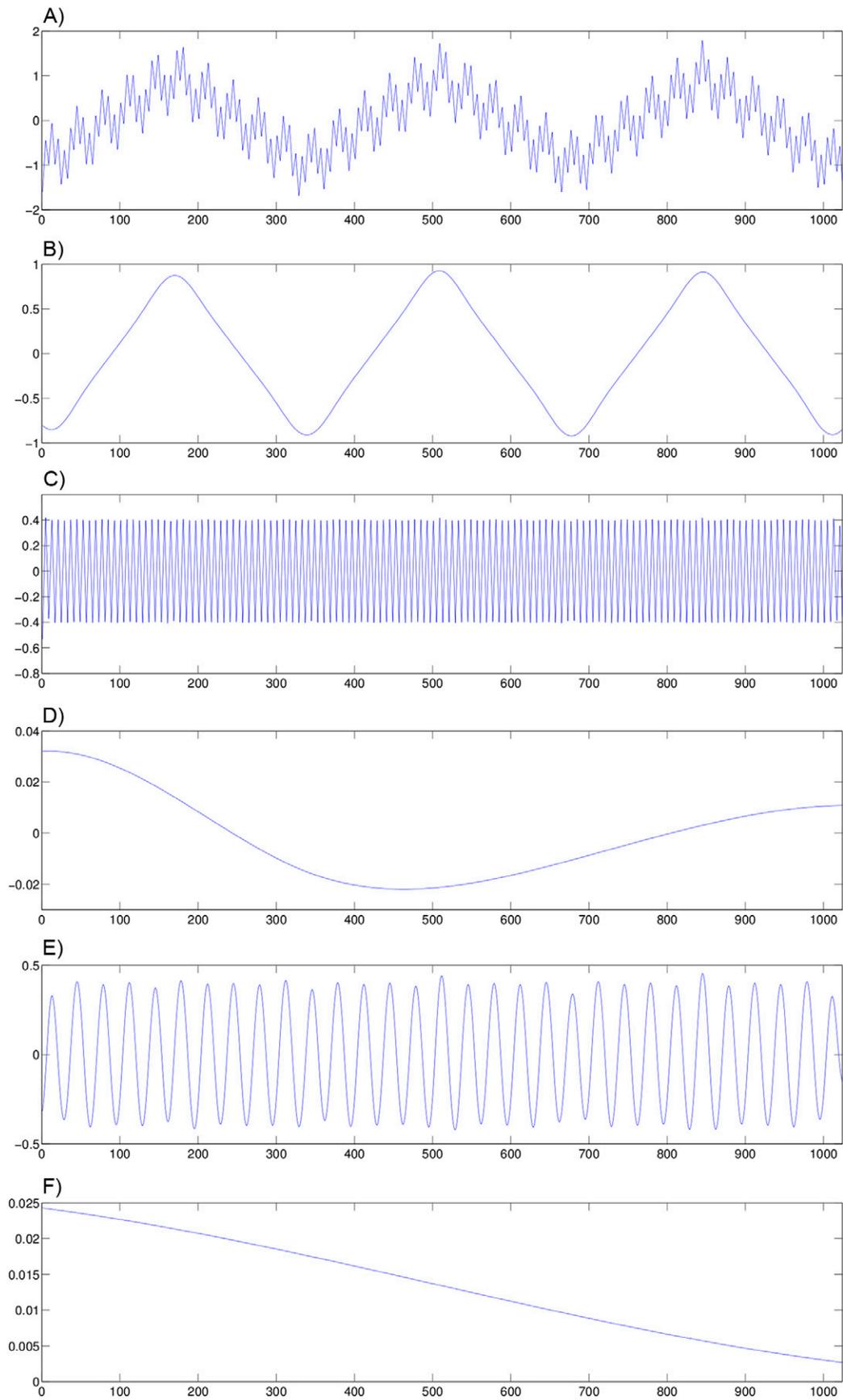


Fig. 3. Empirical mode decomposition (EMD) of (A) synthetic function s into intrinsic mode functions (B) imf_1 , (C) imf_2 , (D) imf_3 , (E) imf_4 , and (F) a monotone residual r_4 . Note the different ordinate scaling for the last two components in comparison with the other ones.

complex-valued function is now used for computation instead. For the analytical signal then, time-dependent (instantaneous) amplitudes and frequencies can be defined. In many applications, this is considered “physically meaningful” (see, e.g., Huang and Shen, 2005).

The Hilbert transform $H[s]$ of a function $s: \mathbf{R} \rightarrow \mathbf{R}$ is defined as the integral transform:

$$H[s](t) := \frac{1}{\pi} \text{PV} \int_{-\infty}^{\infty} \frac{s(u)}{t-u} du \quad [2]$$

$$:= \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_{-\infty}^{t-\epsilon} \frac{s(u)}{t-u} du + \lim_{\epsilon \rightarrow 0} \frac{1}{\pi} \int_{t+\epsilon}^{\infty} \frac{s(u)}{t-u} du, \quad t \in \mathbf{R}$$

where PV is the principal value (see, e.g., Titchmarsh, 1950). We will exploit certain properties of the Hilbert transform in connection with the Fourier transform $F[s]$ of the signal $s: \mathbf{R} \rightarrow \mathbf{R}$, defined by

$$F[s](\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} s(t) \exp(-i\omega t) dt, \quad \omega \in \mathbf{R} \quad [3]$$

where F is the Fourier spectrum of s , often abbreviated $\hat{s} := F[s]$. Recall that the Fourier transform is a technique to represent a signal in the frequency domain and describes which portion of a particular frequency is contained in the signal; however, these frequencies ω are constant with time. In contrast, using the Hilbert transform will allow a frequency to be defined depending on time. This will be achieved by first constructing from the real-valued given data an analytical signal that is by definition complex valued, using the Hilbert transform. For this analytical signal, then, an instantaneous (time-dependent) frequency can be defined.

To do so, we will need a few more technical facts. Together with the property $F[1/s](\omega) = -[i\sqrt{(2\pi)}]/2\text{sign}(\omega)$ [where $\text{sign}(\omega) := 1$ for $\omega \geq 0$ and $\text{sign}(\omega) := -1$ for $\omega < 0$], one can show

$$F[H[s]](\omega) = \frac{\sqrt{2\pi}}{\pi} F\left[\frac{1}{s}\right](\omega) F[s](\omega) = (-i)\text{sign}(\omega)\hat{s}(\omega) \quad [4]$$

One can also show that a function ν has the Fourier coefficients $\hat{\nu}(\omega) = (-i)\text{sign}(\omega)\hat{\omega}(\omega)$ if and only if $\nu(s) = H[\omega](s)$.

For the practical computation of the Hilbert transform, it can be used that for a (real-valued) signal $s: \mathbf{R} \rightarrow \mathbf{R}$,

$$H[s](t) = \sqrt{\frac{2}{\pi}} \Im \left[\int_0^{\infty} \hat{s}(\omega) \exp(i\omega t) d\omega \right] \quad [5]$$

where $\Im(z)$ denotes the imaginary part of a complex number $z \in \mathbf{C}$. For instance, the Hilbert transform of $\tilde{s}(t) := \alpha + \sin(ct)$, $\alpha, c \in \mathbf{R}$ is just $H[\tilde{s}](t) = \Im[-i\exp(ict)] = \cos(ct)$.

The “complexification” of given (real-valued) data $s: \mathbf{R} \rightarrow \mathbf{R}$ can be achieved as follows. Denote by $\nu(t) := H[s](t)$ the Hilbert transform of s and define the analytical signal as

$$s_A(t) := s(t) + i\nu(t), \quad t \in \mathbf{R} \quad [6]$$

Note that the real part \Re of s_A recovers the original signal, i.e., $\Re[s_A(t)] = s(t)$. Equation [6] ensures that the spectrum of the complex-valued signal s_A is zero for negative frequencies ω :

$$\hat{s}_A(\omega) = \begin{cases} 2\hat{s}(\omega), & \omega \geq 0, \\ 0, & \omega < 0 \end{cases} \quad [7]$$

The analytical signal s_A can also be represented as

$$s_A(t) = a(t) \exp[i\varphi(t)] \quad [8]$$

with amplitude $a(t) := \sqrt{[s^2(t) + \nu^2(t)]}$ and phase $\varphi(t) := \arctan[\nu(t)/s(t)]$. We define now the instantaneous frequency $\omega = \omega(t): \mathbf{R} \rightarrow \mathbf{R}$ of s as

$$\omega(t) := \frac{d}{dt} \varphi(t) \quad [9]$$

Specifically, this definition is consistent with the definition of a mean frequency, i.e., it satisfies

$$\int_{-\infty}^{\infty} \omega |\hat{s}_A(\omega)|^2 d\omega = \int_{-\infty}^{\infty} \left[\frac{d}{dt} \varphi(t) \right] |s_A(t)|^2 dt \quad [10]$$

After these preparations, we are in the position to compute instantaneous frequencies for the EMD of the original signal. In fact, recalling the additive decomposition Eq. [1] and applying the Hilbert transform to each of the IMF components, we obtain

$$s(t) = \sum_{j=1}^n \text{imf}_j(t) + r_n(t)$$

$$= \sum_{j=1}^n \Re \left\{ \text{imf}_j(t) + iH[\text{imf}_j](t) \right\} + r_n(t)$$

$$= \Re \left\{ \sum_{j=1}^n \left\{ \text{imf}_j(t) + iH[\text{imf}_j](t) \right\} \right\} + r_n(t) \quad [11]$$

$$= \Re \left\{ \sum_{j=1}^n a_j(t) \exp \left[i \int_{\mathbf{R}} \omega_j(t) dt \right] \right\} + r_n(t)$$

Note here that for the residual r_n , its instantaneous frequency is irrelevant because the residual is either monotone or a constant that only exhibits a trend. Thus, only a possible long-term trend without seasonal influence can be obtained from it.

As intended, Eq. [11] furnishes for each component index j an amplitude $a_j(t)$ as well as a frequency $\omega_j(t)$ depending on time. Comparing this with the Fourier representation of the signal, truncated after the n th term,

$$s(t) = \Re \left\{ \sum_{j=1}^n a_j \exp(i\omega_j t) \right\} + \tilde{r}_n(t) \quad [12]$$

it can be observed that the components of the latter have only constant amplitude and frequency. In this sense, the EMD provides a generalized Fourier representation that is particularly appropriate for nonstationary and nonlinear data.

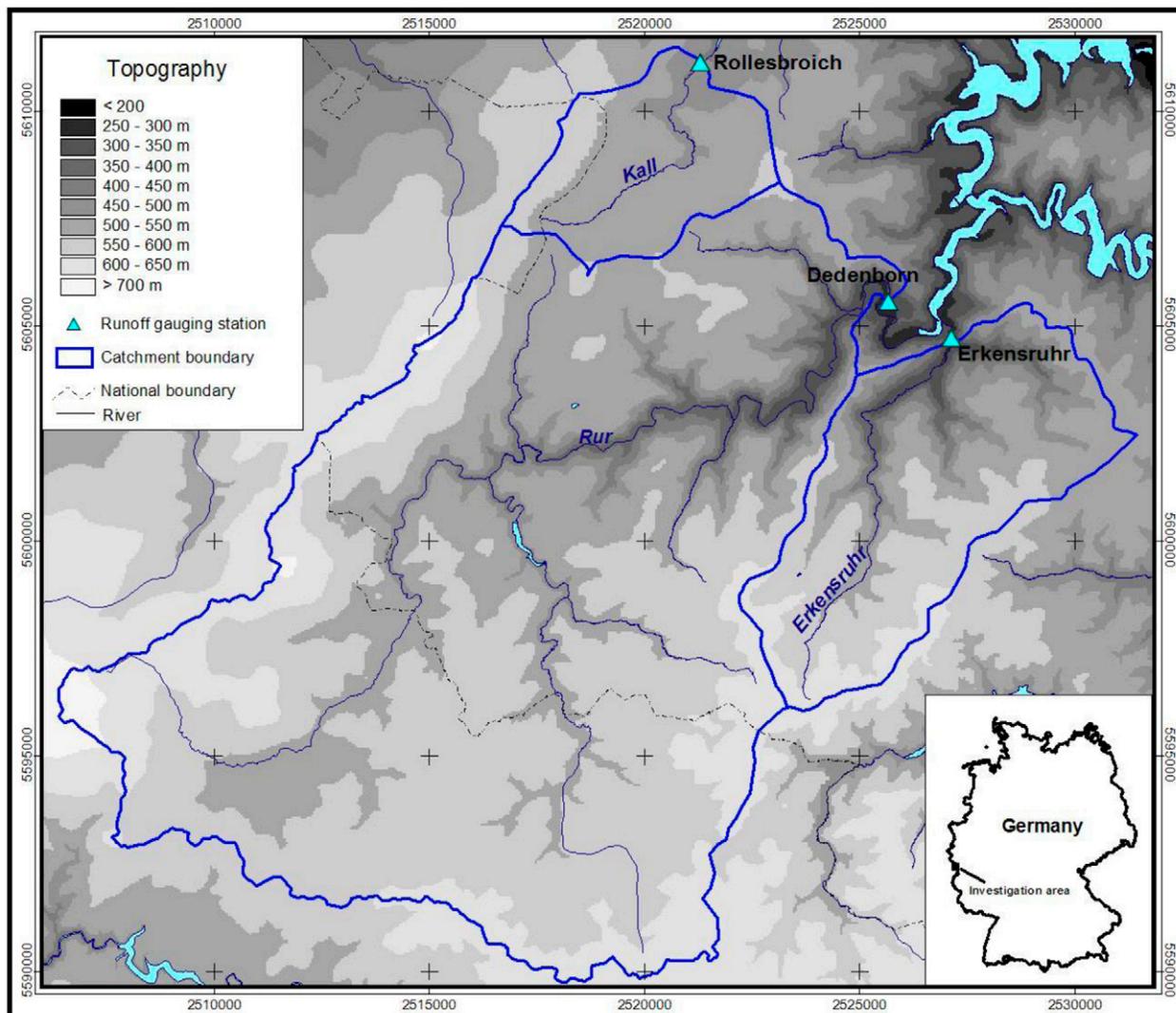


Fig. 4. The location of the runoff gauging stations Rollesbroich, Dedenborn, and Erkersruhr and their associated catchment areas.

We now interpret the amplitude depending on the time and the frequency and denote the time–frequency decomposition of the amplitude as the Hilbert amplitude spectrum $H(\omega, t)$. Formally, this is defined as follows. Let the signal s be represented in the form of Eq. [11]. Then the Hilbert amplitude spectrum is defined as

$$H(\omega, t) = H[\omega(t), t] \quad [13]$$

$$:= \begin{cases} a_1(t) & \text{on the curve } \{[\omega_1(t), t] : t \in \mathbf{R}\} \\ \vdots \\ a_n(t) & \text{on the curve } \{[\omega_n(t), t] : t \in \mathbf{R}\} \\ 0 & \text{elsewhere} \end{cases}$$

Note, however, that one can also define the Hilbert energy spectrum by taking squares of the amplitudes. In fact, “if amplitude squared is more desirable commonly to represent energy density, then the squared values of amplitude can be substituted to produce the Hilbert energy spectrum just as well” (Huang et al., 1998, p. 928, fourth paragraph).

In the subsequent computations below, we have chosen the Hilbert energy spectrum (taking squares of the amplitudes in Eq. [13]) because the results can be better compared with the Fourier spectrogram and the wavelet spectrum.

Application to Hydrologic Time Series Data

The concept described above is now applied to different hydrologic data sets described in Bogena et al. (2005a,b).

Basic Characteristics of the Investigation Area

The catchment of the Rur River was chosen as the regional investigation area for the SFB/TR32 project. The Rur catchment covers a total area of 2354 km² and is situated in western Germany. For this study, we selected data from three runoff gauging stations (Dedenborn, Erkersruhr, and Rollesbroich) located in the southern part of the Rur catchment (see Fig. 4). The runoff discharge was measured during a

Table 1. Basic characteristics of the catchment areas.

Characteristic	Upper Rur River	Kall River	Erkensruhr River
Gauging station	Dedenborn	Rollesbroich	Erkensruhr
Area, km ²	199.7	19.03	41.7
Mean elevation, m	553	532	526
Mean slope, °	36.4	17.3	32.1
Land use, % (2001)	urban, 2; arable land, 4; grassland, 35; raised bogs, 29; forest, 30	urban, 3; arable land, 13; grassland, 45; forest, 39	urban, <1; arable land, 10; grassland, 16; forest, 74
Geology, %	Palaeozoic, 89; Pleistocene, 1; Holocene, 10	Palaeozoic, 94; Holocene, 6	Palaeozoic, 96; Pleistocene, 1; Holocene, 3
Annual precipitation, mm/yr (1979–1999)	mean, 1042; min., 908; max., 1413	mean, 1200; min., 1133; max., 1266	mean, 1081; min., 932; max., 1219
Annual runoff, mm/yr	793	733	633
Daily sampling interval	1961–2000	1982–2000	1961–2000

period of 6940 d (roughly 20 yr between 1981 and 2001). The associated catchments of these stations are different in size (20–200 km²) and exhibit varying catchment characteristics (see Table 1).

Due to the marine air flowing in predominantly from southwest to northeast, a significant precipitation shadowing effect can be observed. The precipitation levels in the peak regions of the Rhenish Massif (“Hohes Venn”) are higher than those on the eastern sides (lee regions). Therefore, the annual precipitation ranges between about 900 mm/yr in the northeastern part of the investigation area and about 1400 mm/yr in the southwest.

The investigation area belongs to the Central European low mountain ranges and is dominated by Palaeozoic solid rocks of the

Rhenish Massif formed in the course of Variscan orogenesis, which occupy the largest area fraction (89–96%). Holocene floodplain deposits and raised bogs occur especially within the Upper Rur River catchment but are of less importance (3–10%). The Upper Rur River comprises also a significantly proportion of raised bogs (29%). The catchment areas of the Upper Rur and Kall rivers are predominated by pastures, whereas the Erkensruhr River is characterized by forest, leading to significant lower mean annual runoff compared with the other catchments (100 and 190 mm/yr, respectively).

Figure 5 shows the runoff discharge of all catchments for the period of 1982 to 1999 on a daily basis. The general runoff characteristic is very similar for all catchments. The highest peak flows have been recorded during winter, whereas during summer, low flow conditions

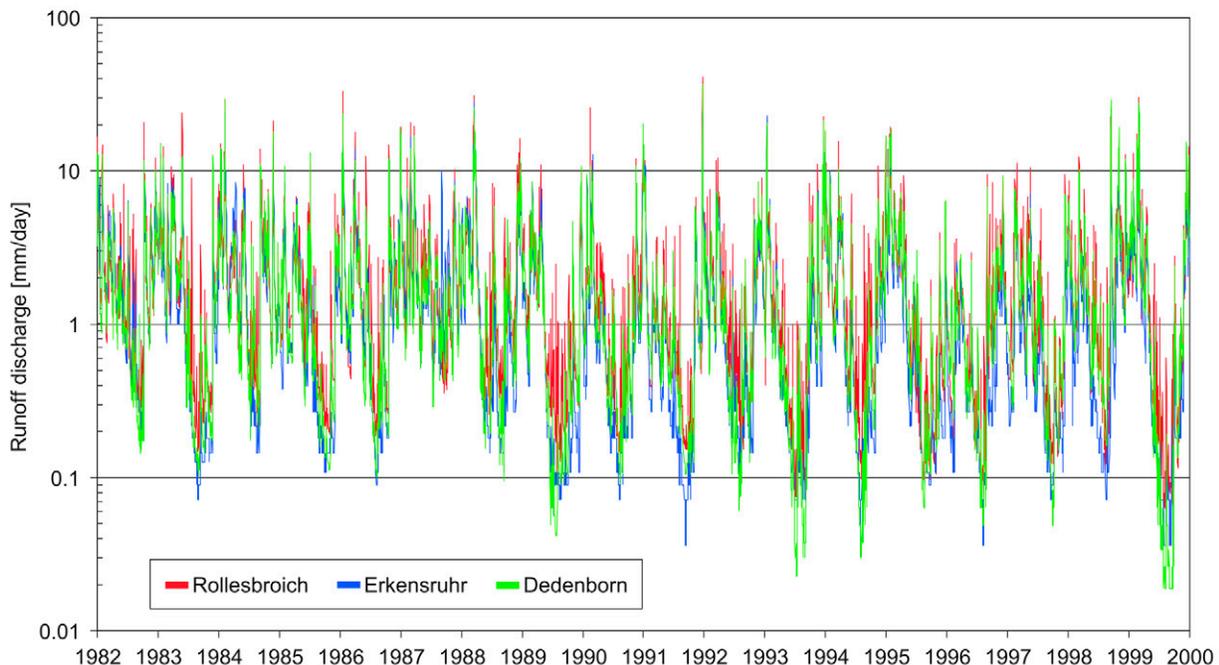


Fig. 5. Runoff discharge data from 1982 to 1999 from the Rollesbroich, Dedenborn, and Erkensruhr gauging stations.

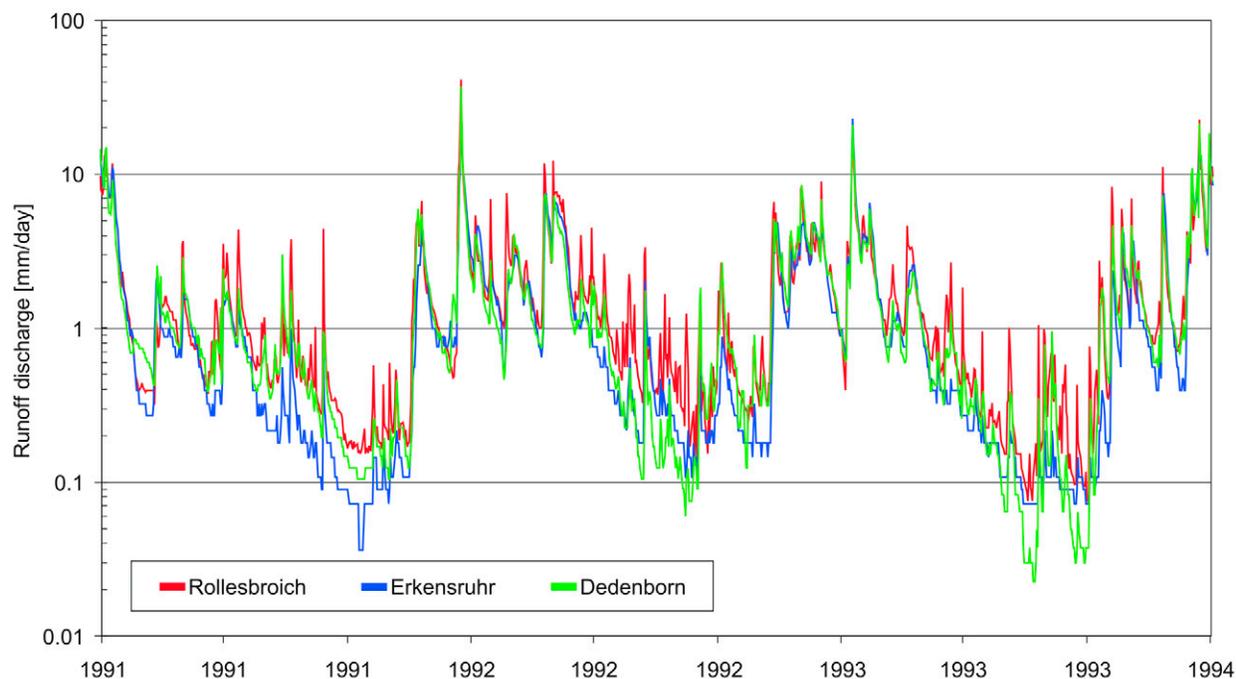


Fig. 6. Runoff discharge data from 1991 to 1993 from the Rollesbroich, Dedenborn, and Erkensruhr gauging stations.

are common. The maximum runoff was recorded on 22 Dec. 1991 (Rollesbroich, 41.34 mm/d), whereas one of the lowest runoffs from 1990 to 1999 was recorded on 11 July 1993 (0.0023 mm/d). Because it contains the highest and one of the lowest runoffs, we selected the period from 1991 to 1993 to present the runoff data in more detail (Fig. 6). From Fig. 6 it becomes apparent that the runoff from the Kall River (Rollesbroich gauging station) shows more peaks as a result of the smaller catchment area. The runoff from the Upper Rur River (Dedenborn gauging station) is to some extent damped due the presence of a small water reservoir (see Fig. 4).

The current study is meant as a preliminary study for the data analysis of a dense soil moisture sensor network for monitoring soil water content changes at high spatial and temporal scales; for a description of the currently used soil moisture network and measurement devices, see Bogaen et al. (2007, 2009, 2010) and Rosenbaum et al. (2010).

Computation of the Empirical Mode Decomposition

For the three data sets and different time ranges, we have computed the corresponding EMDs. The result for the Dedenborn data set in the time range 1990 to 1999 can be seen below; all the others can be found in Rudi (2010).

Figure 7 presents the first three IMF amplitudes for all catchments together with the original runoff discharge data for the 1991 to 1993 time range. (Note that the IMF amplitudes $a_1(t)$, ..., $a_3(t)$ differ from the IMFs $\text{imf}_1(t)$, ..., $\text{imf}_3(t)$ in Eq. [1]). We can make the following observations. The runoff peaks produce at most times high IMF amplitude values, especially for imf_1 , which is the

portion of the signal corresponding to the highest frequencies. The extreme event of 22 Dec. 1991 produced very distinct amplitude peaks for imf_1 , imf_2 , and imf_3 and is clearly visible in all three data sets. Low flow periods mostly are accompanied by low amplitudes for imf_1 , imf_2 , and imf_3 , sometimes with distinct negative peaks.

Figure 8 shows the differences between the different IMF amplitudes for all catchments. In all graphics, we can observe a very similar behavior of each of the IMF amplitudes for the three stations together, indicating that their close proximity produces similar runoff patterns. Their absolute values are different, however, due to higher values of the Dedenborn data (red curve). We also see that the location of extreme points of the three IMF amplitudes, such as at 21 Dec. 1991 or at 9 Jan. 1993, coincide. We can also see seasonal components from these amplitudes as well as from the original data: decreasing values in spring with lower values during summer and afterward increasing amplitudes, resulting in the highest amplitudes in December or January of each year.

Once the EMD is computed as in Fig. 9, an approximation of the original data can be defined as follows. Because the EMD is an exact additive representation of the signal, the first IMFs in the representation Eq. [1], containing the high frequencies, can be skipped. A coarse approximation to the original data can then be defined by setting

$$\tilde{s}(t) := \sum_{j=m}^n \text{imf}_j(t) + r_n(t) \quad [14]$$

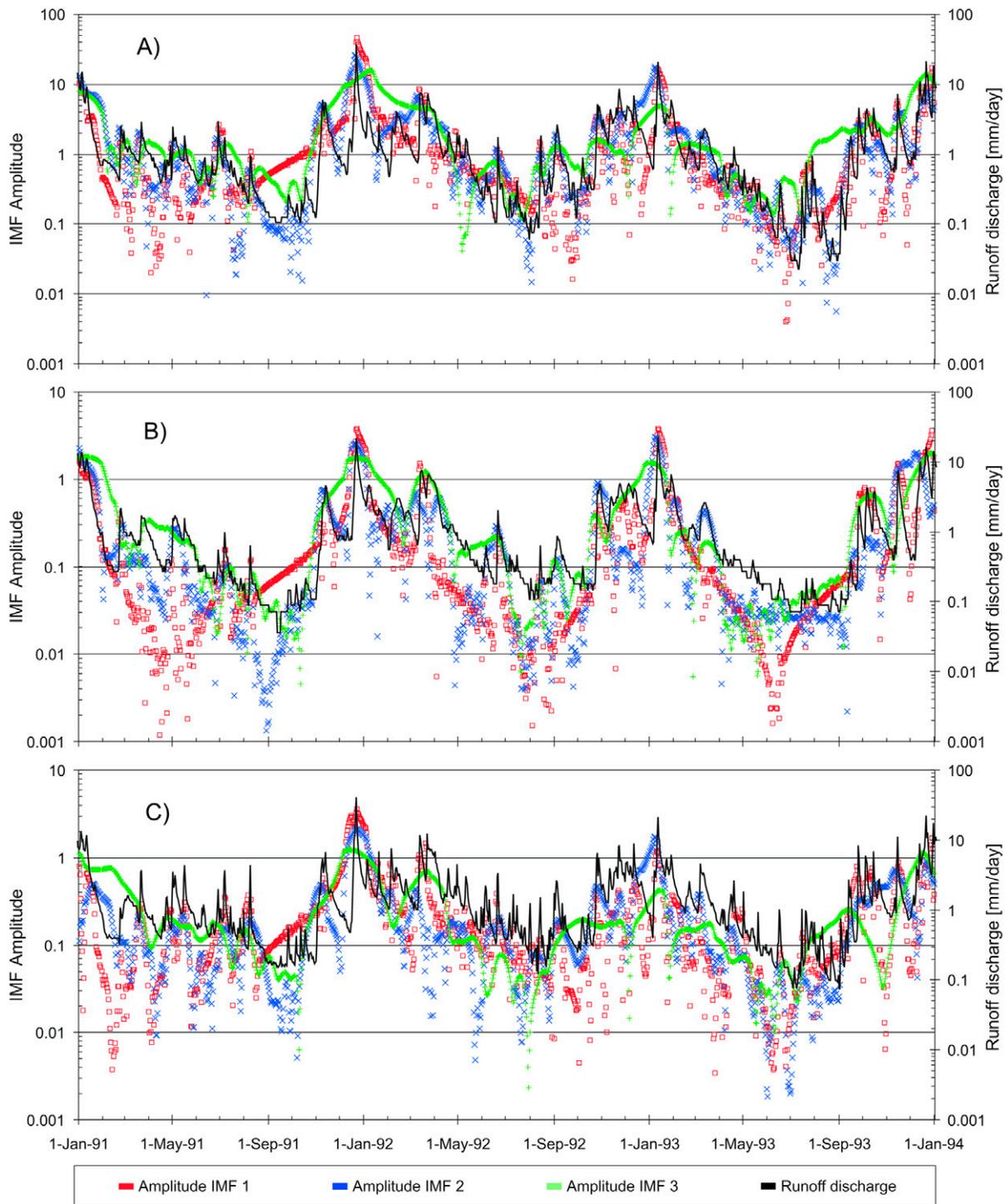


Fig. 7. Runoff discharge data from 1991 to 1993 from the three gauging stations compared with intrinsic mode function (IMF) amplitudes 1, 2, and 3: (A) Dedenborn, (B) Rollesbroich, and (C) Erkensruhr.

where $m > 1$ stands for the IMF with the lowest index. Frequencies should be taken into account from this index onward. Because this approximation contains all the relevant information but much less data because the high-frequency components have been taken out, this representation may be used for further processing within a model for parameterization.

Results and Discussion

Different Spectra

We now compare the Hilbert energy spectrum as defined above with other spectra. For this, we have first used the Dedenborn data

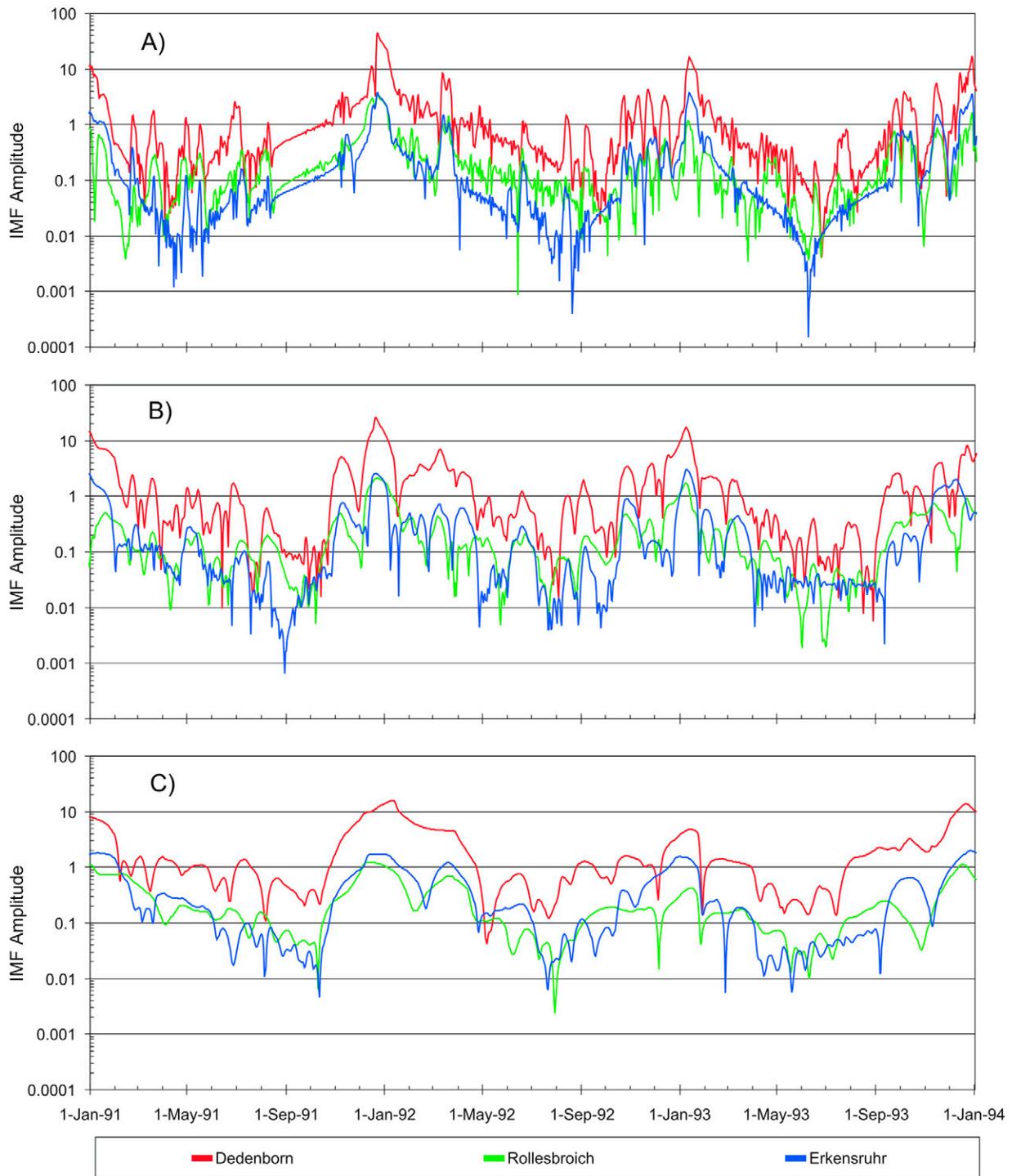


Fig. 8. Comparison of intrinsic mode function (IMF) amplitudes for all catchments: (A) imf_1 , (B) imf_2 , and (C) imf_3 .

from 1990 to 1999 displayed in Fig. 9A. The corresponding EMD is shown in Fig. 9B.

The corresponding Hilbert energy spectrum is shown in Fig. 10A. The ordinate axis displays the time period of the corresponding frequencies in days on a logarithmic scale, and the strength of the colors reveal the energy (i.e., the value of the amplitude squared) according

to the values shown on the right. The Hilbert spectrum, therefore, shows the distribution of the squares of the amplitudes and frequency as functions of time. Figures 10B and 10C display the wavelet energy spectrum using the continuous wavelet transform with the Morlet wavelet and the Mexican hat wavelet, respectively. Figure 10D is the spectrogram using the short-time (or windowed) Fourier transform

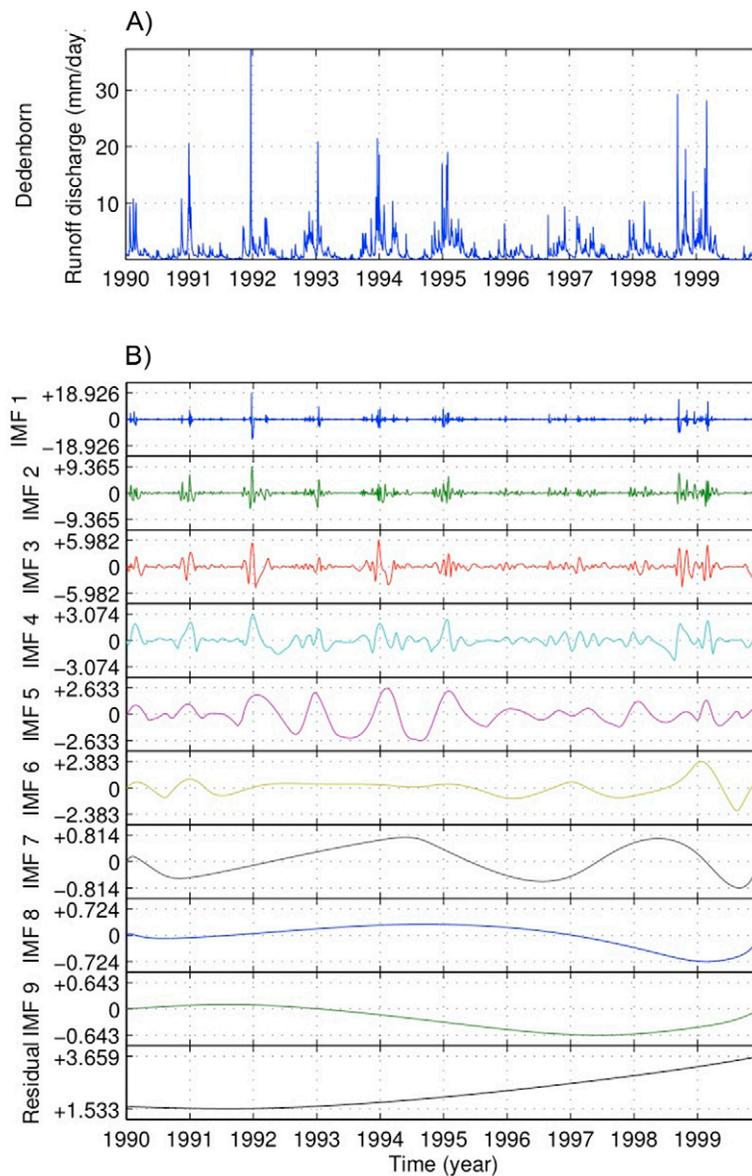


Fig. 9. Dedenborn site: (A) original data from 1990 to 1999, and (B) empirical mode decomposition into intrinsic mode functions (IMF) and residual.

(STFT) with a window width of 64 measured data values. All spectra are energy spectra with a linear scale of the colors.

The same computations were made with an enlargement of the Dedenborn data for the time period 1991 to 1993 in Fig. 11 and 12.

In the comparison in Fig. 10 between the Hilbert energy spectrum and the Fourier spectrogram, we see a much more localized behavior of the Hilbert energy spectrum. Moreover, in comparison with the wavelet spectra, the Fourier spectrogram even seems to emphasize regions of importance such as in 1994 or 1995 where the other spectra are not so dominant. In addition, the Fourier spectrogram fails to provide information for periods essentially smaller than 16 d. This effect is even more visible in Fig. 12 between the beginning of 1992 and 1993, or around February 1993—here the STFT exhibits

features that the other spectra do not show so that we can conclude that they may be faulty. For these reasons, we will not consider the STFT in subsequent comparisons. In Fig. 10, boundary effects can clearly be detected for the two wavelet transforms due to their periodic nature. This cannot be seen for the Hilbert energy spectrum.

We further enlarge the Dedenborn original data centered at the extreme event of 22 Dec. 1991 with visualization of period lengths, for which we show the Hilbert energy spectrum and two wavelet spectra in Fig. 13. The Hilbert energy spectrum shows a very clear and quite unusual nonlinear pattern while the two wavelet spectra essentially just exhibit the location. It is remarkable that both these spectra display high amplitude values only for periods <64 d while the Hilbert energy spectrum shows high amplitudes also for periods up to 256 d.

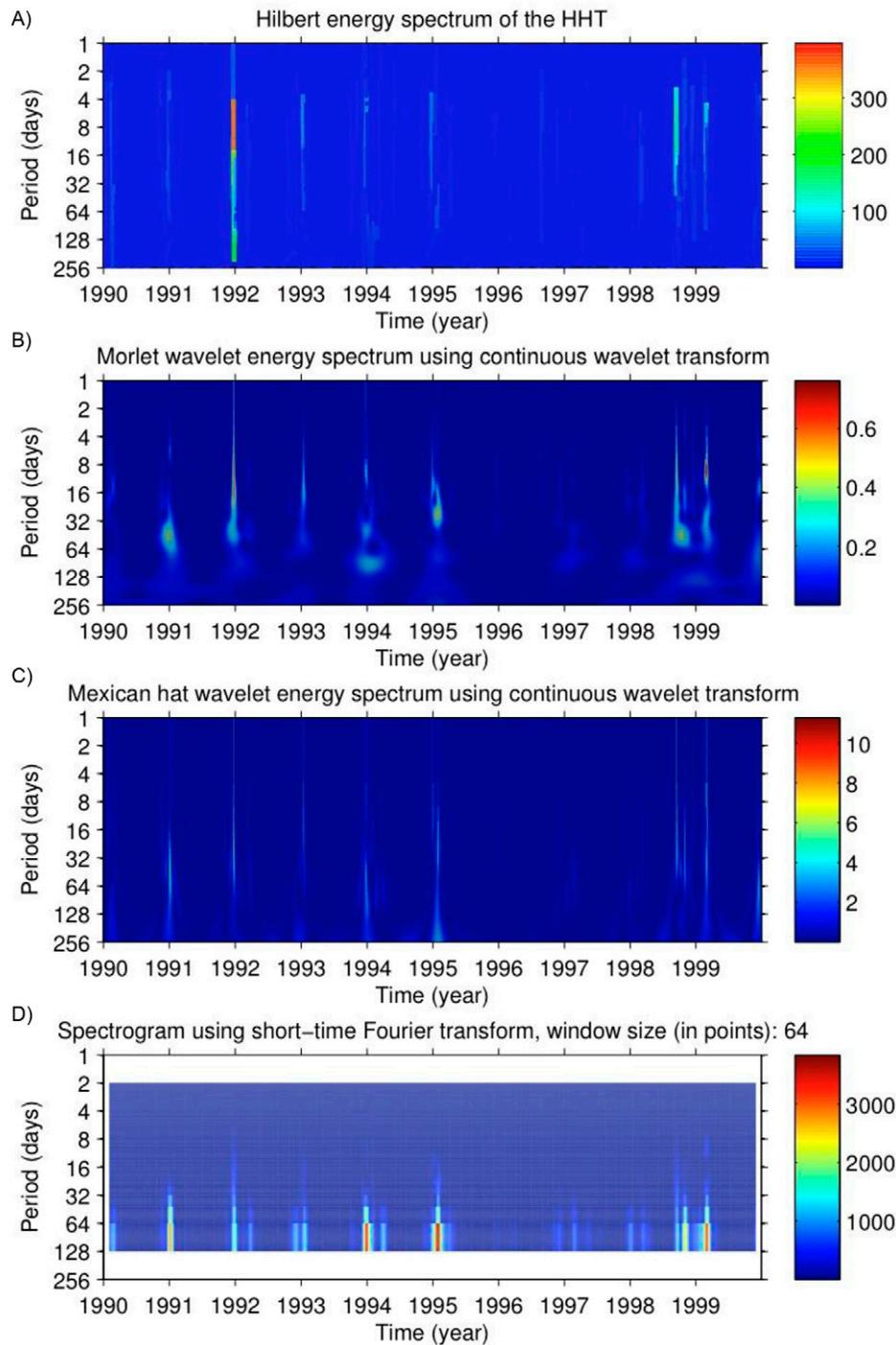


Fig. 10. Dedenborn site data from 1990 to 1999: (A) Hilbert energy spectrum of the Hilbert–Huang transform (HHT), (B) Morlet and (C) Mexican hat wavelet spectra, and (D) Fourier spectrogram.

Finally, we show in Fig. 14 the Hilbert energy spectra for all three sites together for the sample period 1991 to 1993. These have been scaled so that they display the same energy. It is apparent that all three of them exhibit the same patterns at the same locations. The higher values for Dedenborn account for higher values in the energy at the end of 1991.

It is apparent that the Hilbert energy spectra show very localized information that is even stronger than the wavelet spectra; the latter have a tendency to smear out, especially for increasing periods >64 d. Seasonal components with strong peaks in their energies around the turn of the years are most apparent for the Hilbert energy spectrum in Fig. 14.

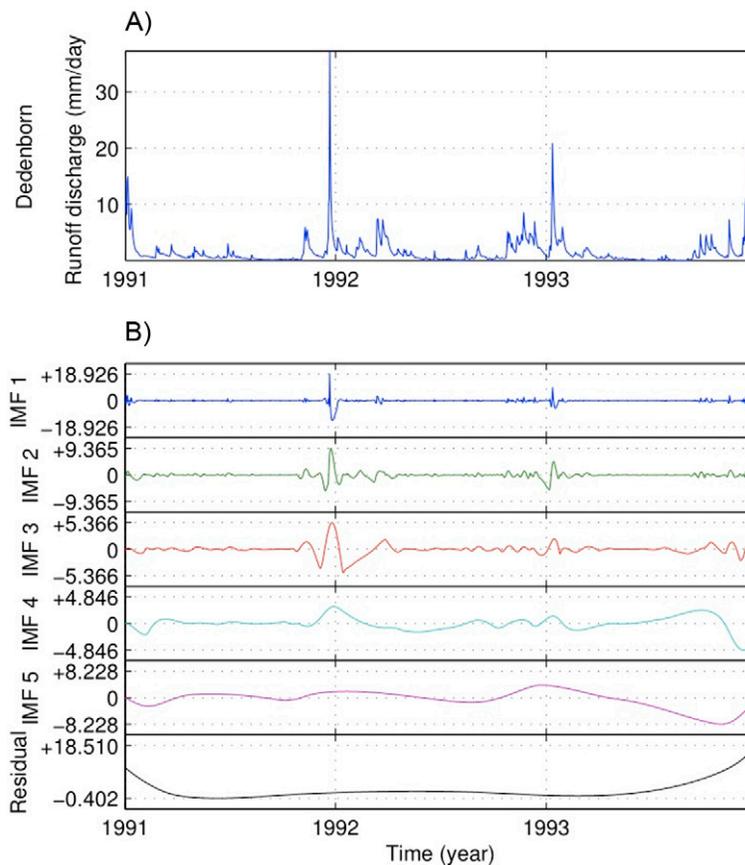


Fig. 11. Dedenborn site: (A) original data from 1990 to 1999, and (B) empirical mode decomposition into intrinsic mode functions (IMF) and residual.

The Hilbert energy spectrum clearly shows certain periodic appearances of similar horizontal distances that can be interpreted as years and seasons. The strength of the amplitudes, visualized by the different colors, indicate different amounts of drainage with time. From this, a time-dependent impact of higher amounts of data can be derived. We have displayed in Fig. 10 and 12 a comparison between the Hilbert spectrum and the Fourier spectrogram for details of the Dedenborn data. It can be seen that the Hilbert spectrum provides much more localized spectrum information than the short-time Fourier transform. This effect apparently becomes stronger the more one zooms into the data. In this way, one can interpret the Fourier spectrogram as a strongly “smeared-out” version of the Hilbert spectrum. Even when compared with wavelet spectra using continuous wavelet transforms, the Hilbert energy spectrum exhibits a stronger localized and less smeared-out energy.

Moreover, the Hilbert energy spectra of the three measurement stations in Fig. 14 appear very similar, indicating little local variability in drainage.

All computations of the EMD and the HHT spectra have been performed in Matlab, Version 7.9.0 (The MathWorks, Natick, MA). The Fourier spectrograms using the short-time Fourier transform were generated with the spectrogram function of the Matlab Signal

Processing Toolbox. The continuous wavelet transforms using the Morlet and the Mexican hat wavelets have been implemented using the Matlab Wavelet Toolbox according to the guide to wavelet analysis by Torrence and Compo (1998). All programs were written by Johann Rudi. More data and comparisons can be found in Rudi (2010).

Conclusions and Future Work

The HHT method presented here provides a refined analysis of hydrologic data when compared with classical Fourier analysis and wavelet analysis. Specifically, the possibility of introducing a time-dependent frequency and the computation of a localized time–frequency spectrum provides seasonal components together with their energies. Compared with wavelet spectra using continuous wavelet transforms or with Fourier spectrograms, the Hilbert energy spectrum exhibits a stronger localized and less smeared-out energy.

Because the EMD provides an additive decomposition of the original data, the different IMFs display different portions of the measured data with time-dependent frequencies that become larger for larger indices j . Clearly, a coarse approximation of the data could be obtained by summing only the IMFs for, say, $j \geq 3$. In view of the detection and the parameterization of multiscale patterns and shapes, the method presented here provides a way to

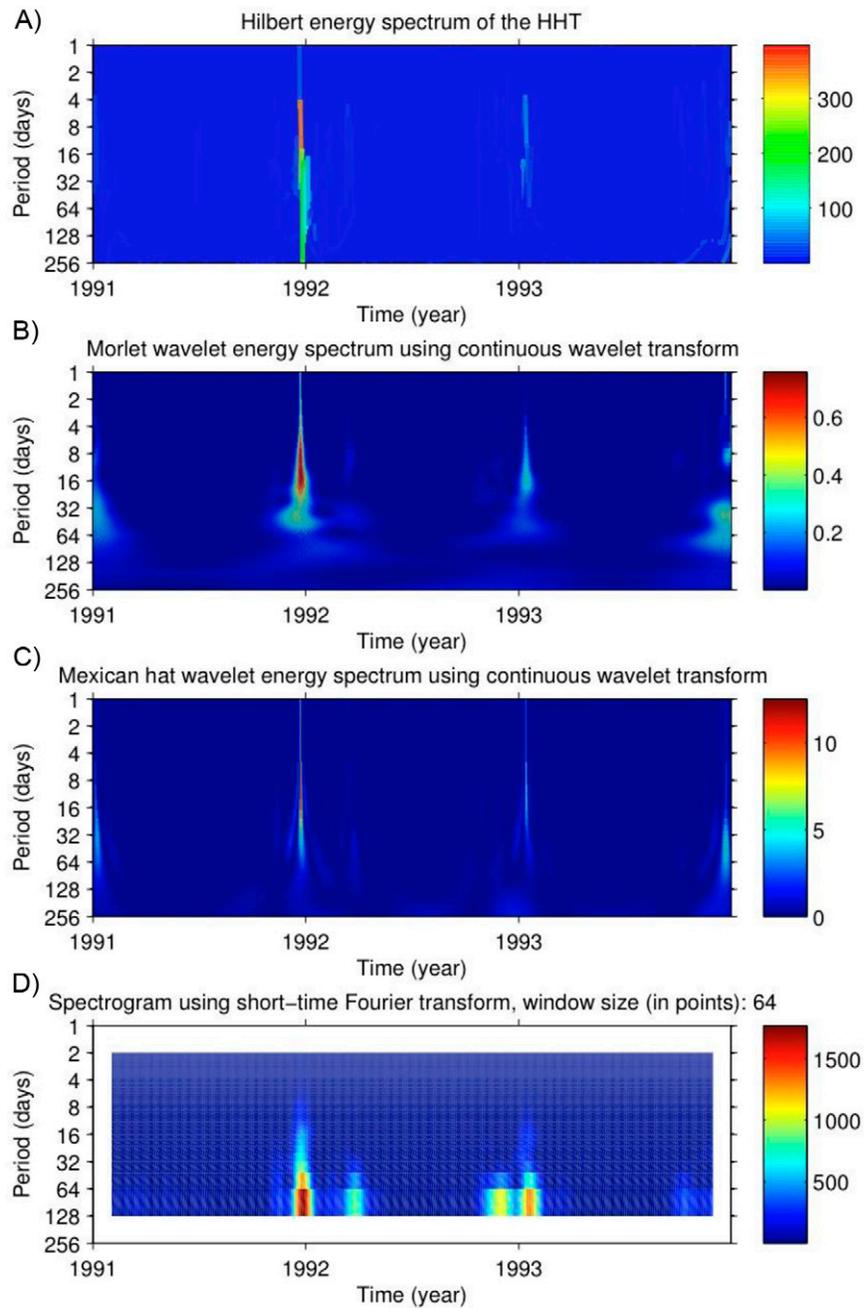


Fig. 12. Dedenborn site data from 1991 to 1993: (A) Hilbert energy spectrum of the Hilbert–Huang transform (HHT), (B) Morlet and (C) Mexican hat wavelet spectra, and (D) Fourier spectrogram.

characterize the data set consisting of thousands of data points in terms of very few parameters (years and seasonal components and the distribution of their amplitudes and their energies). These may then be utilized in the numerical simulation models for fluxes.

Although the method has proved to be successful in many applications and has perhaps provided more insight into the data than conventional methods, it should be mentioned that theoretical aspects of the EMD, like a mathematical convergence theory, the complexity of the iterative scheme in terms of floating point operations, or the appropriate handling of boundary data is still not

understood. For the complexity issue, the problem is the EMD algorithm, which is based on two nested iterations (see above). Both need the specification of thresholding parameters to terminate the iterations. The difficulty of an appropriate choice of these parameters has been discussed before, e.g., in Huang and Shen (2005, p. 8–9). It has been observed by us and others that a slightly different choice of these thresholding parameters may yield a quite different number of IMFs and also different patterns of these. Accordingly, the total amount of outer and inner iterations cannot be determined beforehand and, in the worst case, the algorithm may not converge at all. To overcome the parameter choice in

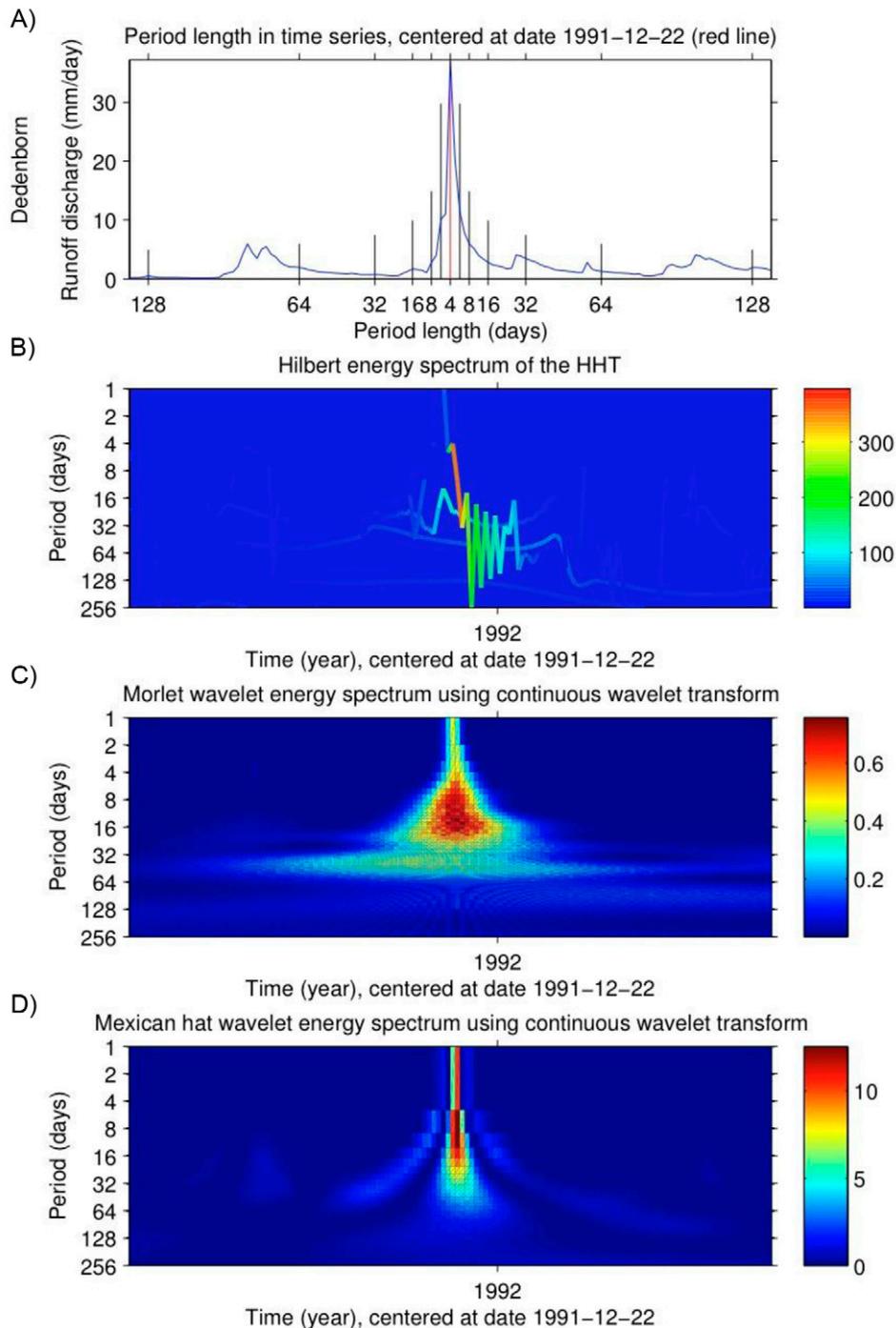


Fig. 13. Dedenborn site: original data centered at 22 Dec. 1991 with (A) visualization of period lengths, (B) Hilbert energy spectrum, and (C) Morlet and (D) Mexican hat wavelet spectra.

the sifting process, an alternative that works with linear functionals for B splines instead of upper and lower hulls of the data and that is cheaper to compute was proposed in Chen et al. (2006) and further elaborated in Koch (2008) and Jager et al. (2010).

Of course, once the EMD is determined, the complexity of computing the Hilbert spectrum can be determined easily because it

just requires application of the fast Fourier transform. First steps were undertaken toward a theoretical understanding of the overall HHT process in Sharpley and Vatchev (2006).

When it comes to the analysis of multivariate data sets, the EMD process can become numerically quite expensive or even prohibitive for space dimensions $n \geq 3$. Different approaches for the

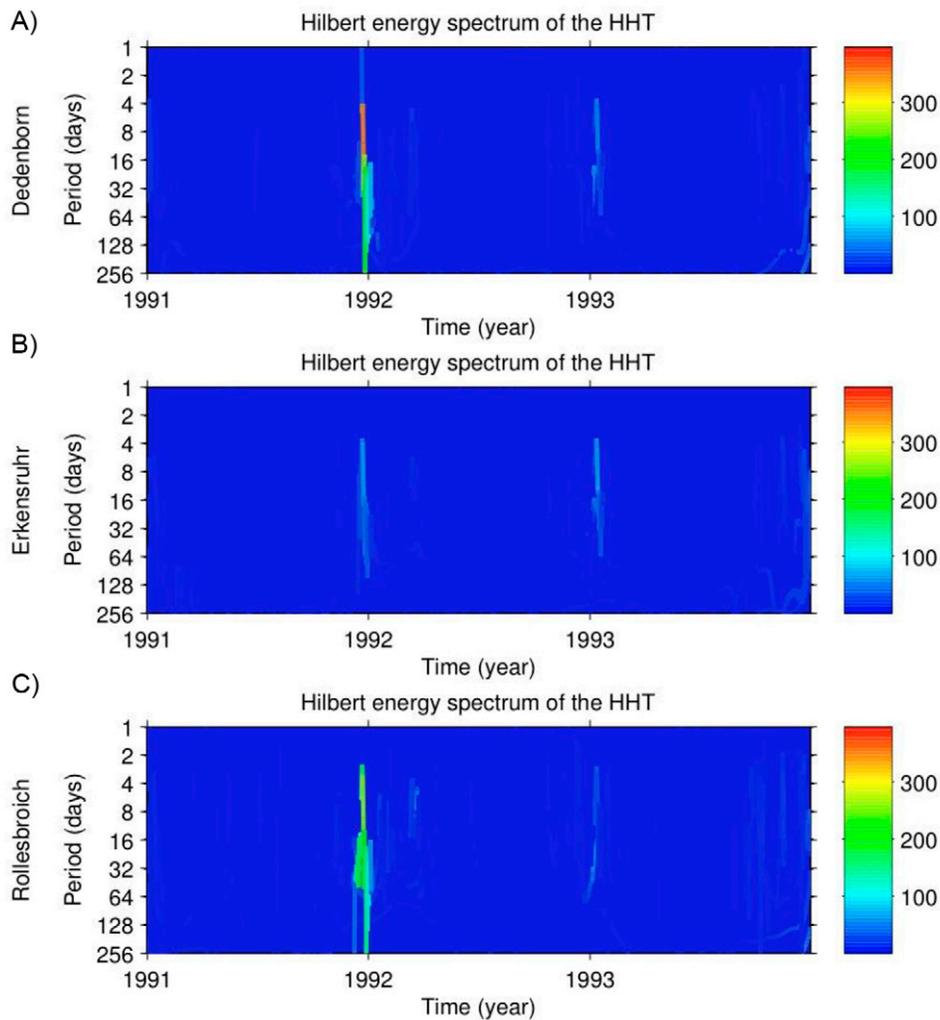


Fig. 14. Hilbert energy spectra of the Hilbert–Huang transform (HHT) with the same energy scaling for the Dedenborn, Erkenstruhr, and Rollesbroich gauging stations from 1991 to 1993.

two-dimensional case based on finite elements on Delaunay triangulations or thin plate splines or (in the univariate case) by solving a time-dependent partial differential equation have been proposed by Deléclle et al. (2005), Damerval et al. (2005), and Xu et al. (2006). A systematic comparison for the two-dimensional case that focuses on the quality of the EMD and the numerical performance was given in Koch (2008) and Jager et al. (2010). Specifically, a new method based on adaptive spline wavelets was proposed that ultimately yields the best results when it comes to the quality of the EMD and the speed of the computations. Equally important, an analysis of the theoretical setup for the general, n -dimensional case by means of Clifford algebra and monogenic functions, and a definition of a generalized Hilbert transform for the multivariate case, was provided.

We applied the EMD methodology to compare long-term daily runoff discharge measurements taken from three rivers of very different size. We demonstrated that these time series can be

successfully separated into several IMF modes. The runoff peaks mostly resulted in high IMF amplitude values, especially for imf_1 , which is the portion of the signal corresponding to the highest frequencies. Extreme events produced also distinct amplitude peaks for imf_2 and imf_3 . In contrast, low-flow periods are usually accompanied by low amplitudes for imf_1 , imf_2 , and imf_3 , sometimes even with distinct negative peaks. The similarity of the IMF modes indicate a similar runoff pattern due the close locality and similar environmental setting of the catchments. Because Huang et al. (2009) also were able to apply the EMD methodology for hydrologic time series data, we are confident that the EMD will also be useful for the analysis of other time series data that are similar to runoff data (e.g., soil moisture). Furthermore, EMD should also be helpful for the validation of rainfall–runoff models, which has already been demonstrated for the wavelet domain by Schaeffli and Zehe (2009). The next step will be to use the HHT method for the scale-dependent characterization of soil moisture patterns as measured by the sensor network deployed in the framework of the TR32 project (Bogena et al., 2010).

Acknowledgments

We gratefully acknowledge financial support by the SFB/TR 32 "Pattern in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling, and Data Assimilation" (www.tr32.de), funded by the Deutsche Forschungsgemeinschaft (DFG).

References

- Bachner, S., A. Kapala, and C. Simmer. 2006. Niederschlagsvariabilität während der letzten 100 Jahre in Deutschland. Teil III: Zeitskalen der Variabilität hydrometeorologischer Größen. In German Climate Conf. 7th, Munich, Germany. 9–11 Oct. 2006.
- Bogena, H.R., J.F. Hake, M. Herbst, R. Kunkel, C. Montzka, T. Puetz, H. Vereecken, and F. Wendland. 2005a. MOSYRUR: Water balance analysis in the Rur basin. Schrift. Forschungszentrums Jülich, Reihe Umwelt 52.
- Bogena, H.R., M. Herbst, J.A. Huisman, U. Rosenbaum, A. Weuthen, and H. Vereecken. 2010. Potential of wireless sensor networks for measuring soil water content variability. *Vadose Zone J.* 9:1014–1024 (this issue).
- Bogena, H.R., J.A. Huisman, H. Meier, U. Rosenbaum, and A. Weuthen. 2009. Hybrid wireless underground sensor networks: Quantification of signal attenuation due to soil adsorption. *Vadose Zone J.* 8:755–761.
- Bogena, H.R., J.A. Huisman, C. Oberdörster, and H. Vereecken. 2007. Evaluation of a low-cost soil water content sensor for wireless network applications. *J. Hydrol.* 344:32–42.
- Bogena, H., R. Kunkel, T. Schöbel, H.P. Schrey, and F. Wendland. 2005b. Distributed modelling of groundwater recharge at the macroscale. *Ecol. Modell.* 187:15–26.
- Castaño, D., and A. Kunothe. 2006. Robust regression of scattered data with adaptive spline-wavelets. *IEEE Trans. Image Process.* 15:1621–1632.
- Castaño Diez, D., G. Jager, and A. Kunothe. 2009. Multiscale analysis of multivariate data. In K. Gürlebeck and C. Könke (ed.) *Proc. Int. Conf. on Appl. of Comput. Sci. and Math. in Architecture and Civ. Eng.*, 18th, Weimar, Germany. 7–9 July 2009 [CD-ROM]. Univ. of Weimar, Weimar, Germany.
- Chen, Q., N.E. Huang, S. Riemenschneider, and Y. Xu. 2006. A B-spline approach for empirical mode decomposition. *Adv. Comput. Math.* 24:171–195.
- Cohen, L. 1994. *Time-frequency analysis*. Prentice Hall, Englewood Cliffs, NJ.
- Damerval, Ch., S. Meignen, and V. Perrier. 2005. A fast algorithm for bidimensional EMD. *IEEE Signal Process. Lett.* 12:701–704.
- Deléclhelle, E., J. Lemoine, and O. Niang. 2005. Empirical mode decomposition: An analytical approach for sifting process. *IEEE Signal Process. Lett.* 12:764–767.
- DeVore, R.A., and A. Kunothe (ed.). 2009. *Multiscale, nonlinear and adaptive approximation*. Springer-Verlag, Berlin.
- Felsberg, M., and G. Sommer. 2000. The multidimensional isotropic generalization of quadrature filters in geometric algebra. *Lect. Notes Comput. Sci.* 1888:175–185.
- Felsberg, M., and G. Sommer. 2001. The monogenic signal. *IEEE Trans. Signal Process.* 49:3136–3144.
- Flandrin, P. 1999. *Time-frequency/time-scale analysis*. Academic Press, San Diego.
- Flandrin, P., and P. Gonçalves. 2004. Empirical mode decompositions as data-driven wavelet-like expansions. *Int. J. Wavelets Multires. Inf. Process.* 2:477–496.
- Flandrin, P., G. Rilling, and P. Gonçalves. 2004. Empirical mode decompositions as a filter bank. *IEEE Signal Process. Lett.* 11:112–114.
- Grayson, R., and G. Blöschl (ed.). 2001. *Spatial patterns in catchment hydrology: Observations and modelling*. Cambridge Univ. Press, Cambridge, UK.
- Hahn, S. 1995. *Hilbert transforms in signal processing*. Artech House, Norwood, MA.
- Huang, N.E., and S.S.P. Shen. 2005. Hilbert–Huang transform and its applications. *World Sci. Publ.*, London.
- Huang, N.E., Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zhang, N.-C. Yen, C.C. Tung, and H.H. Liu. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London Ser. A* 454:903–995.
- Huang, Y.X., F.G. Schmitt, Z.M. Lu, and Y.L. Liu. 2009. Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis. *J. Hydrol.* 373:103–111.
- Kang, S., and H. Lin. 2007. Wavelet analysis of hydrological and water quality signals in an agricultural watershed. *J. Hydrol.* 338:1–14.
- Koch, R. 2008. *Analyse multivariater Daten: Konstruktion monogener Clifford-Algebra-wertiger Funktionen mittels Empirical Mode Decomposition basierend auf adaptiven kubischen Spline-Wavelets und der Riesz-Transformation* (in German). Diploma thesis. Inst. of Numer. Simul., Univ. of Bonn, Bonn, Germany.
- Linderhed, A. 2004. Image compression based on empirical mode decomposition. p. 110–113. In *Proc. SSAB Symp. on Image Analysis*, Uppsala, Sweden. March 2004.
- Pabel, R., R. Koch, G. Jager, and A. Kunothe. 2010. Fast empirical mode decompositions of multivariate data based on adaptive spline-wavelets and a generalization of the Hilbert–Huang-Transform (HHT) to arbitrary space dimensions. *Adv. Adaptive Data Anal.* (in press).
- Rilling, G., P. Flandrin, and P. Gonçalves. 2003. On empirical mode decompositions and its algorithms. In *IEEE-EURASIP Worksh. on Nonlinear Signal and Image Processing (NSIP-03)*, Grado, Italy.
- Rosenbaum, U., J.A. Huisman, A. Weuthen, H. Vereecken, and H.R. Bogena. 2010. Sensor-to-sensor variability of the ECH₂O EC-5, TE, and STE sensors in dielectric liquids. *Vadose Zone J.* 9:181–186.
- Rudi, J. 2010. *Empirical mode decomposition via adaptiver wavelet-approximation*. (In German.) Diploma Thesis. Inst. für Mathematik, Universität Paderborn, Paderborn, Germany.
- Schaeffli, B., and E. Zehe. 2009. Hydrological model performance and parameter estimation in the wavelet-domain. *Hydrol. Earth Syst. Sci.* 13:1921–1936.
- Sharpley, R.C., and V. Vatchev. 2006. Analysis of the intrinsic mode functions. *Constr. Approx.* 24:14–47.
- Si, B.C. 2008. Spatial scaling analyses of soil physical properties: A review of spectral and wavelet methods. *Vadose Zone J.* 7:547–562.
- Titchmarsh, E.C. 1950. *Introduction into the theory of Fourier integrals*. Oxford Univ. Press, Oxford, UK.
- Torrence, C., and G.P. Compo. 1998. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.* 79:61–78.
- Western, A.W., S.-L. Zhou, R.B. Grayson, T.A. McMahon, G. Blöschl, and D.J. Wilson. 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant spatial hydrological processes. *J. Hydrol.* 286:113–134.
- Xu, Y., B. Liu, J. Liu, and S. Riemenschneider. 2006. Two-dimensional empirical mode decomposition by finite elements. *Proc. R. Soc. London Ser. A* 462:3081–3096.