Multiscale Analysis of Hydrologic Time Series Data using the Hilbert-Huang-Transform (HHT)

Johann Rudi[1], Roland Pabel[1], Gabriela Jager[1], Robin Koch[1], Angela Kunoth[1]
Heye Bogena[2]

For the analysis of time series data from hydrology, we employ a technique recently developed in Huang et al., 1998, which 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 employing 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, one computes an additive decomposition called *empirical mode decomposition (EMD)* of the data into certain multiscale components. 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 paper, we recall the necessary components of the HHT and apply it to hydrological time series data from the Upper Rur Catchment Area, mostly German territory, taken over a period of twenty years. Our first observation is 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 reveals in a particular apparent way seasonal events together with their energy. We provide a comparison of the Hilbert spectra with Fourier spectrograms and wavelet spectra in order to demonstrate a better localization of the energy components which also exhibit strong seasonal components. Finally, the Hilbert energy spectrum of the three measurement stations appear to be very similar, indicating little local variability in drainage.

1 Introduction

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

This study is part of a SFB/TR32 project (www.tr32.de) 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

[1]Institut für Mathematik, Universität Paderborn, Warburger Str. 100, 33098 Paderborn, Germany,
{johannr,pabel,gconst,kunoth}@math.uni-paderborn.de
http.//www2.math.uni-paderborn.de/ags/agkunoth/group/angelakunoth.html
[2]Forschungszentrum Jülich GmbH, ICG-4 52425 Jülich, Germany,
h.bogena@fz-juelich.de, www.fz-juelich.de/icg/icg-4/index.php?index=484

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 are interacting in a complex nonlinear way, see, e.g. Grayson & 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 scale has been set up, see Bogena et al., 2007, Bogena et al., 2009, Bogena et al., 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 river streams with low management influence located in the southern part of the Rur catchment, see Section 3.1. Since runoff discharge and soil moisture time series data exhibit similar temporal patterns, this paper can be considered as a preliminary study for the analysis of the soil moisture sensor network data.

For the detection of structures and patterns living on different scales, the method of choice is to transform and thereby decompose measurement data into multi-scale components. Classical and widely established methods are the short-time Fourier transform (STFT) 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 & Kunoth, 2006, Castaño et al., 2009, DeVore & Kunoth, 2009.

An example for the analysis of hydrometeorologic data by means of wavelets is given in Bachner et al., 2006. Here the goal was the identification and filtering of dominant time scales in statistical indices of daily time series. The wavelet analysis provided a temporally varying power spectrum and does not require stationarity of the data. In Kang & Lin, 2007, a wavelet analysis of hydrological and water quality signals in an agricultural watershed was performed, using a weighted wavelet-Z-transform. Schaefli & 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 is nonlinear and non-stationary, these transforms often do not lead to physically meaningful results. The reasons are fundamental construction ingredients: they only allow certain linear combinations of predefined bases and for each component the same frequency over the whole time domain is employed. Although wavelets have the advantage over the Fourier basis that they allow a localized identification of the significant frequency components, the motivation for a new approach developed in Huang et al., 1998, was to find a better decomposition into components which are adapted to the specific data set.

In this paper, we follow this approach to analyse nonlinear and non-stationary 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 non-stationary time series in one dimension. The principle of the HHT is as follows: first, one decomposes iteratively the time series into empirical adaptive nonlinear modes (intrinsic mode functions, abbreviated as IMFs)

which 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 provides then a localized time-frequency spectrum and a possible extraction of instantaneous frequencies.

By now this method has been employed for many physical long- and short-term univariate data, see, e.g., Huang & Shen, 2005, and references therein. A very recent paper where Hilbert spectral analysis was applied to hydrological data is Huang et al., 2009; here daily river flow fluctuations are analyzed.

The original scheme has been modified and refined by several authors during the past decade, see, e.g., Chen et al., 2006, Deléchelle et al., 2005, Flandrin & Gonçalvès, 2004a, Flandrin et al., 2004b. First extensions to derive an empirical mode decomposition of two--dimensional data may be found in Damerval et al., 2005, Xu et al., 2006, and applications to adaptive image compression in Linderhed 2004. In Koch, 2008, several methods for computing the EMD for two-dimensional data are compared with respect to their numerical performance. This is extended in 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 has to be generalized to extend analytical signals to arbitrary dimensions. Based on ideas employing Clifford algebras and Riesz transforms from Felsberg & Sommer, 2000, Felsberg & Sommer, 2001, a corresponding fast numerical scheme and results on two--dimensional data can be found in Koch, 2008, Jager et al., 2010.

In this paper, we present the Hilbert-Huang-Transform (HHT), following the collection of material in Koch, 2008, and apply it to hydrological time series data. In Section 2.1, we introduce a basic iterative scheme for computing the empirical mode decomposition of univariate time series data. Section 2.2 is devoted to some mathematical ingredients for deriving the Hilbert transform, including the concept of *instantaneous frequencies* and the *Hilbert energy spectrum*. Section 3 is devoted to computing the EMD and appropriate approximations of some long-term hydrological time series data. We analyze the data with the HHT in Section 3.3 and include a comparison with the Fourier spectrogram and the wavelet spectrum. Finally, Section 4 summarizes our main conclusions and provides directions for future work, both for the analysis of hydrological data as well as for improvements of computations for the Hilbert-Huang-Transform and extensions to higher spatial dimensions.

2 The Hilbert-Huang-Transform (HHT)

2.1 Empirical Mode Decompositions (EMDs)

The advantage of the Hilbert-Huang-Transform (HHT) developed in Huang et al., 1998, over the Fourier decomposition or a wavelet decomposition is that it can be used to analyze nonlinear and non-stationary data over irregular time grids. The HHT works by iteratively decomposing the time series into a finite number of intrinsic mode functions (IMFs) through an empirical mode decomposition (EMD) process. After the data-driven additive decomposition is obtained, one applies in a second step detailed below in Section 2.2 to each IMF component 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 in Huang et al., 1998, with a

synthetic example from Rilling et al., 2003. One calls a set of real-valued time series data $\{(t_\ell, z_\ell)\}_{\ell \in \mathsf{Z}}$ a *nonlinear* and *non-stationary* data set if there exists a number $m \in \mathsf{Z}$ such that the common probability distribution of $z_\ell, \ldots, z_{\ell+m}$ depends on the time index ℓ. All sorts of measurement data from physical processes usually satisfy this condition. It will be convenient to describe the methods by considering instead of discrete time series data a continuous function $s : \mathsf{R} \to \mathsf{R}$. 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 allow later to define 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 $imf : \mathsf{R} \to \mathsf{R}$ is an *intrinsic mode function* (IMF) if (i) the number of local extrema and zero points of IMF differ mostly by one; (ii) at any point, the mean value of the cubic spline which interpolates all local maxima and of the cubic spline which interpolates all local minima is zero. Recall that a cubic spline is a function consisting of piecewise polynomials of degree three joined together such that their second derivative is still a continuous function. Thus, an IMF represents a basic oscillation which is symmetrically localized around the ordinate axis. The collection of IMFs may be viewed as a data-adapted basis which 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 instantaneous amplitude and frequency. In contrast, the Fourier basis and the wavelet basis are known a-priori.

For illustration purposes, we will describe the subsequent EMD process using a synthetic function displayed in Figure 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.

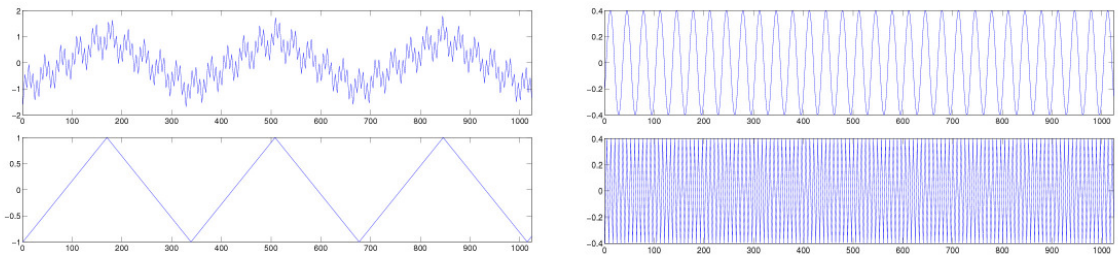


Figure 1: Synthetic function s (top left), additively composed from a sine wave (top right) and two piecewise linear continuous functions (bottom) with different periodicities.

The decomposition is based on two iterations, one inner loop called *sifting* which has generated after its completion a single IMF, and an outer loop which consists of the decomposition into the different IMFs. For computing a single IMF, one first computes an upper and lower envelope s_+, s_- of the local maxima, resp. minima, consisting of cubic splines which

contain in its convex hull the original signal s, see Figure 2 (top).

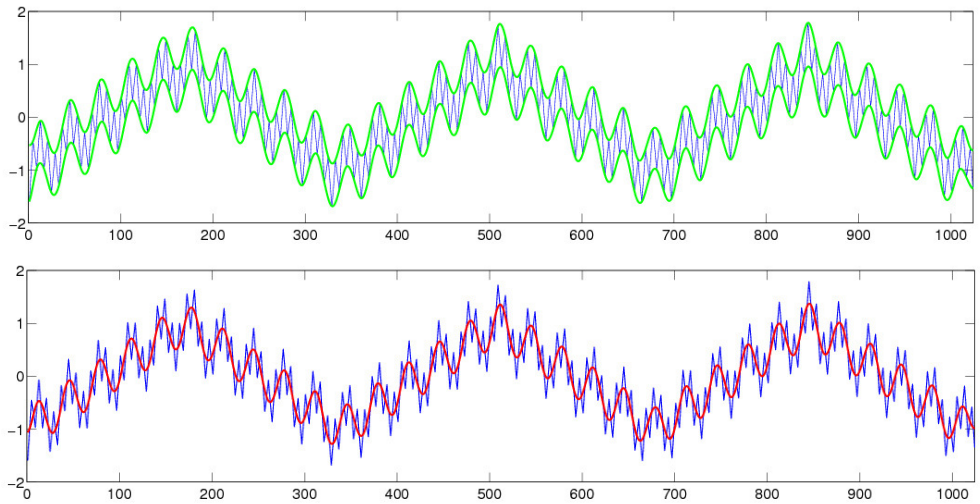
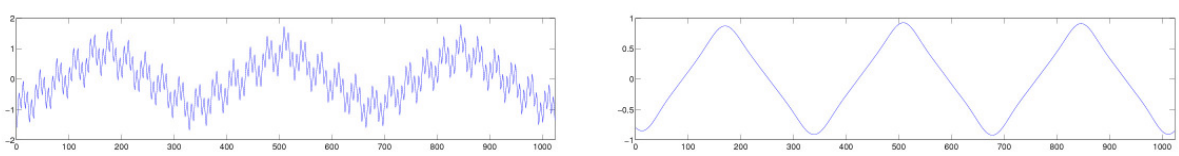


Figure 2: Synthetic function s between its upper envelope s_+ and its lower envelope s_- (top), synthetic function s and its mean function $m_{1,1}$ (bottom).

For the difference between the original signal $s(t)$ and the mean $m_{1,1}(t) := \frac{1}{2}(s_+(t) + s_-(t))$ which is shown in Figure 2 (bottom), 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 criteria is met; for details and a discussion of appropriate termination criteria, see Huang et al., 1998, Koch, 2008. The result is called $imf_1(t)$. The next inner sifting process then starts with the residual $r_1(t) := s(t) - imf_1(t)$, yielding the second $imf_2(t)$. After termination of the last outer iteration, combining all components finally results in an additive decomposition

$$s(t) = \sum_{j=1}^{n} imf_j(t) + r_n(t) \qquad (1)$$

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 Figure 1, the result of the iterative process, the empirical mode decomposition (EMD), is shown in Figure 3.



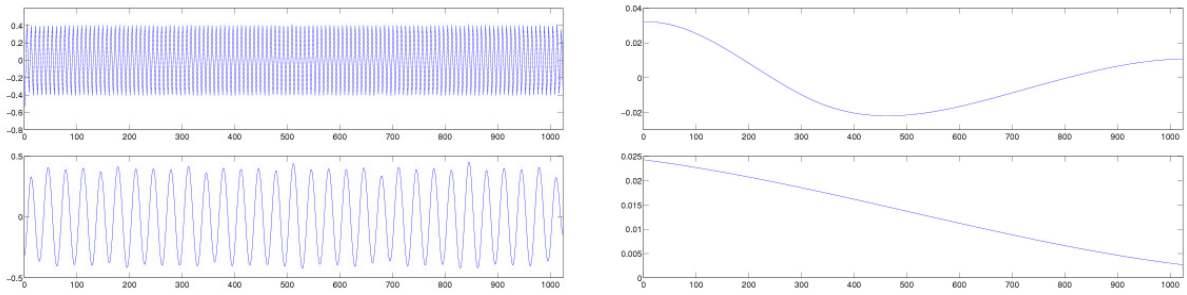


Figure 3: Empirical mode decomposition (EMD) of synthetic function s (top left) into imf_1 (middle left), imf_2 (bottom left), imf_3 (top right), imf_4 (middle right) and a monotone residual r_4 (bottom right). Note the different ordinate scaling for the last two components in comparison to the other ones.

Although synthetic, this example is very illustrative: one sees 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 to 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 in this order.

Note that the different IMFs are approximations of linear combinations of cubic B-splines: they are approximations since the process is iterative. This can be seen from the results for the synthetic example by comparing the plots in Figure 1 with the ones in Figure 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 we do 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 (1) (or (11) below) and e.g. the Fourier representation (12) below is that the Fourier representation has a fixed basis $\exp(i\omega_j t)$, 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 (11) allow for a time-dependent coefficient which is constructively adapted to the data.

2.2 Hilbert Spectral Analysis

Once the data is decomposed into an EMD according to (1), one can apply to each IMF component the Hilbert transform, and one can compute instantaneous frequencies 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, *analytic 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 (7) below. Then these negative frequency components can be discarded without loosing information, under the condition that one accepts to compute now with a complex-valued function 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 & Shen, 2005.

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

$$\mathsf{H}[s](t) := \frac{1}{\pi} PV\int_{-\infty}^{\infty}\frac{s(u)}{t-u}du := \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{-\infty}^{t-\varepsilon}\frac{s(u)}{t-u}du + \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{t+\varepsilon}^{\infty}\frac{s(u)}{t-u}du, \quad t\in\mathsf{R}, \tag{2}$$

see, e.g., Titchmarsh, 1950. We will exploit certain properties of the Hilbert transform in connection with the *Fourier transform* $\mathsf{F}[s]$ of a signal $s:\mathsf{R}\to\mathsf{R}$, defined by

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

One denotes F as the *Fourier spectrum* of s and often abbreviates $\hat{s} := \mathsf{F}[s]$. 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 over time. In contrast, employing the Hilbert transform will allow to define a frequency depending on time. This will be achieved by first constructing from the real-valued given data an analytical signal which is by definition complex-valued, using the Hilbert transform. For this analytical signal, then an instantaneous (time-dependent) frequency can be defined.

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

$$\mathsf{F}[\mathsf{H}[s]](\omega) = \frac{\sqrt{2\pi}}{\pi}\mathsf{F}[\frac{1}{s}](\omega)\mathsf{F}[s](\omega) = (-i)sign(\omega)\hat{s}(\omega). \tag{4}$$

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

For the practical computation of the Hilbert transform, one can use that for a (real-valued) signal $s:\mathsf{R}\to\mathsf{R}$ one has

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

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

The `complexification' of given (real-valued) data $s:\mathsf{R}\to\mathsf{R}$ can be achieved as follows. Denote by $v(t) := \mathsf{H}[s](t)$ the Hilbert transform of s and define the *analytic signal* as

$$s_A(t) := s(t) + iv(t), \quad t\in\mathsf{R}. \tag{6}$$

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

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

The analytical signal s_A can also be represented as

$$s_A(t) = a(t)e^{i\phi(t)} \tag{8}$$

with *amplitude* $a(t) := \sqrt{s^2(t) + v^2(t)}$ and *phase* $\phi(t) := \arctan\left(\frac{v(t)}{s(t)}\right)$. We define now the *instantaneous frequency* $\omega = \omega(t) : \mathbb{R} \to \mathbb{R}$ of s as

$$\omega(t) := \frac{d}{dt}\phi(t). \tag{9}$$

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

$$\int_{-\infty}^{\infty} \omega \, | \, \hat{s}_A(\omega) \, |^2 \, d\omega = \int_{-\infty}^{\infty} \left(\frac{d}{dt}\phi(t) \right) | \, s_A(t) \, |^2 \, dt. \tag{10}$$

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

$$\begin{aligned} s(t) &= \sum_{j=1}^{n} imf_j(t) + r_n(t) = \sum_{j=1}^{n} \Re\left(imf_j(t) + i\mathrm{H}[imf_j](t) \right) + r_n(t) \\ &= \Re\left(\sum_{j=1}^{n} \left(imf_j(t) + i\mathrm{H}[imf_j](t) \right) \right) + r_n(t) \\ &= \Re\left(\sum_{j=1}^{n} a_j(t) \exp\left(i \int_{\mathbb{R}} \omega_j(t) dt \right) \right) + r_n(t). \end{aligned} \tag{11}$$

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

As intended, the representation (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,

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

one observes that the components of the latter only have constant amplitude and frequency. In this sense, the EMD provides a generalized Fourier representation which is particularly appropriate for non-stationary and nonlinear data.

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 (11) . Then the *Hilbert amplitude spectrum* is defined as

$$H(\omega,t) = H(\omega(t),t) := \begin{cases} a_1(t) & \text{on the curve } \{(\omega_1(t),t) : t \in \mathbb{R}\}, \\ \vdots & \\ a_n(t) & \text{on the curve } \{(\omega_n(t),t) : t \in \mathbb{R}\}, \\ 0 & \text{elsewhere.} \end{cases} \tag{13}$$

Note, however, that one can also define the *Hilbert energy spectrum* by taking squares of the amplitudes. In fact, as it is written in Huang et al., 1998, p. 928, 4th paragraph, `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.'

In the subsequent computations in Section 3.3 below, we have chosen the Hilbert energy spectrum (taking squares of the amplitudes in (13)) since then the results can be compared better to the Fourier spectrogram and to the wavelet spectrum.

3 Application to Hydrological Time Series Data

The concept described in the previous section is now applied to different hydrological data sets described in Bogena et al., 2005a, Bogena et al., 2005b.

3.1 Basic Characteristics of the Investigation Area

The catchment of the Rur River has been chosen as the regional investigation area for the SFB/TR32 project. The Rur catchment covers a total area of 2354 km^2 and is situated in Western Germany. For this study we selected data from three runoff gauging stations (Dedenborn, Erkensruhr and Rollesbroich) located in the southern part of the Rur catchment (see Figure 4). The runoff discharge was measured over a period of 6940 days (roughly twenty years between 1981 and 2001). The associated catchments of these stations are different in size (from 20 to 200 km^2) and exhibit varying catchment characteristics (see Table 1).

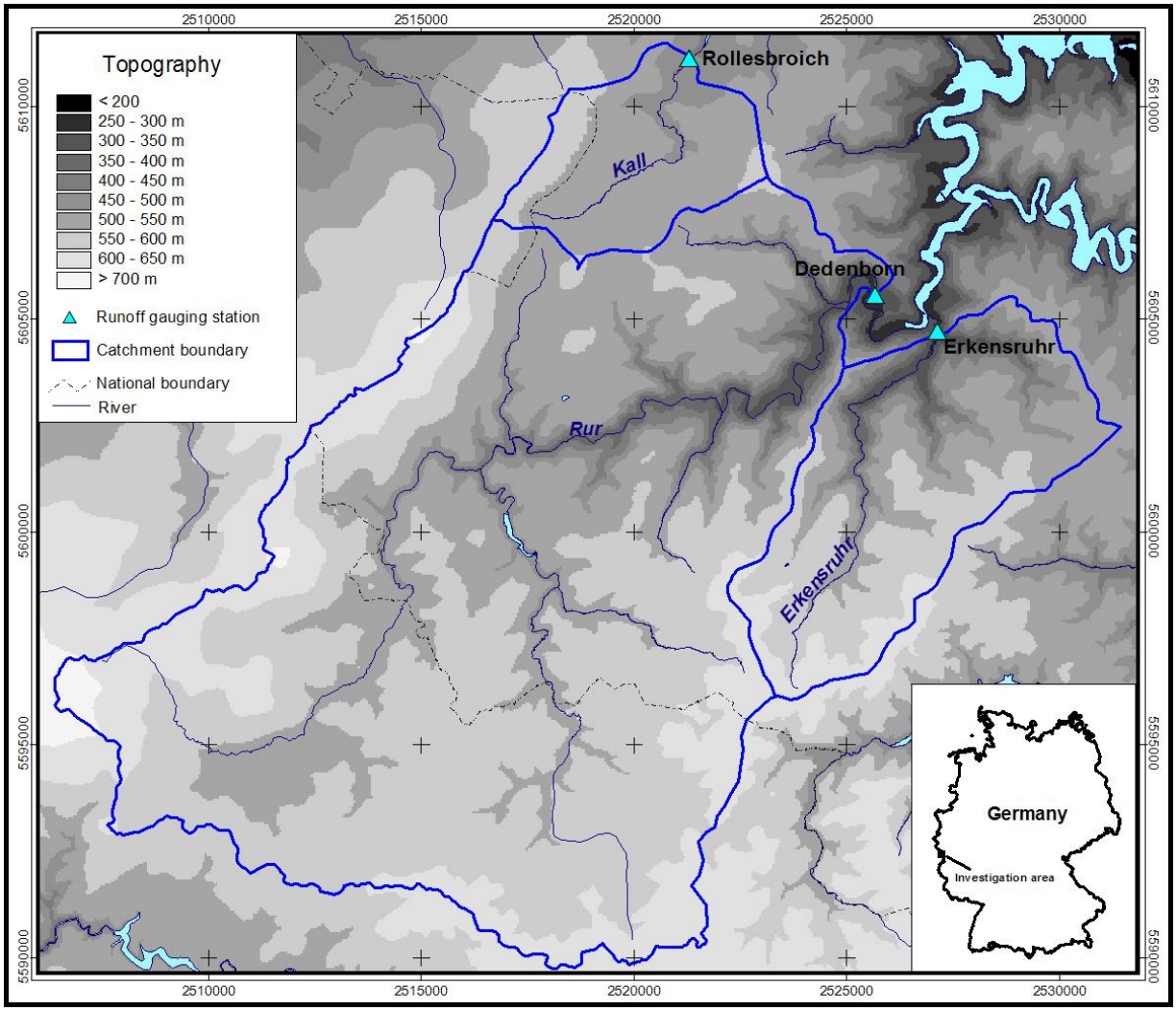


Figure 4: The location of the runoff gauging stations Rollesbroich, Dedenborn and Erkensruhr and the associated catchment areas.

Table 1 gives an overview of the basic catchment characteristics, see Bogena et al., 2005a, Bogena et al., 2005b.

	Upper Rur River		**Kall River**		**Erkensruhr River**	
Gauging station	Dedenborn		Rollesbroich		Erkensruhr	
Area	199.7 km ^{2}		19.03 km ^{2}		41.7 km ^{2}	
Mean elevation	553 m		532 m		526 m	
Mean slope	36.4 $^\circ$		17.3 $^\circ$		32.1 $^\circ$	
Landuse percentage	urban:	2 %	urban:	3 %	urban:	< 1 %
(2001)	arable land:	4 %	arable land:	13 %	arable land:	10 %

	grassland:	35 %	grassland:	45 %	grassland:	16 %
	raised bogs:	29 %				
	forest:	30 %	forest:	39 %	forest:	74 %
Geology (HK100)	Palaeozoic:	89 %	Palaeozoic:	94 %	Palaeozoic:	96 %
	Pleistocene:	1 %			Pleistocene:	1 %
	Holocene:	10 %	Holocene:	6 %	Holocene:	3 %
Annual precipitation	mean:	1042 mm/a	mean:	1200 mm/a	mean:	1081 mm/a
(1979-99)	min:	908 mm/a	min:	1133 mm/a	min:	932 mm/a
	max:	1413 mm/a	max:	1266 mm/a	max:	1219 mm/a
Annual runoff	793 mm/a		733 mm/a		633 mm/a	
(Daily sampling interval)	(1961-2000)		(1982-2000)		(1961-2000)	

Table 1: Basic characteristics of the catchment areas.

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 crestal 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/a in the northeastern part of the investigation area and about 1400 mm/a 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, occupying the largest area fraction (89-96 %). Holocene floodplain deposits and raised bogs are especially occurring 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 characterised by forest, leading to significant lower mean annual runoff compared to the other catchments (100 and 190 mm/a, respectively).

Figure 5 shows the runoff discharge of all catchments for the period of 1982 to 1999 for a time period of one day. 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 are common. The maximum runoff was recorded on 22nd of December 1991 (Rollesbroich: 41.34 mm/day), whereas one of the lowest runoffs from 1990 til 1999 was recorded on 11th of July 1993 (0.0023 mm/day). Since it contains the highest and and one of the lowest runoffs, we selected the period from 1991 to 1993 to present the runoff data in more detail (Figure 6). From Figure 6 it becomes apparent that the runoff from the Kall River (gauging station Rollesbroich) shows more peaks as a result of the smaller catchment area. The runoff from the Upper Rur River (gauging station Dedenborn) is to some extent damped due the presence of a small water reservoir (see Figure 4).

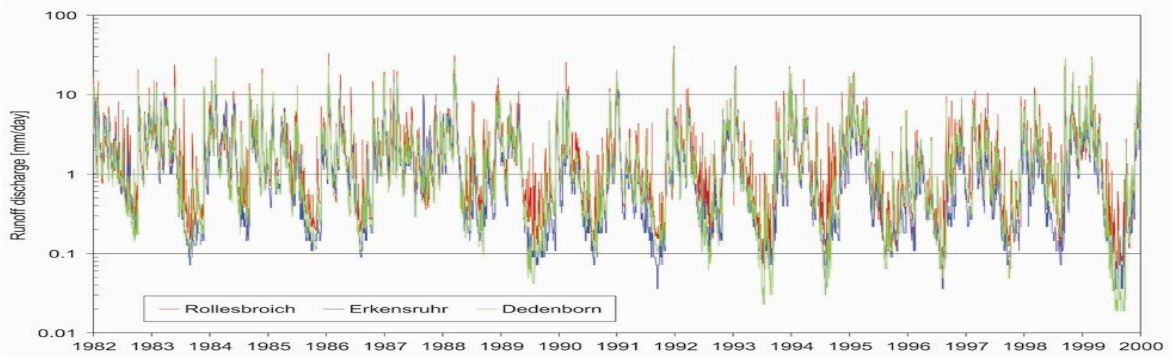


Figure 5: Runoff discharge data from 1982 to 1999 from the three gauging stations.

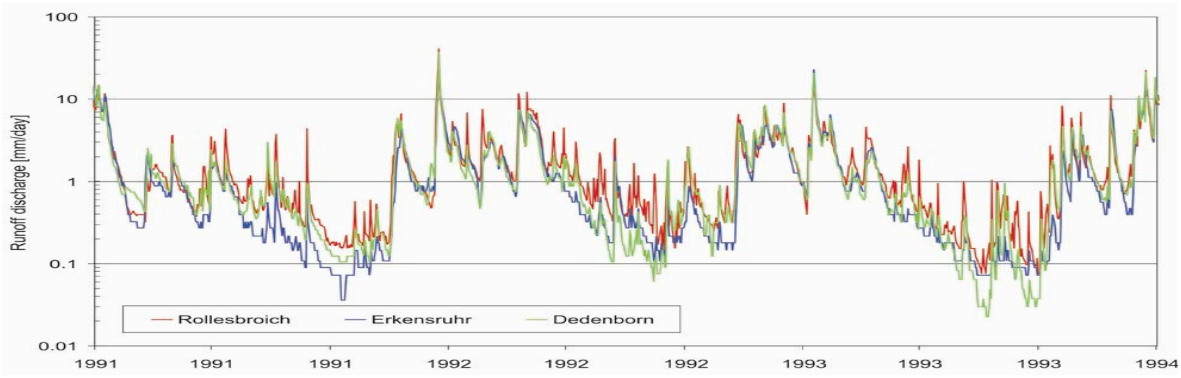


Figure 6: Runoff discharge data from 1991 to 1993 from the three gauging stations.

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 scale; for a description of the currently used soil moisture network and measurement devices, see Bogena et al., 2007, Bogena et al., 2009, Bogena et al., 2010, and Rosenbaum et al., 2010.

3.2 Computation of the Empirical Mode Decomposition (EMD)

For the three data sets and different time ranges, we have computed the corresponding Empirical Mode Decomposition (EMD). The result for the Dedenborn data set in the time range 1990-1999 can be seen in Figure 9 below; all the others can be found in Rudi, 2010.

Figure 7 presents the first 3 IMFs amplitudes for all catchments together with the original runoff discharge data for the 1991-1993 time range. (Note that the IMF amplitudes $a_1(t),\ldots,a_3(t)$ differ from the IMFs $imf_1(t),\ldots,imf_3(t)$ in (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 December 22, 1991, produces very distinct amplitude peaks for imf_1, imf_2 and imf_3 and is clearly visible in all three data sets. Low flow periods go along mostly with low amplitudes for imf_1, imf_2 and imf_3, sometimes with distinct negative peaks.

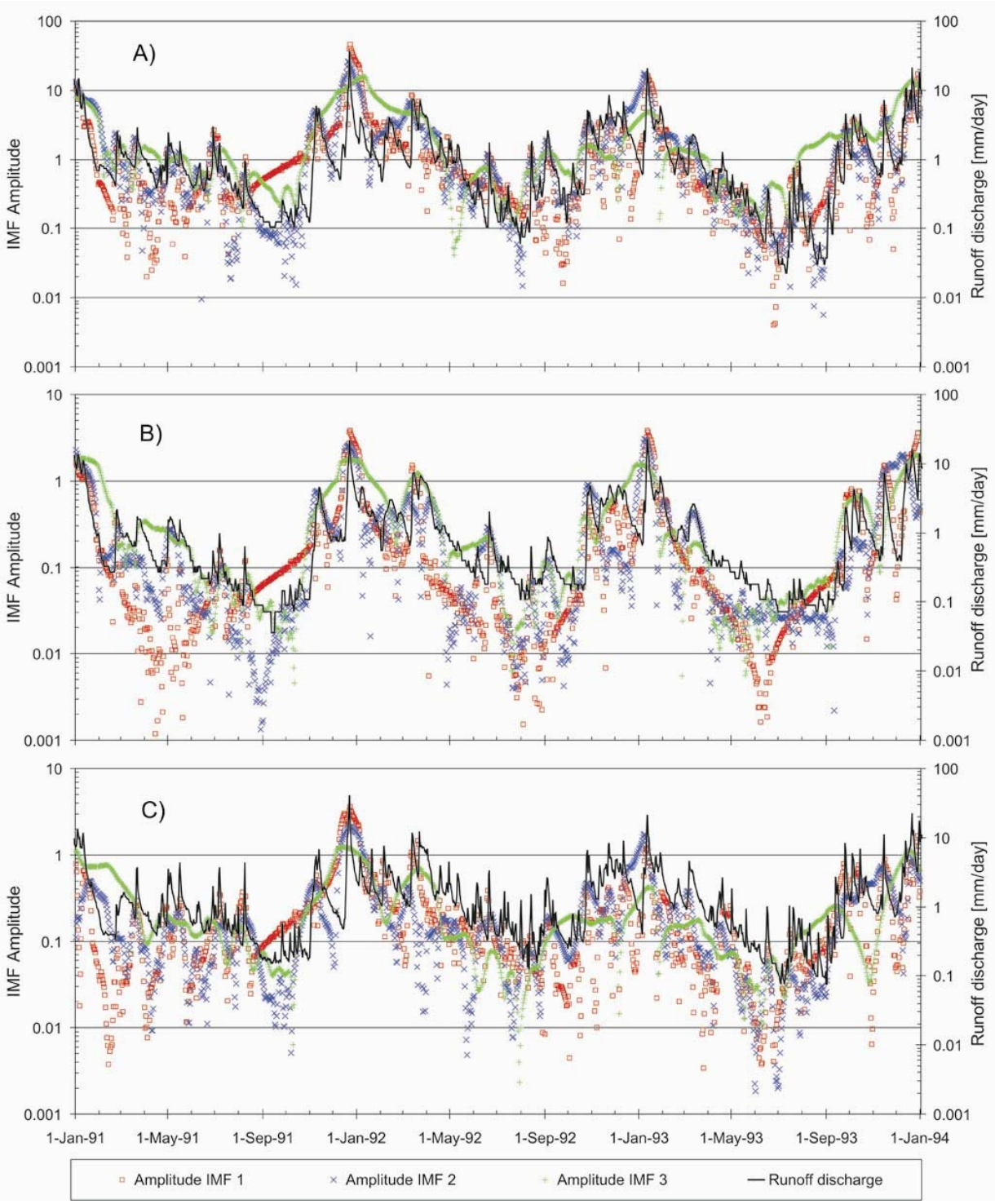


Figure 7: Runoff discharge data from 1991 to 1993 from the three gauging stations compared with IMF Amplitudes 1, 2 and 3 (A) Dedenborn, (B) Rollesbroich and (C) Erkensruhr.

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 locality exhibits similar runoff patterns. However, their absolute values are different, due to higher values of the Dedenborn data (red curve). We also see that the location of extreme points of the three IMF amplitudes, like at December 21, 1991, or at January 9, 1993, coincides. We also see from these amplitudes as well as from the original data seasonal components: decreasing values in spring with lower values during summer and afterwards increasing amplitudes, resulting in the highest amplitudes in December/January of each year.

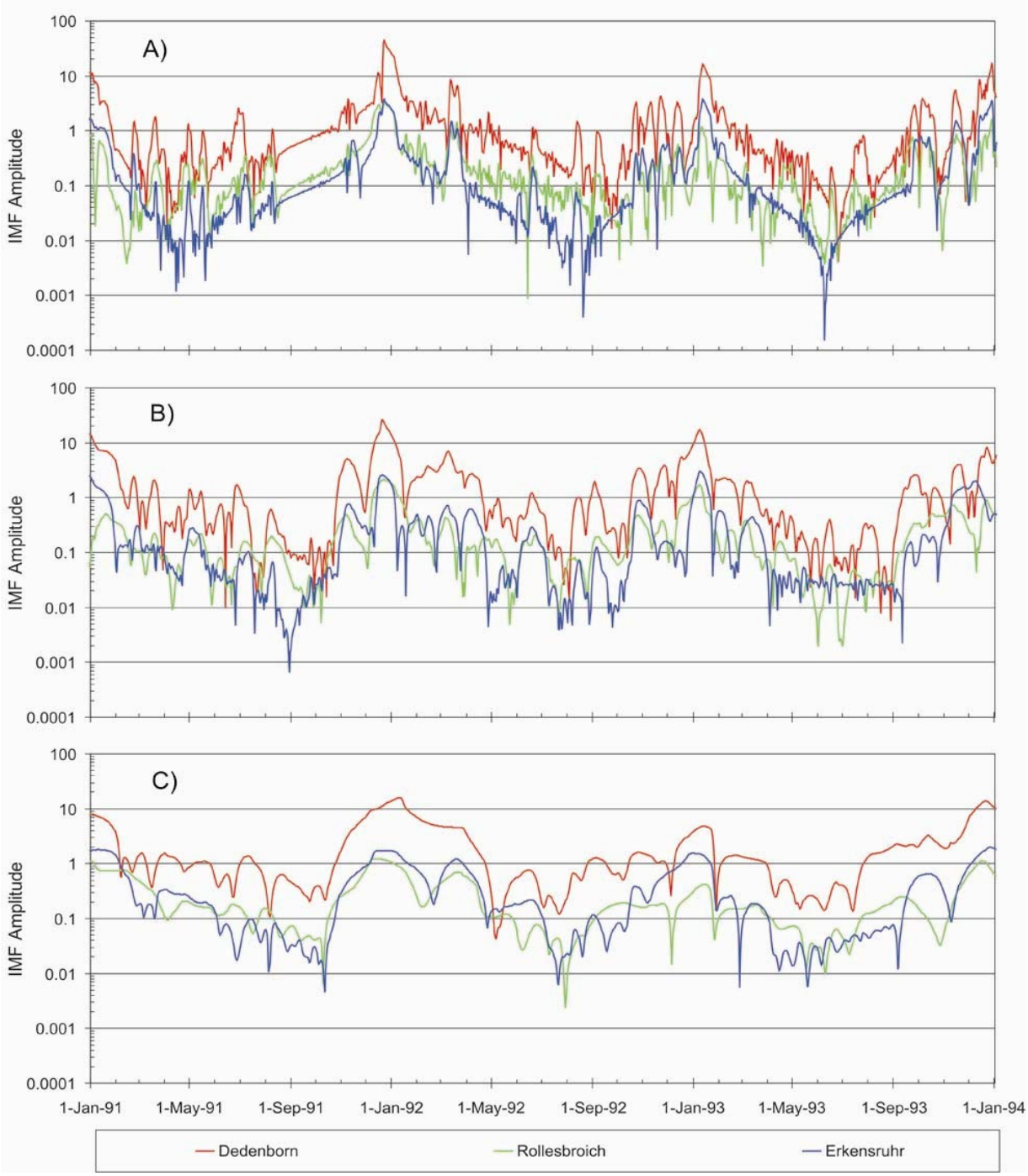


Figure 8: Comparison of IMF amplitudes for all catchments (A) imf_1, (B) imf_2 and (C) imf_3.

Once the EMD is computed like in Figure 9 below, one can define an approximation of the original data as follows. Since the EMD is an exact additive representation of the signal, one can skip the first IMFs in the representation (1) , containing the high frequencies. One can then define a coarse approximation to the original data by setting

$$\tilde{s}(t) := \sum_{j=m}^{n} imf_{j}(t) + r_{n}(t) \tag{14}$$

where $m > 1$ stands for the IMF with the lowest index from which on frequencies should be taken into account. Since this approximation contains all relevant information but much less data since high frequency components are taken out, this representation may be used for further processing within a model for parametrization.

3.3 Different Spectra: Results and Discussion

In this section, we compare the Hilbert energy spectrum as defined in Section 2.2 with other spectra. For this, we have first used the Dedenborn data from 1990-1999 displayed at the top in Figure 9. The corresponding EMD is shown at the bottom.

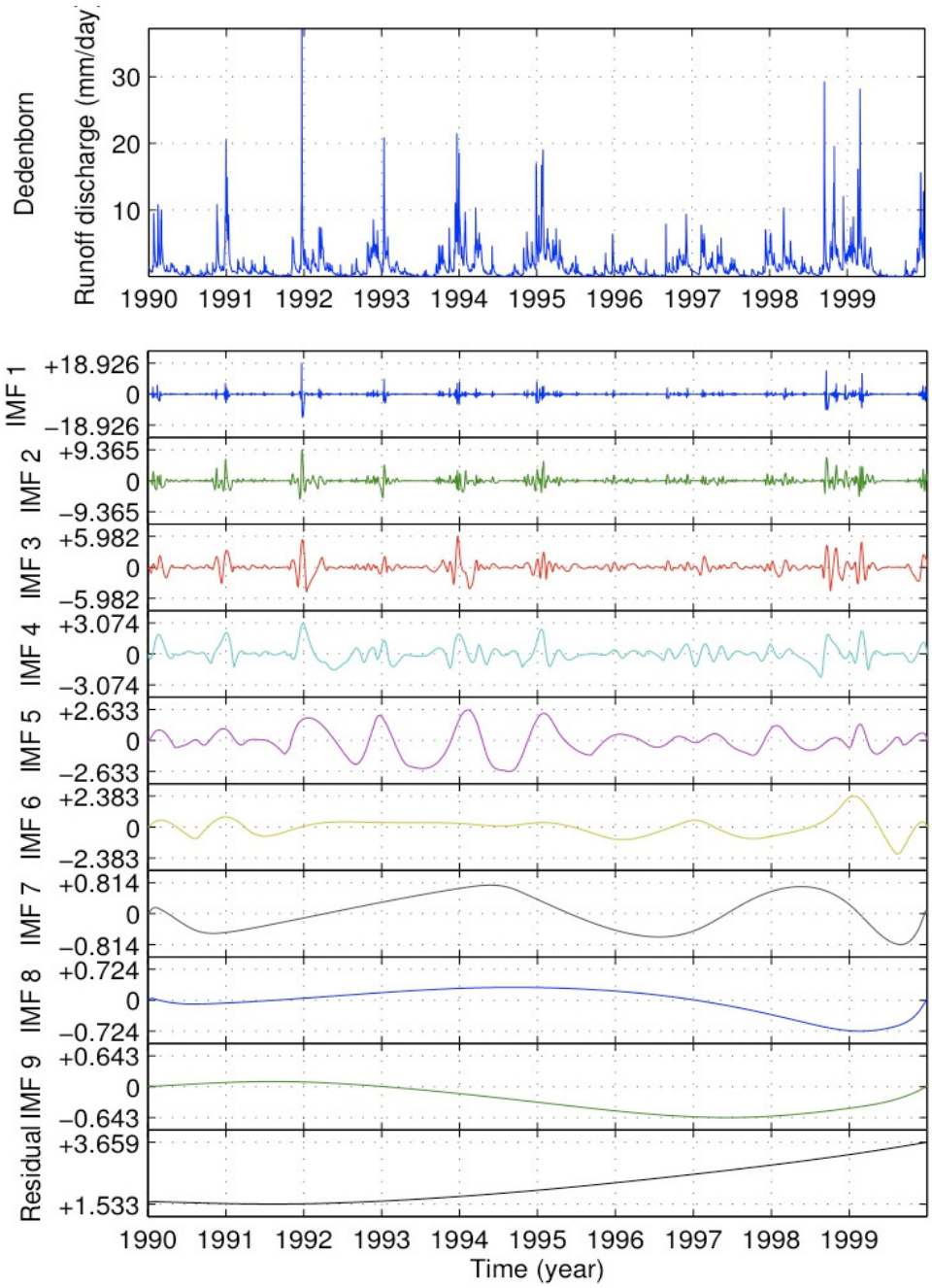


Figure 9: Dedenborn site: original data 1990-1999 (top), EMD (bottom).

The corresponding Hilbert energy spectrum is shown at the top in Figure 10. The ordinate axis displays the time period of the corresponding frequencies in days in 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. So the Hilbert spectrum shows the distribution of the squares of the amplitudes and frequency as functions of time. The two graphics below display the wavelet energy spectrum using the continuous wavelet transform with the Morlet wavelet and the Mexican hat wavelet, respectively. The last graphic is the spectrogram using the short-time (or windowed) Fourier transform (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 a zoom-in of the Dedenborn data into the time period 1991-1993 in Figures 11 and 12.

In the comparison in Figure 10 between the Hilbert energy spectrum and the Fourier spectogram, we see a much more localized behaviour of the Hilbert energy spectrum. Moreover, in comparison with the wavelet spectra the Fourier spectogram even seems to emphasize regions of importance like in 1994 or 1995 where the other spectra are not so dominant. In addition, the Fourier spectogram fails to provide information for periods essentially smaller than 16 days. This effect is even more visible in Figure 12 between beginning of 1992 and 1993, or around February 1993: here the STFT exhibits features which the other spectra do not show so that we can conclude that they may be faulty. For these reasons, we will not consider the Short-time Fourier Transform in the subsequent comparisons. In Figure 10, one can clearly detect boundary effects for the two wavelet transforms due to their periodic nature. This cannot be seen for the Hilbert energy spectrum.

We further zoom into the Dedenborn original data centered at the extreme event of 1991-12-22 with visualization of period lengths, for which we show the Hilbert energy spectrum and two wavelet spectra in Figure 13. The Hilbert energy spectrum shows a very clear and quite unusual nonlinear pattern while the two wavelet spectra just exhibit essentially the location. It is remarkable that both these spectra display high amplitude values only for periods smaller than 64 while the Hilbert energy spectrum shows high amplitudes also for periods up to 256.

Finally, we show in Figure 14 the Hilbert energy spectra for all three sites together for the sample period 1991-1993. 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 which is even stronger than the wavelet spectra; the latter have a tendency to smear out especially for increasing periods over 64 days. Seasonal components with strong peaks in their energies around the turn of the years are most apparent for the Hilbert energy spectrum in Figure 14.

The Hilbert energy spectrum clearly shows certain periodic appearances of similar horizontal distances which can be interpreted as years and seasons. The strength of the amplitudes, visualized by the different colors, indicate different amounts of drainage over time. From this, a time-dependent impact of higher amounts of data can be derived. We have displayed in Figures 10 and 12 a comparison between the Hilbert spectrum and the Fourier spectrogram for zoom-ins of the Dedenborn data. One sees 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 in 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 Figure 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 (R2009b). The Fourier spectrograms employing 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 & Compo, 1998. All programs were written by Johann Rudi. More data and comparisons can be found in Rudi, 2010.

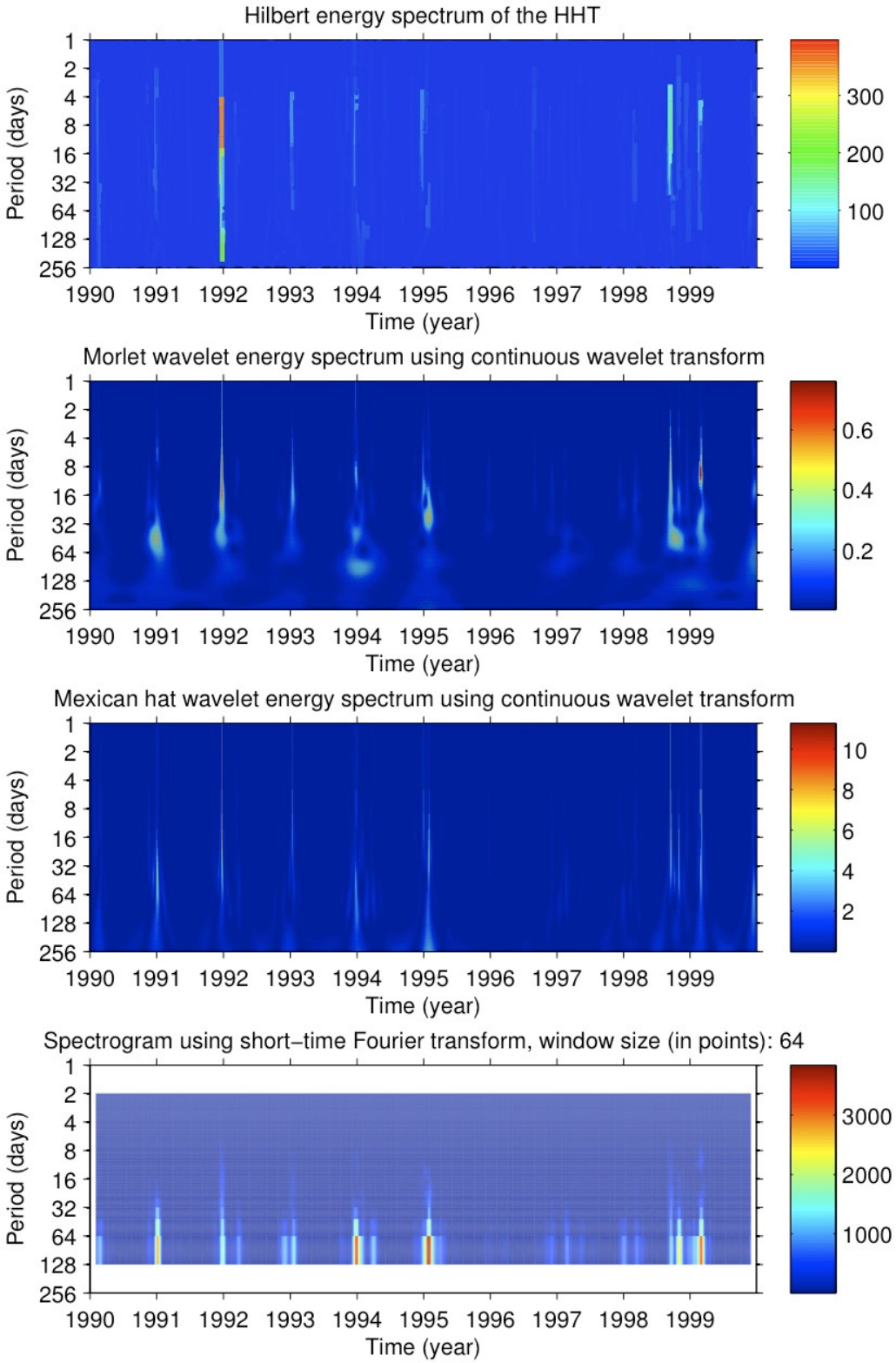


Figure 10: Dedenborn site for data 1990-1999: Hilbert energy spectrum (top), two wavelet spectra (middle) and Fourier spectrogram (bottom).

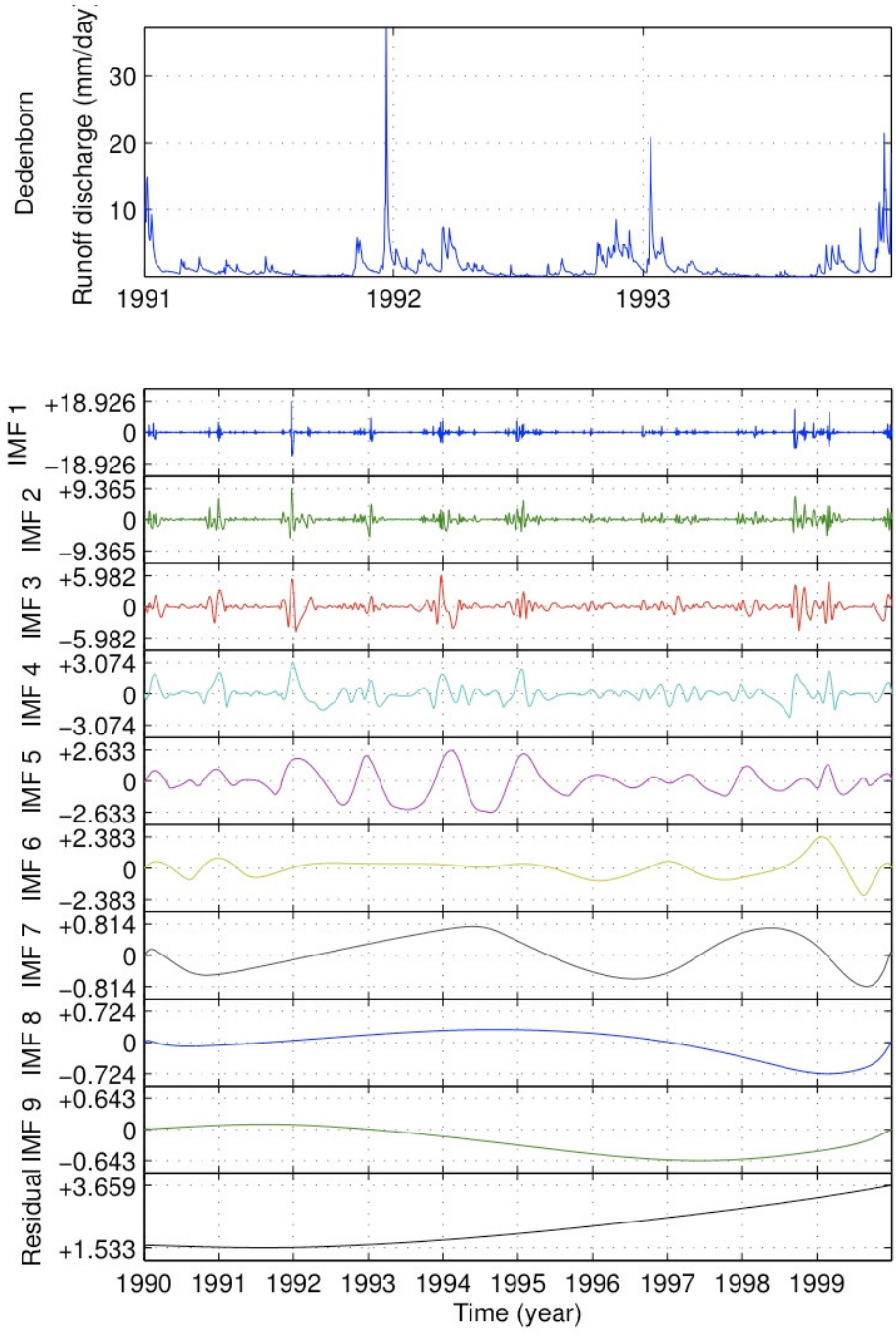


Figure 11: Dedenborn site: original data 1991-1993 (top), EMD (bottom).

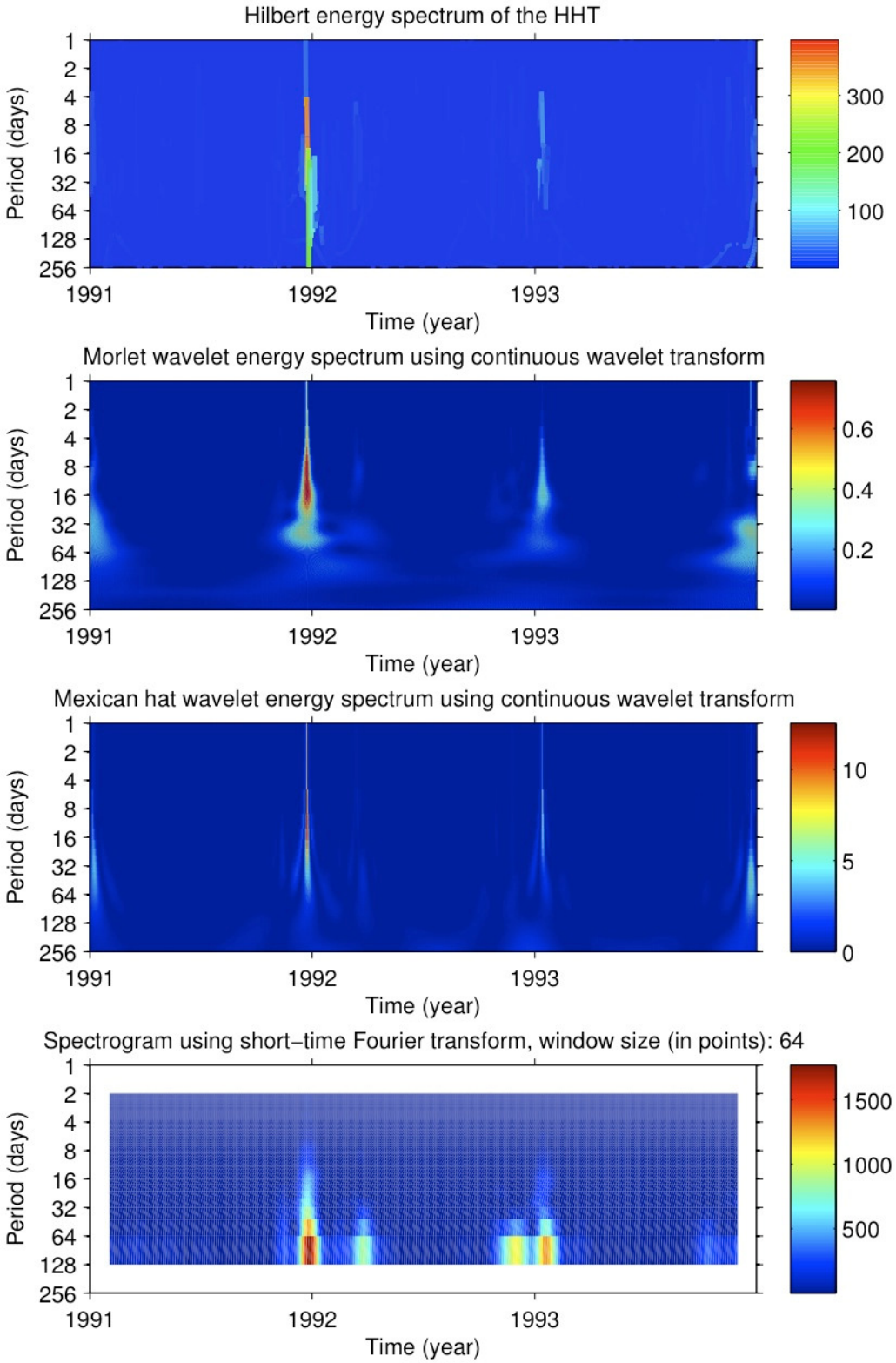


Figure 12: Dedenborn site for data 1991-1993: Hilbert energy spectrum (top), two wavelet spectra (middle) and Fourier spectrogram (bottom).

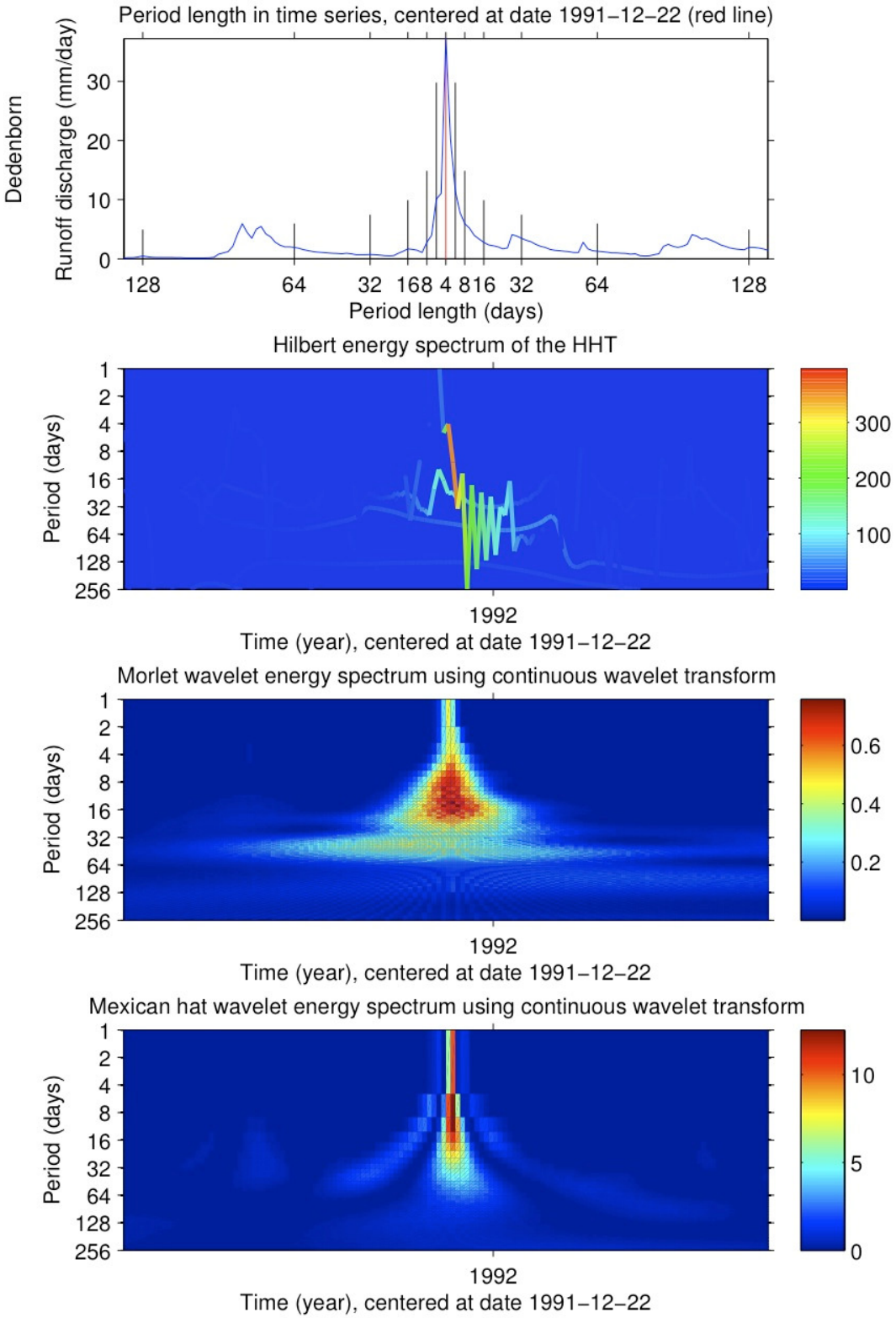


Figure 13: Dedenborn site: original data centered at 1991-12-22 with visualization of period lengths (top), Hilbert energy spectrum and two wavelet spectra.

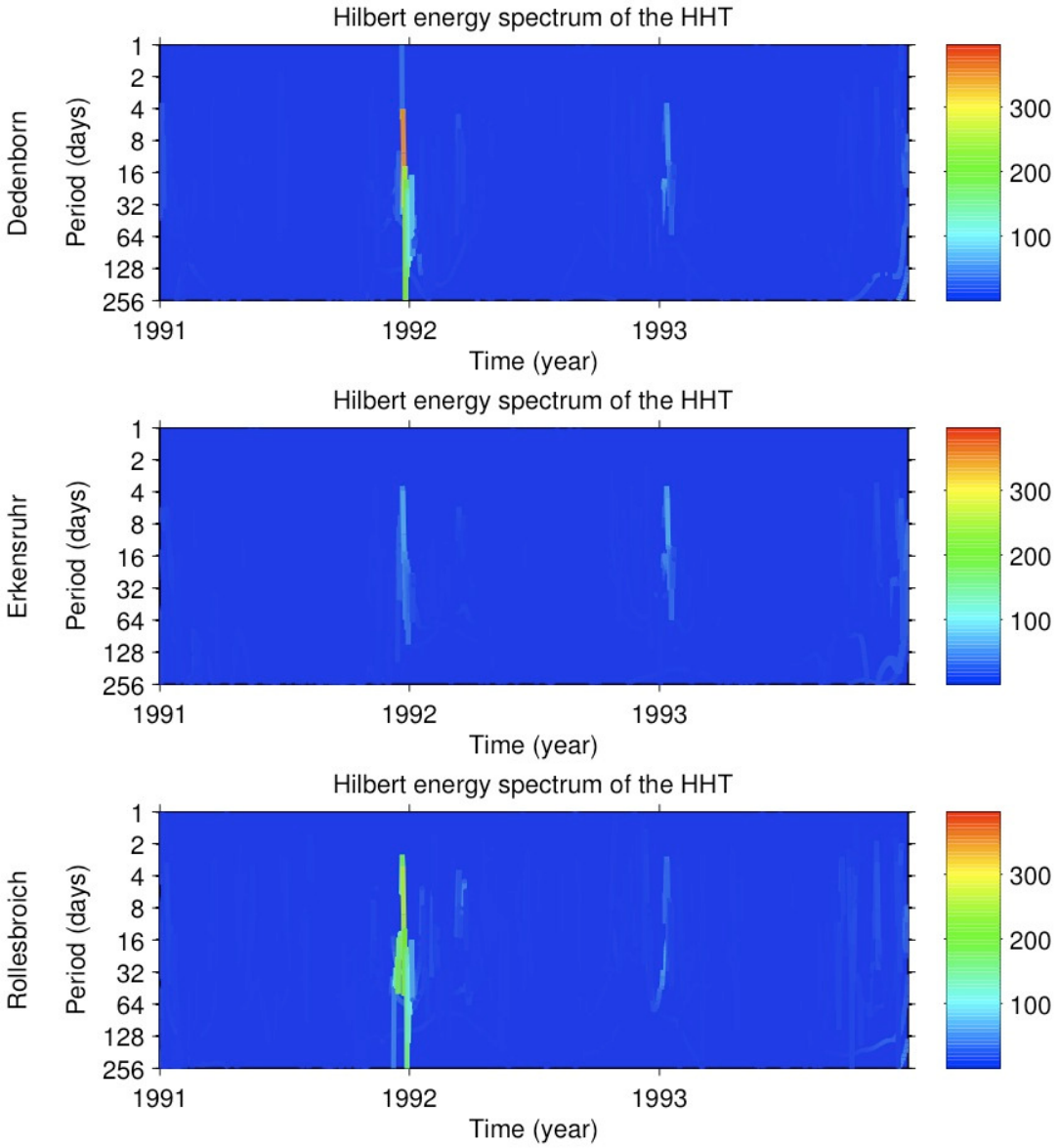


Figure 14: Hilbert spectra with same energy scaling of Dedenborn, Erkensruhr and Rollesbroich from 1991 to 1993

4 Conclusions and Future Work

The HHT method presented here provides a refined analysis of hydrological data when compared to 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 spectograms, the Hilbert energy spectrum exhibits a stronger localized and less smeared-out energy.

Since the EMD provides an additive decomposition of the original data, the different IMFs display different portions of the measured data with time-dependent frequencies which 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 parametrization of multiscale patterns and shapes, the method presented here provides a way to characterize the data set consisting of thousands of data points in terms of very few parameters (years and seasonal components, distribution of their amplitudes and their energies). These may then be employed 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 Section 2.1. Both need the specification of thresholding parameters to terminate the iterations. The difficulty of an appropriate choice of these parameters has been thematized, e.g., in the introductory chapter by Huang in Huang & 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. In order to overcome the parameter choice in the sifting process, an alternative which works with linear functionals for B-Splines instead of upper and lower hulls of the data and which is cheaper to compute has been proposed in Chen et al., 2006, and is 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 since it just requires applications of the Fast Fourier Transform. Towards a theoretical understanding of the overall HHT process, first steps have been undertaken in Sharpley & 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 dimension $n \geq 3$. Different approaches for the 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 in Deléchelle et al., 2005, Damerval et al., 2005, Xu et al., 2006. A systematic comparison for the two-dimensional case which focusses on the quality of the EMD and the numerical performance is given in Koch, 2008, and Jager et al., 2010. Specifically, a new method based on adaptive spline wavelets has been proposed there which 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 is given there.

We applied the EMD methodology to compare long-term daily runoff discharge measurements taken from three river streams 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. Since Huang et al. (2009) also were able to apply the EMD methodology for hydrological time series data we are confident that the EMD will also be usefull 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 characterisation of soil moisture patterns as measured by the sensor network deployed in the framework of the TR32 project (Bogena et al., 2010).

Acknowledgment

We very much want to thank two referees for their substantial and extensive remarks and suggestions which led to a significant improvement of this paper.

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., Kapala, A. and Simmer, C., 2006. Niederschlagsvariabilität während der letzten 100 Jahre in Deutschland. Teil III: Zeitskalen der Variabilität hydrometeorologischer Größen, 7. Deutsche Klimatagung, 9.-11. Oktober, München, extended abstract, 3 pp.

Bogena, H.R., Hake, J.F., Herbst, M., Kunkel, R., Montzka, C., Puetz, T., Vereecken, H. and Wendland, F., 2005. MOSYRUR-Water balance analysis in the Rur basin. Schriften des Forschungszentrums Jülich, Reihe Umwelt 52, 155 p.

Bogena, H., Kunkel, R., Schöbel, T., Schrey, H.P. and Wendland, F., 2005. Distributed modelling of groundwater recharge at the macroscale, Ecological Modelling 187: 15-26.

Bogena, H.R., Huisman, J.A., Oberdörster, C. and Vereecken, H., 2007. Evaluation of a low-cost soil water content sensor for wireless network applications, Journal of Hydrology 344, 32â€“42.

Bogena, H.R., Huisman, J.A., Meier, H., Rosenbaum, U. and Weuthen, A., 2009. Hybrid wireless underground sensor networks - Quantification of signal attenuation due to soil adsorption. Vadose Zone J. 8:755-761.

Bogena H.R., Herbst, M., Huisman, J.A., Rosenbaum, U., Weuthen, A. and H. Vereecken, 2010. Potential of wireless sensor networks for measuring soil water content variability. Accepted for publication in Vadose Zone J., this issue.

Castaño Diez, D. and Kunoth, A., 2006. Robust regression of scattered data with adaptive spline-wavelets, IEEE Trans. Image Proc.15 (6), 1621-1632.

Castaño Diez, D., Jager, G. and Kunoth, A., 2009. Multiscale analysis of multivariate data, in: Proc. of 18th Intern. Conf. on Appl. of Computer Science & Mathematics in Architecture & Civil Engineering, K. Gürlebeck and C. Könke (Eds.).

Chen, Q., Huang, N.E., Riemenschneider, S. and Xu, Y., 2006. A B-spline approach for empirical mode decomposition, Adv. Comput. Math. 24, 171-195.

Cohen, L., 1994. Time-frequency analysis, Prentice Hall PTR, 320 p.

Damerval, Ch., Meignen, S. and Perrier, V., 2005. A fast algorithm for bidimensional EMD, IEEE Signal Process. Lett. 12(10), 701--704.

Deléchelle, E., Lemoine, J. and Niang, O., 2005. Empirical mode decomposition: An analytical approach for sifting process, IEEE Signal Process. Lett. 12(11), 764--767.

DeVore, R.A. and Kunoth, A. (Eds.), 2009. Multiscale, Nonlinear and Adaptive Approximation, Springer.

Felsberg, M. and Sommer, G., 2000. The multidimensional isotropic generalization of quadrature filters in geometric algebra. In: Lecture Notes in Computer Science 1888, 175-185.

Felsberg, M. and Sommer, G., 2001. The monogenic signal, IEEE Transactions on Signal Proc. 49, 3136--3144.

Flandrin, P., 1999. Time-frequency/time-scale analysis, Academic Press Inc., 386 p.

Flandrin, P. and Gonçalvès, P., 2004. Empirical mode decompositions as data-driven wavelet-like expansions, Int. J. Wavelets Multiresolut. Inform. Process. 2, 477-496.

Flandrin, P., Rilling, G. and Gonçalvès, P., 2004. Empirical mode decompositions as a filter bank, IEEE Signal Process. Lett. 11, 112--114.

Grayson, R. and Blöschl, G. (Eds.), 2001. Spatial patterns in catchment hydrology --- observations and modelling. Cambridge University Press. 404 p.

Hahn, S., 1995. Hilbert Transforms in Signal Processing, Artech House, 442 p.

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zhang, Q., Yen, N.-C., Tung, C.C. and Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. A 454:903-995.

Huang, N.E. and Shen, S.S.P., 2005. Hilbert-Huang Transform and its Applications, World Scientific Publishing.

Huang, Y.X., Schmitt, F.G., Lu, Z.M. and Liu, Y.L., 2009. Analysis of daily river flow

fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analyis, Journal of Hydrology 373, 103-111.

Jager, G., Koch, R., Kunoth, A. and Pabel, R., 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, accepted for publication in Advances in Adaptive Data Analysis (AADA).

Kang, S. and Lin, H., 2007. Wavelet analysis of hydrological and water quality signals in an agricultural watershed, Journal of Hydrology 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), Diplomarbeit, Institut für Numerische Simulation, Universität Bonn.

Linderhed, A., 2004. Image compression based on empirical mode decomposition, In: Proceedings of SSAB 04 symposium on image analysis, 110--113.

Rilling, G., Flandrin, P. and Gonçalvès, P., 2003. On empirical mode decompositions and its algorithms, in: IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP--03.

Rosenbaum, U., Huisman, J.A., Weuthen, A., Vereecken, H. and Bogena, H.R., 2010. Quantification of sensor-to-sensor variability of the ECH2O EC-5, TE and 5TE sensors in dielectric liquids. Vadose Zone J. 9: 181-186.

Rudi, J., 2010. Empirical Mode Decomposition via adaptiver Wavelet-Approximation (in German), Diploma Thesis in preparation, Institut für Mathematik, Universität Paderborn.

Schaefli, B. and Zehe, E., 2009. Hydrological model performance and parameter estimation in the wavelet-domain, Hydrol. Earth Syst. Sci., 13, 1921-1936.

Sharpley, R.C. and Vatchev, V., 2006. Analysis of the intrinsic mode functions, Constr. Approx. 24 (1), 14--47.

Si, B.C., 2008. Spatial scaling analyses of soil physical properties: A review of spectral and wavelet methods, Vadose Zone J. 7(2): 547--562.

Titchmarsh, E.C., 1950. Introduction into the Theory of Fourier Integral, Oxford University Press.

Torrence, C. and Compo, G.P., 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79 (1): 61-78.

Western, A. W., Zhou, S.-L., Grayson, R.B., McMahon, T.A., Blöschl, G. and Wilson, D.J., 2004. Spatial correlation of soil moisture in small catchments and its relationship to dominant

spatial hydrological processes. Journal of Hydrology 286:113-134.

Xu, Y., Liu, B., Liu, J. and Riemenschneider, S., 2006. Two-dimensional empirical mode decomposition by finite elements, Proc. R. Soc. Lond. A, 462, 3081--3096.