

## Characterization of a cavern conduit system in Vietnam by time series correlation, cross-spectrum and wavelet analyses

VU THANH TAM<sup>1</sup>, FLORIMOND DE SMEDT<sup>2</sup>,  
OKKE BATELAAN<sup>2</sup> & ALAIN DASSARGUES<sup>3</sup>

<sup>1</sup> *Research Institute of Geology and Mineral Resources, Ministry of Natural Resources and Environment, Km 9 Nguyen Trai Street, Thanh Xuan, Hanoi, Vietnam*

<sup>2</sup> *Department of Hydrology and Hydraulic Engineering, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium*  
[batelaan@vub.ac.be](mailto:batelaan@vub.ac.be)

<sup>3</sup> *Département Géoresources, Géotechnologies et Matériaux de Construction (GEOMAC), Université de Liège, B.52 Sart-Tilman, B-4000 Liege, Belgium*  
and

*Hydrogeology and Engineering Geology, Department of Geography-Geology, K.U. Leuven, Redingenstraat 16, B-3000 Leuven, Belgium*

**Abstract** Time series analyses are applied to characterize the transient flow regimes of the Nam La cavern conduit, northwest Vietnam. The conduit transforms the input signal to an output signal, and the degree of transformation provides information on the nature of the flow system. The input for the analysis is net precipitation and the flow hydrograph at the cave entrance, while the output series is the flow hydrograph at the resurgence. Cross-correlation and cross-spectrum analysis are used to investigate the stationarity and linearity of the input–output transformation of the system, resulting in hydrodynamic properties such as system memory, response time, and mean delay between input and output. It is shown that during high flow periods, the flow in the conduit is pressurized. Consequently, the linear input–output assumption holds only for low flows. To highlight the hydrodynamics of the cavern conduit for the high flow periods, wavelet spectrum and wavelet cross-spectrum analyses are applied.

**Key words** karst; carbonate rocks; cross-correlation; wavelet analysis; Vietnam

### Caractérisation du système du conduit d'une grotte au Vietnam par des analyses corrélatoires, spectrales-croisées et en ondelettes de séries temporelles

**Résumé** Des analyses de séries temporelles sont réalisées pour caractériser les écoulements en régime transitoire dans le conduit de la grotte Nam La, au Vietnam. Le conduit transforme le signal d'entrée en un signal de sortie. Le degré de cette transformation fournit des informations sur la nature du système d'écoulement. Les données d'entrée sont la pluviométrie nette et l'hydrogramme à l'entrée de la grotte, alors que le signal de sortie est l'hydrogramme au niveau de la resurgence. L'analyse de corrélation croisée et l'analyse spectrale croisée sont employées pour étudier la stationnarité et la linéarité de la transformation entrée-sortie du système, en renseignant sur certaines propriétés hydrodynamiques comme la mémoire du système, le temps de réponse, et le délai moyen entre entrée et sortie. Il apparaît que, pendant les périodes de hautes eaux, l'écoulement dans le conduit est sous pression. En conséquence, l'hypothèse de linéarité entrée-sortie n'est valable que pour les basses eaux. Afin de mettre en évidence l'hydrodynamique du conduit pour les périodes de hautes eaux, des analyses spectrales et spectrales croisées en ondelettes sont réalisées.

**Mots clefs** karst; roches carbonatées; corrélation croisée; ondelette; Vietnam

## INTRODUCTION AND OBJECTIVES

Within a karst carbonate aquifer the groundwater is mostly stored in the bedrock matrix and in the fractures, while the cavern conduits play the role of collectors and conveyers that eventually drain almost all of the karst basin water to a resurgence. Two notions are of particular concern regarding the hydrogeological functioning of cavernous conduits: (a) the system acts as a low resistance drain, so that the flow field in the surrounding matrix and fractures is directed toward the conduit rather than toward groundwater discharge zones on the surface (White, 1969; Smart & Hobbs, 1986; Ford & Williams, 1989; Palmer, 1991), and (b) the conduit flow is often turbulent and can display a free surface or be under pressure. The free surface flow mostly occurs during low-flow conditions when groundwater recharged from diffuse infiltration and from storage in the fractures and matrix porosity drains toward the conduit system. However, during storms the conduit flow may flood to the roof and establish a substantial pressure head. Hence, during intervals of storm flow, water can move from the conduit into the surrounding fractures and matrix. The coupling of these systems, combined with the intrinsic hydraulic conductivity of the matrix and fractures, determines the rate of movement of groundwater into and out of storage and also the baseflow of karst springs. These specific features should be identified and considered in the estimation of intrinsic storage of cavernous conduit systems.

Time series analyses, as developed by Jenkins & Watts (1968), have been widely applied in hydrology (Delleur & Rao, 1971; Priestley, 1981, 1988; Dreiss, 1983; Mangin, 1984; Erskine & Papaioannou, 1997; Molénat *et al.*, 1999; Bayazit *et al.*, 2001). Time series analyses are especially useful in karstic environments because they are easy to implement and can provide basic understanding of the aquifer of interest. A majority of applications in this field is based on black-box systems in which the aquifer is considered as a filter that transforms the input signal into an output signal (Padilla & Pulido-Bosch, 1995; Estrela & Sahuquillo, 1997; Larocque *et al.*, 1998, 2000; Long & Derickson, 1999; Amraoui *et al.*, 2003). These applications are based on the linear and time-invariant system assumption, which is often not satisfactory for karstic basins (Labat *et al.*, 2000a,b). Considering nonstationarity being as important as nonlinearity in the formation of the basin response, a more realistic modelling approach based on wavelet analysis of the input and output signals was suggested, which deals better with the relatively infrequent events of the time series.

In this study, time series analysis is used to retrieve fundamental parameters such as system memory, response time, mean delay, carrying capacity, and conduit/fracture coupling effect of the Nam La cavern conduit system in northwest Vietnam. The sinkhole stream discharge and the net precipitation as input and resurgence discharge as output are considered as two autocorrelated and cross-correlated stochastic processes, while the conduit is considered as a filter that transforms the input into the output. The system storage is conceptualized to include not only the conduit storage itself (which is assumed not significant) but also the storage in the fractures and rock matrix around the conduit. In the first part of this study, analyses are made of the classical correlation and cross-spectral functions that are applied to the time series to retrieve the desired parameters, assuming that the input–output relationship is linear and stationary. In the second part, the wavelet transform techniques are applied to the time series to achieve a representation of transient phenomena occurring at different time scales. This allows determination of the transition between linear and nonlinear

states of the system. The system conveying capacity is therefore defined as the outflow at the resurgence when the input–output relationship is nonlinear and the inflow exceeds the outflow.

**THEORY**

Many aspects of the random process theory that are used here have been developed in the classical texts by Jenkins & Watts (1968), Yevjevich (1972), Oppenheim & Schafer (1975), Box & Jenkins (1976), Liu (1995), Chui & Montefusco (1994), and others. These texts develop several aspects of the identification and simulation of joint random processes based on the concepts of cross-correlation (in the time domain), cross-spectrum (in the frequency domain), and cross-wavelet spectrum (in both time and scale/frequency domains). A short introduction of these theories is presented here to facilitate comprehension of the study result.

**Cross-correlation and cross-spectrum**

Let there be two discretized chronological series: the input,  $x_i$  ( $x_1, x_2, \dots, x_n$ ), is the cause of the output,  $y_i$  ( $y_1, y_2, \dots, y_n$ ), where  $n$  is the total of number of data pairs available. The autocorrelation function or correlogram quantifies the linear dependency of successive values of the time series  $x_i$  and is defined as:

$$r_x(k) = \frac{C_x(k)}{C_x(0)} \tag{1}$$

where  $k$  is the time lag ( $k = 0, m$ ), with  $m$  representing system memory length, and  $C_x(k)$  is the autocovariance given by:

$$C_x(k) = \frac{1}{n} \sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x}) \tag{2}$$

with  $\bar{x}$  the mean of the series  $x_i, i = 1, \dots, n$ . The correlogram  $r_x(k)$  also outlines the memory of the system, i.e. if an event has a long-term influence on the time series, the slope of the autocorrelation function decreases slowly as a function of the time lag  $k$ . The cross-correlation function (CCF), represents the impulse response of the system under investigation when the input series is considered to be white noise and is described by the following expressions:

$$r_{+k} = r_{xy}(k) = \frac{C_{xy}(k)}{\sigma_x \sigma_y} \tag{3}$$

$$r_{-k} = r_{yx}(k) = \frac{C_{yx}(k)}{\sigma_x \sigma_y} \tag{4}$$

where  $C_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y})$  (5)

$$C_{yx}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \bar{y})(x_{t+k} - \bar{x}) \tag{6}$$

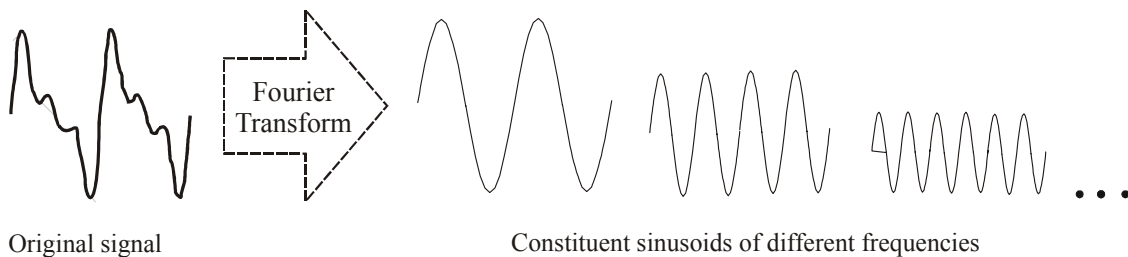
$$\sigma_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \sigma_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

with  $\bar{y}$  the mean of the series  $y_i$ ,  $i = 1, \dots, n$ . The cross-correlation function is not symmetrical:  $r_{xy}(k) \neq r_{yx}(k)$ , i.e. if  $r_{xy}(k) > 0$  for  $k > 0$ , series  $x_i$  influences series  $y_i$ , while if  $r_{xy}(k) > 0$  for  $k < 0$ , series  $y_i$  influences series  $x$ . The delay time, defined as the time lag between  $k = 0$  and the maximum of  $r_{xy}(k)$ , gives an estimation of the peak impulse response time of the system.

Spectrum (Fourier) analysis decomposes a complex time series with cyclical components into underlying sinusoidal functions of particular wavelength (or equivalently frequency,  $f$ ) (Fig. 1). Components of high frequencies of a streamflow series can be considered as high discharge and rapidly variable flows that often occur during extensive rainy periods or flood events. By identifying the important underlying cyclical components, one learns something about the phenomenon of interest. Change from a time mode to a frequency mode is done through a Fourier transform:

$$F(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-i\omega t} dt \quad (8)$$

resulting in a Fourier function  $F(\omega)$ , which when multiplied by a sinusoid of frequency  $\omega = 2\pi f/f_s$  ( $f_s$  is the sampling frequency), yield the constituent sinusoidal components of the original signal.



**Fig. 1** Decomposition of a signal into constituent sinusoidal components by Fourier transform.

The spectral density function,  $S_x(f)$ , of the time series  $x_i$ , corresponds to the discrete-time Fourier transform of the autocorrelation function:

$$S_x(f) = \frac{1}{2} \left[ 1 + \sum_{k=1}^m D(k) r_x(k) \cos(2\pi f k) \right] \quad (9)$$

where  $D(k)$  is a weighted filter necessary to overcome the problem of occasional chaotic spikes in the time series. Similar, the complex cross-spectral density function,  $S_{xy}(f)$ , corresponds to the change from time to frequency mode for the cross-correlation function:

$$S_{xy}(f) = h_{xy}(f) - i\lambda_{xy}(f) \quad (10)$$

of which the real part is the cospectrum:

$$h_{xy}(f) = 2 \left[ r_{xy}(0) + \sum_1^m (r_{xy}(k) + r_{yx}(k)) D(k) \cos(2\pi fk) \right] \quad (11)$$

and the imaginary part is the quadrature spectrum:

$$\lambda_{xy}(f) = 2 \left[ r_{xy}(0) + \sum_1^m (r_{xy}(k) + r_{yx}(k)) D(k) \sin(2\pi fk) \right] \quad (12)$$

The interpretation of these functions through the identification of the different peaks representing periodical phenomena, leads to the characterization of the system and gives an indication of the length of the impulse response of the system.

The magnitude of the cross-spectral density function is called the cross-amplitude function (CAF),  $\alpha_{xy}(f)$ :

$$\alpha_{xy}(f) = \sqrt{h_{xy}^2(f) + \lambda_{xy}^2(f)} \quad (13)$$

The CAF can be interpreted as a measure of covariance between the respective frequency components in the two series, identifying the way in which the input signal has been modified by the system (Padilla & Pulido-Bosch, 1995). The CAF can also be associated with the duration of the impulse response function.

The phase shift or phase function (PHF),  $\Phi_{xy}(f)$ , estimates the extent to which each frequency component of one series leads the other (i.e. the delay of output with respect to input):

$$\phi_{xy}(f) = \arctan \frac{\lambda_{xy}(f)}{h_{xy}(f)} \quad (14)$$

The mean delay in the time mode,  $d$ , can be obtained by the slope of the line of best fit to:

$$d \approx \frac{\phi_{xy}(f)}{2\pi f} \quad (15)$$

The cross-coherency function (COF),  $CO_{xy}(f)$ , is given by:

$$CO_{xy}(f) = \frac{\alpha_{xy}(f)}{\sqrt{S_x(f)S_y(f)}} \quad (16)$$

and can be interpreted as the squared correlation between the cyclical components in the two series at the respective frequency. It shows whether variations in the output series respond to the same type of variations in the input series, and thereby indicates the linearity of the input–output relationship. A system is linear,  $COF \approx 1$ , when a change in the input series creates a proportional change in the output series. If the COF shows a very chaotic pattern with values far beyond unity, nonlinearity very likely exists in the system and other factors should be considered in the definition of the system. However, the coherency value should not be interpreted by itself, but in conjunction with other functions.

The cross-gain function (CGF),  $G_{xy}(f)$ , of the output series  $y_i$  over the input series  $x_i$  is:

$$G_{xy}(f) = \frac{\alpha_{xy}(f)}{\sqrt{S_x(f)}} \quad (17)$$

and can be interpreted as the standard least squares regression coefficients for the respective frequencies. It expresses an amplification ( $>1$ ) or attenuation ( $<1$ ) of the output signal in comparison with the input signal. In karstic environments this can be respectively related to the storage of water during a high flow periods and the release of water during low flow periods (Laroque *et al.*, 2000).

### Wavelet spectrum and cross-wavelet spectrum

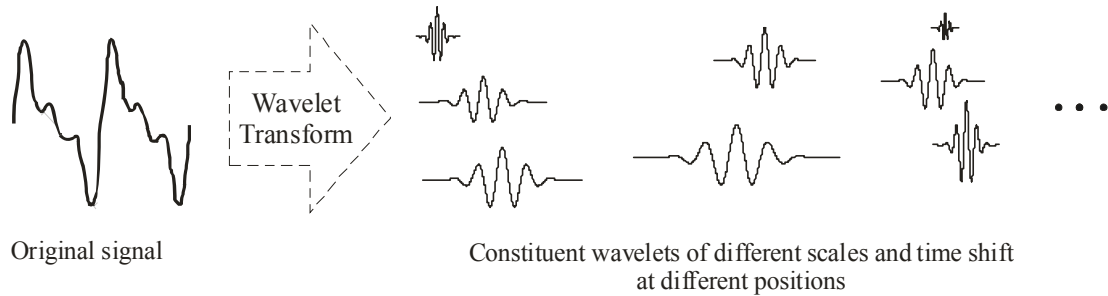
The drawback of Fourier analysis is that temporal information is lost and it is impossible to tell when a particular event took place. Wavelets can be used to perform local analysis of sharply changing time series data as they tend to be irregular and asymmetric. Similar to Fourier analysis, wavelet transform allows breaking up of a signal into shifted (delayed or hastened) and scaled (stretched or compressed) versions of the original wavelet (Fig. 2(a)). Therefore, wavelet analysis is complementary to Fourier analysis because it is capable of representing time series data with fast local variations that Fourier analysis misses. The continuous wavelet transform is defined as:

$$C_x(a, \tau) = \int_{-\infty}^{+\infty} x(t) \psi_{a, \tau}^*(t) dt \quad \text{with} \quad \psi_{a, \tau}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t - \tau}{a}\right) \quad (18)$$

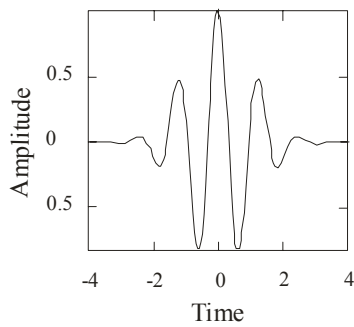
where  $C_x(a, \tau)$  are the wavelet transform coefficients,  $a$  is a wavelet scale factor (expansion if  $a > 1$  or contraction if  $a < 1$ ),  $\tau$  is the wavelet temporal translation or phase shift, and \* means the complex conjugate of a function. The function  $\psi(t)$ , which can be real or complex, plays the role of a convolution-kernel and is called a wavelet. For detection of abrupt changes in time series the most important properties of a wavelet are the regularity and the number of vanishing moments. Figure 2(b) and (c) shows the time representation of the Morlet and the fourth-order Daubechies wavelet, which are used in this study (Chui, 1992; Kaiser, 1994; Torrence & Crampo, 1998). Coifman *et al.* (1992) and Chui & Montefusco (1994) showed that the scale factor is inversely related to the radian frequency; hence, the lower the scale, the more compressed the wavelet and the higher the frequency, and *vice versa*. A plot of wavelet coefficients of a time-scale diagram will help to identify events of particular frequency at a particular time. This time-scale localization feature is the greatest advantage of wavelet analysis over its Fourier counterpart. Thus, the wavelet transform of a signal  $x(t)$  is the family  $C_x(a, \tau)$ , which depends on two indexes  $a$  and  $\tau$ . Intuitively, the wavelet decomposition could be thought of as calculation of a “resemblance” index, i.e. the wavelet transform coefficient, between the signal and the wavelet located at position  $\tau$  and of scale  $a$ . If the index is large the resemblance is strong, otherwise it is weak.

With the wavelet transform, a continuous time signal  $x(t)$  can be reconstructed as:

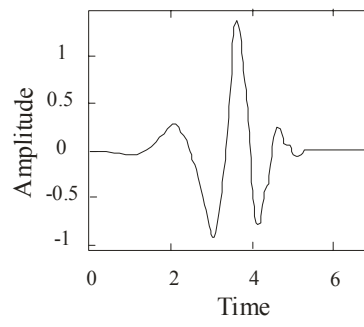
$$x(t) = \frac{1}{K_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} C_x(a, \tau) \psi_{a, \tau}(t) \frac{dad\tau}{a^2} \quad (19)$$



(a)



(b)



(c)

**Fig. 2** (a) Decomposition of a signal into constituent wavelet components by Morlet wavelet transform, and time representation of (b) Morlet wavelet and (c) fourth-order Daubechies wavelet.

where  $K_\psi$  is a constant depending only on the chosen wavelet function. By analogy with the Fourier analysis, one can define a wavelet spectrum of a continuous-time signal and a wavelet cross-spectrum between two signals. Furthermore, for hydrological signals that are often discrete in time one can also use discrete wavelet transform for a non-redundant discrete time-scale representation. In the discrete wavelet transform, both the time shift,  $\tau$ , and the scale domain,  $a$ , are discretized (Daubechies, 1992).

In multilevel wavelet analysis (Mallat, 1989), the signal  $x(i)$  is split into an approximation,  $A_1^x$ , and a detail,  $D_1^x$ . The approximation  $A_1^x$  is then itself split into a second-level approximation  $A_2^x$  and detail  $D_2^x$ , and the process is repeated until the desired decomposition level  $j$ . The signal can be easily retrieved from the progression of successive approximations,  $A_j^x$ , and details,  $D_j^x$ , corresponding to different decomposition level  $j$ , by the fundamental relationship:

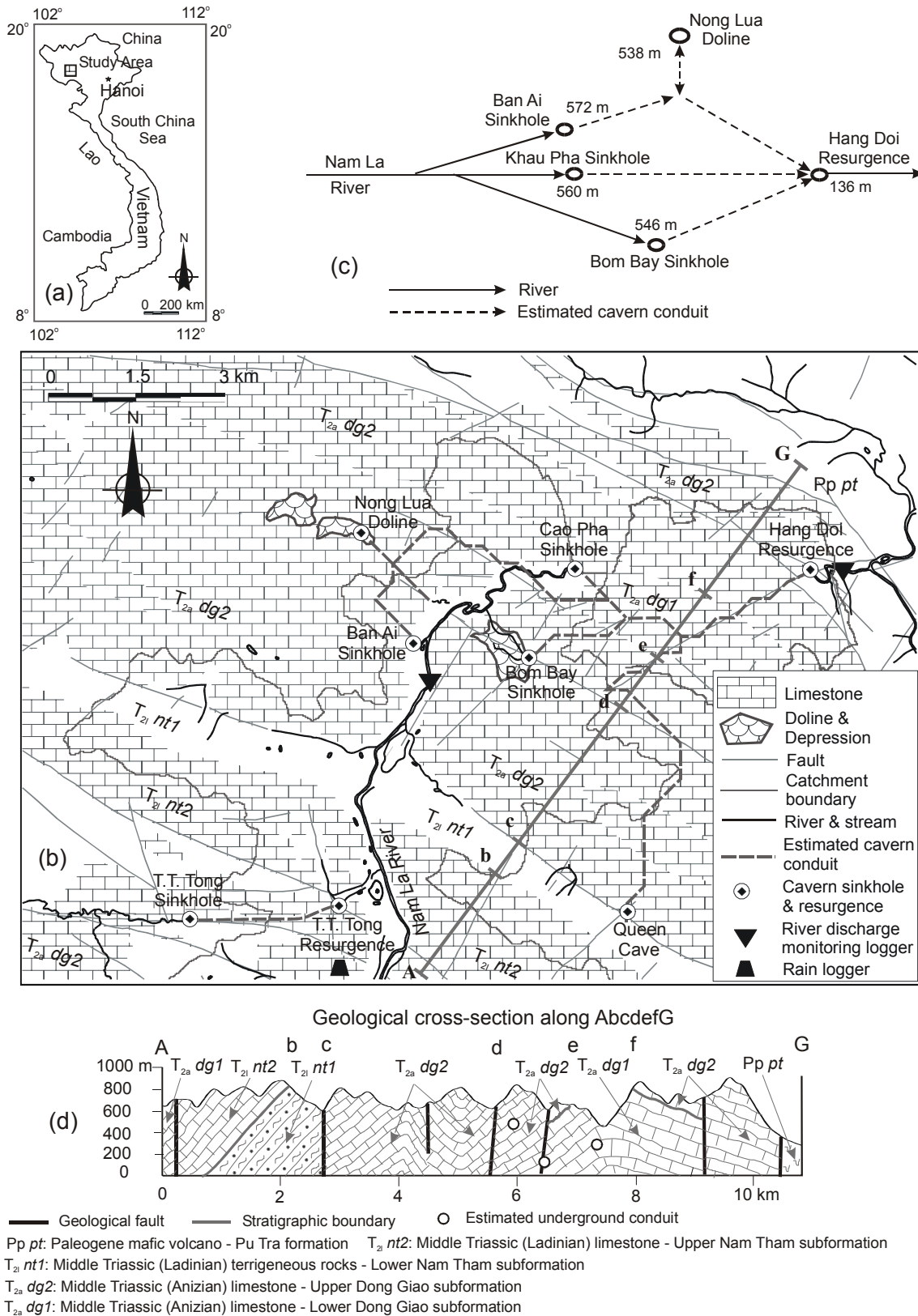
$$x(i) = A_j^x + \sum_{k=1}^j D_k^x \tag{20}$$

Note that the detail corresponds to the difference between two successive levels of approximation  $j$  and  $j + 1$ . Intuitively, one can think of each level of decomposition as a filtering step that outputs the high-scale, low-frequency content (approximation) through a low-pass filter and the low-scale, high-frequency content (detail) through a high-pass filter. Therefore, multilevel wavelet analysis is particularly useful for the

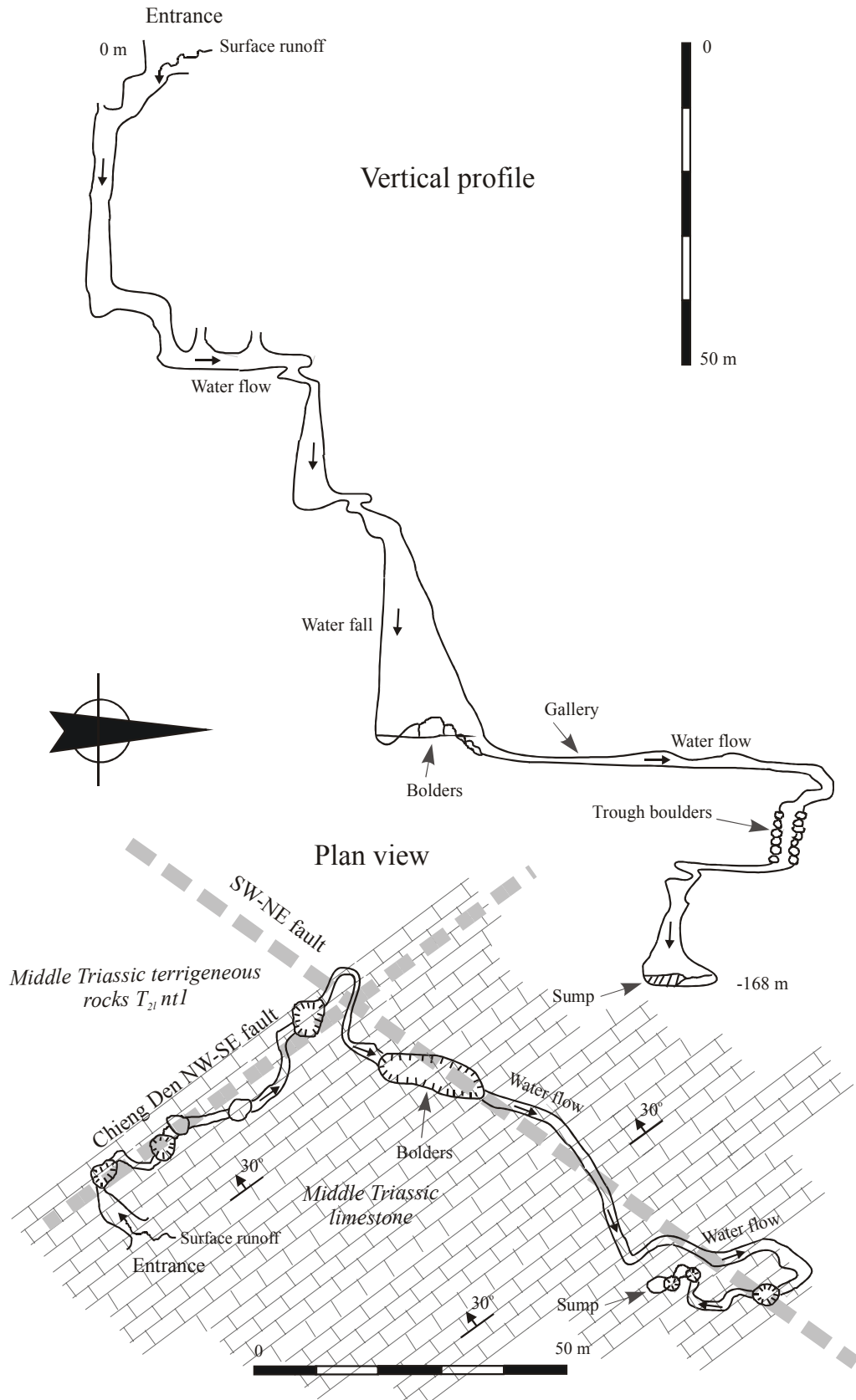
identification, in both time and frequency domains, of tiny and abrupt changes in the time series.

## HYDROGEOLOGICAL SETTING

The cavern conduit system under study is part of the River Nam La, which disappears underground in the sinkholes of Ban Ai, Bom Bay and Cao Pha and reappears at the Hang Doi resurgence (Fig. 3(b) and (c)). The River Nam La drains a basin of approximately 380 km<sup>2</sup> that is located in a high mountain plateau at an elevation of 1700 m (in the south) to 540 m a.s.l. (in the north). Geologically, the basin is composed of: (a) metamorphic sedimentary rocks of Proterozoic-early Devonian age and carbonate rocks of early-middle Devonian age outcropping in the south; (b) continental sedimentary rocks of late Permian-middle Triassic and carbonate rocks of middle Triassic age outcropping in the central and northern parts; and (c) volcanic rocks of Paleogene and Jurassic-Cretaceous age that outcrop in the north. The groundwater in the region is limited to the carbonate rocks that are strongly fractured and karstified and distributed as a 10-km wide band stretching across the study area in a NW–SE direction. Each of these carbonate formations is 800 to 1200 m in thickness. However, based on well-log of boreholes in the study area, the thickness of the superficial and strongest fractured part of these formations is estimated to be less than 100 m. Figure 3(b) shows only the central and northern parts of the basin, where the Triassic carbonate rocks, which make up about one third of the entire basin, dominate. The rocks are regionally northeasterly and southwesterly dipping and form local NW–SE trending anticlines and synclines (Fig. 3(d)). Within the study area, three fault systems exist: (a) the NW–SE striking system includes pre-Cambrian deep faults, which control the structural geology of the region (b) the NE–SW striking system includes younger faults that are less significant in the regional structural geology; and (c) the sub-longitudinal striking system includes the youngest faults. Based on statistical work of pumping well capacity, distribution of springs and their discharge value, Xuyen (1998) shows that the most prominent groundwater occurrences can be found along the NE–SW and sub-longitudinal striking faults. The basin drains to the outlet points of Cao Pha sinkhole and Hang Doi resurgence (Fig. 3(b)). Within the basin, two karstic groundwater units can be defined. The first unit is the carbonate rocks of the Dong Giao formation in the centre, confined in the northeast by the impervious mafic volcanic rocks of the Pu Tra formation and in the southwest by impermeable siltstones and shales of the Lower Nam Tham subformation. The second unit, separated from the first unit by the Lower Nam Tham subformation, is composed of peak karst towers of the Dong Giao formation, the Upper Nam Tham subformation as well as narrow bands of the early-middle Devonian limestones in the southwest. Numerous cavern conduits have been found in the study area, whose development coincides with the development direction of geological faults and fractured zones (November, 1999). Vertical shafts, which connect horizontal conduits, are often found as a consequence of the intersection of two fault systems (Fig. 4). Also, due to tectonic–neotectonic activities, many closed depressions and dolines were formed, most of them located in the highland recharge areas. At the bottom of these depressions and dolines, swallow holes and cavern shaft are often found with entrances partly blocked by sediments, boulders and debris.



**Fig. 3** (a) Location of the study area.; (b) geological sketch of the northern part of the Nam La basin and the estimated underground cavernous conduit system; (c) schematic presentation of the Nam La underground cavernous conduit system; and (d) geological cross-section along the line AbcdefG.



**Fig. 4** Development pattern of the Queen Cave in relation to fault occurrence (adapted with permission of the Speleo Club of the K.U.Leuven).

The karstic groundwater in the study area is mainly stored in fractures, crushed zones and cavern conduits and circulates in consistence with the hydrodynamic conditions and the network of high hydraulic conductivity zones (Tam *et al.*, 2001). Based on cave surveying data (Dusar *et al.*, 1994; Coessens *et al.*, 1996; Lagrou *et al.*, 2002), it is conceptualized that the aquifer consists of perched karst groundwater bodies, each of which develops within the fractured zone along a horizontal cavern conduit. The actual karst groundwater table is likely to exist at the level of the Hang Doi resurgence. The karst aquifer receives water mainly by regional groundwater flow, with additional, important *in situ* recharge by precipitation, surface water and exotic water from higher lying non-karstic denuded areas.

The basin is characterized by a humid subtropical climate with two distinct seasons: an extensive rainy summer from May to October and a dry winter from November to April. The mean total yearly precipitation is 1450 mm, of which 85% falls during the rainy season. Part of the precipitation rapidly recharges the karst groundwater through swallow holes and shafts in closed karst valleys/dolines scattered in the basin. The remaining infiltrates through the soil mantle into dendritic series of fractures in the limestone. The soil mantle, up to 17 m thick, consists of semi-permeable Quaternary deposits and debris of limestone, and occurs mainly in the dolines and valleys and along the river plain. It is assumed that the recharge through the soil mantle lasts much longer than the recharge through the shafts and swallows holes, which takes only a couple of hours. The basin is also known for its very rapid groundwater level fluctuations and sharp flow rate variations at the sinkholes and resurgence, which can be interpreted as evidence of the existence of underground conducting dissolution channels.

Based on cave expeditions, studies on the geological structure, remote sensing-based lineament density measurements, and tracer tests, it is possible to outline the underground conduit system as shown in Fig. 3(b). Two tracer tests were carried out at the end of the rainy season of the years 2001 and 2002; each with 2 kg of uranine injected in the Cao Pha and the Bom Bay sinkholes. The tests revealed that it takes as long as 19 h for the first arrival at the Hang Doi resurgence. Field observation also revealed that, during intensive precipitation periods, the groundwater level in the cavern shaft in the Nong Lua doline, which is assumed to be connected to the underground cavern conduit system, rises as high as 107 m above the dry-season level, resulting in temporal water ponding in the doline for several days.

Two automatic, hourly based, water-level loggers were installed, one to capture the total streamflow, which disappears into the sinkholes, and the other the discharge at the resurgence (Fig. 3(b)). The precipitation in the area was measured 4 km upstream of the sinkholes. In addition, potential evapotranspiration data are monitored by a nearby gauging station and were collected for further calculation of net precipitation, estimated as the difference between the precipitation and the potential evapotranspiration if the difference is more than zero and zero otherwise. The data were monitored hourly, but were afterwards converted to a daily basis. The daily time series of the sinkhole ( $Q_{\text{sink}}$ ), resurgence ( $Q_{\text{resurg}}$ ) and net precipitation ( $P_n$ ), from 16 January 2000 to 21 May 2002, are the baseline data for this study.

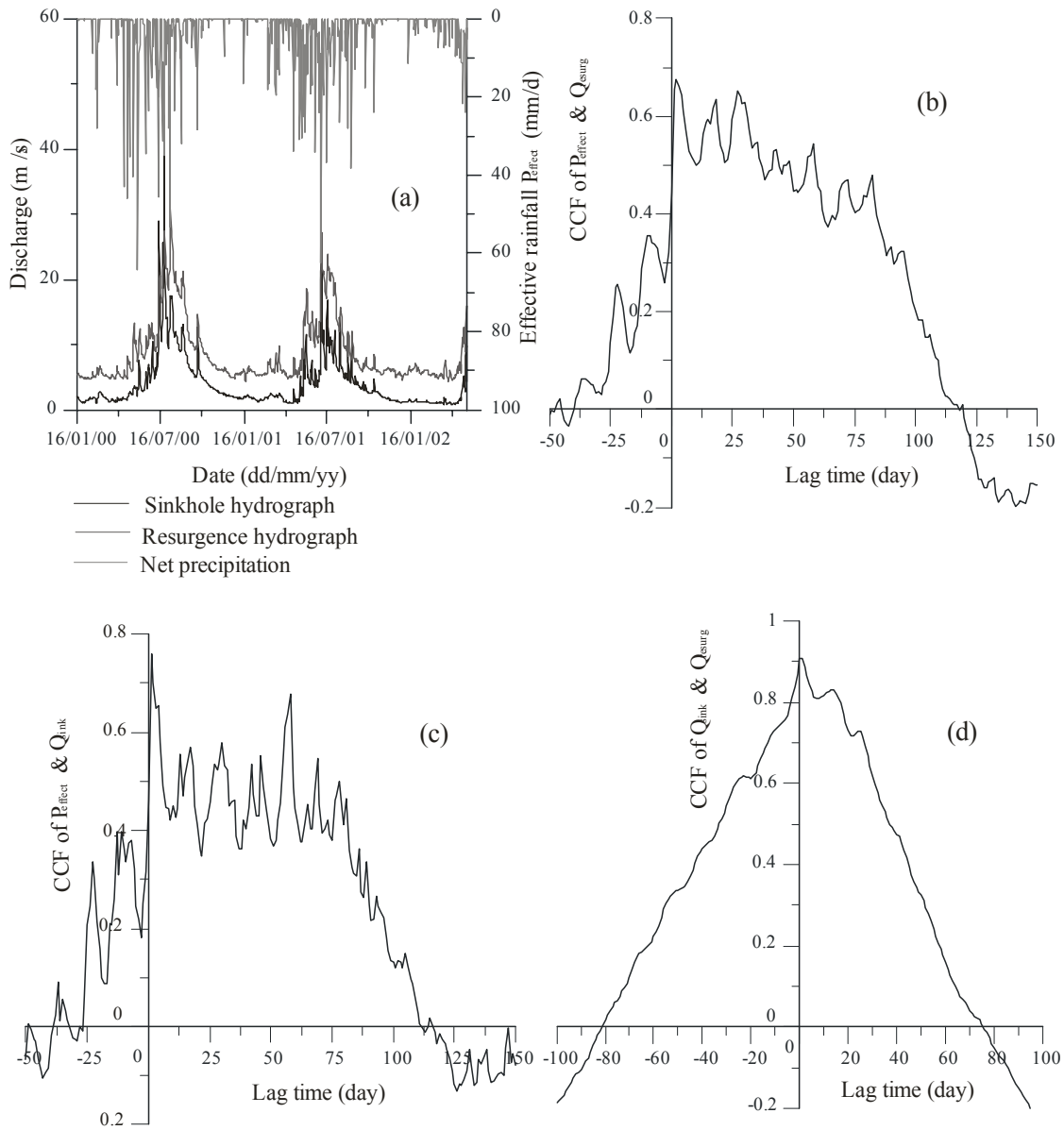
## APPLICATION OF CROSS-CORRELATION AND CROSS-SPECTRUM ANALYSIS

The most important input to the Nam La underground cavern conduit system is the total streamflow,  $Q_{\text{sink}}$ , which flows into the sinkholes of Ban Ai, Bom Bay and Cao Pha. Another input to the cavern conduit system is direct recharge from precipitation within the surface drainage area,  $A$ , between the Cao Pha sinkhole and the Hang Doi resurgence (the 21 km<sup>2</sup> basin in the northeast in Fig. 3(b)). This area is almost completely forested and has no permanent drainage network. Also, no cavern shafts or swallow holes, except a few dolines and depressions, were found in the area. Therefore, the net precipitation,  $P_n$ , percolates as direct recharge to the underlying cavern conduit system through the soil mantle and the dendritic fractures. Since this study focuses on the transient phenomena during the rainy season, and the area between the sinkhole and the resurgence for the estimated direct recharge is relatively small, this recharge will be insignificant in the interpretation of the results. This will be verified by means of cross-correlation and cross-spectrum analyses. Output of the system is the flow rate,  $Q_{\text{resurg}}$ , measured at the Hang Doi resurgence, which is also the main discharge from the entire Nam La basin. Figure 5(a) shows the hydrographs of the sinkholes and resurgence, which show quick responses to the net precipitation. The rate of change of the system storage is:

$$\frac{dS}{dt} = Q_{\text{sink}} - Q_{\text{resurg}} \quad (21)$$

The underground conduit system is considered as a filter that converts the inputs (the flow rate at the sinkholes and the additional recharge) into the outputs (internal storage and the flow rate at the resurgence). A daily time basis was chosen for the analyses, taking into account the proven transfer time (19 h) of the cavern conduits and the mean concentration time (33 h) of the Nam La basin. Because the measured hourly records are averaged to yield daily records, the daily time basis used is also a measure to smooth the time series.

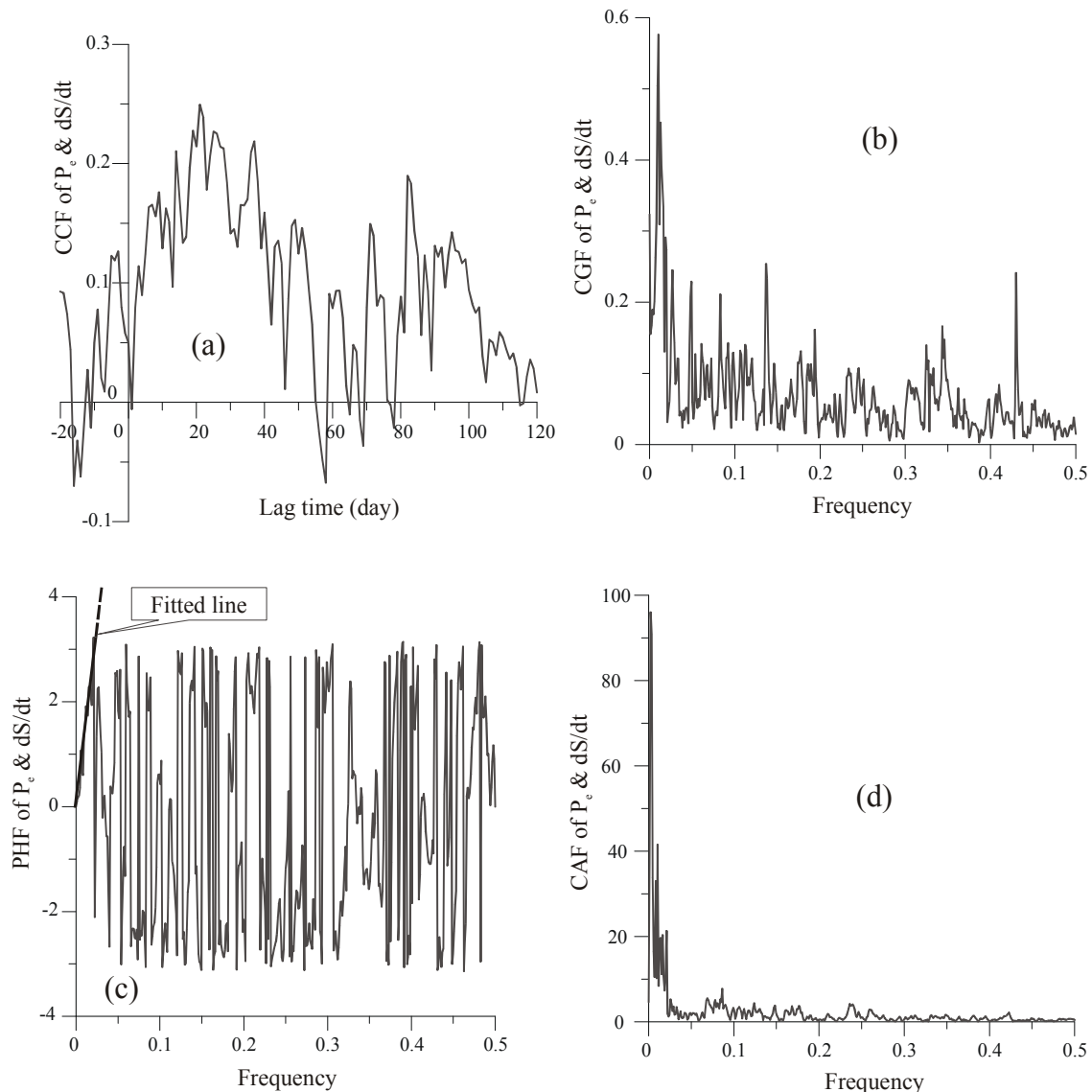
Comparison of the cross-correlations between the net precipitation and the streamflow at the sinkholes (Fig. 5(c)) on the one hand and the net precipitation and the resurgence (Fig. 5(b)) on the other, reveals the filtering function of the Nam La basin and the conduit system. For the cross-correlation  $P_n-Q_{\text{sink}}$ , the net precipitation is the cause of the streamflow at the sinkholes and the Nam La basin is considered as a filter. However, for the cross-correlation  $P_n-Q_{\text{resurg}}$ , the net precipitation for the basins belonging to the Cao Pha sinkhole and the area between the sinkhole and the Hang Doi resurgence is the cause of the outflow measured at the resurgence. The result shows that the net precipitation is related to the flow rate at the sinkholes and the resurgence (Fig. 5(b) and (c)). Both CCFs show a multi-peaked curve reflecting the complicated cyclical structure of the precipitation and discharge series. The lower maxima and smoother curve of the CCF for the net precipitation and resurgence discharge compared to the CCF of the net precipitation and sinkhole discharge reflects the presence of an additional filter system, i.e. the cavern conduit system. Hence, the translation of the precipitation into the river discharge at the resurgence is filtered not only by the basin (which is the case for the river discharge at the sinkhole) but also by the conduit system itself. Note that the precipitation–river discharge delay is 1 day for



**Fig. 5** (a) Net precipitation and hydrographs of sinkholes. Resurgence and cross-correlation functions of (b) net precipitation–resurgence discharge, (c) net precipitation–sinkhole discharge, and (d) sinkhole discharge–resurgence discharge.

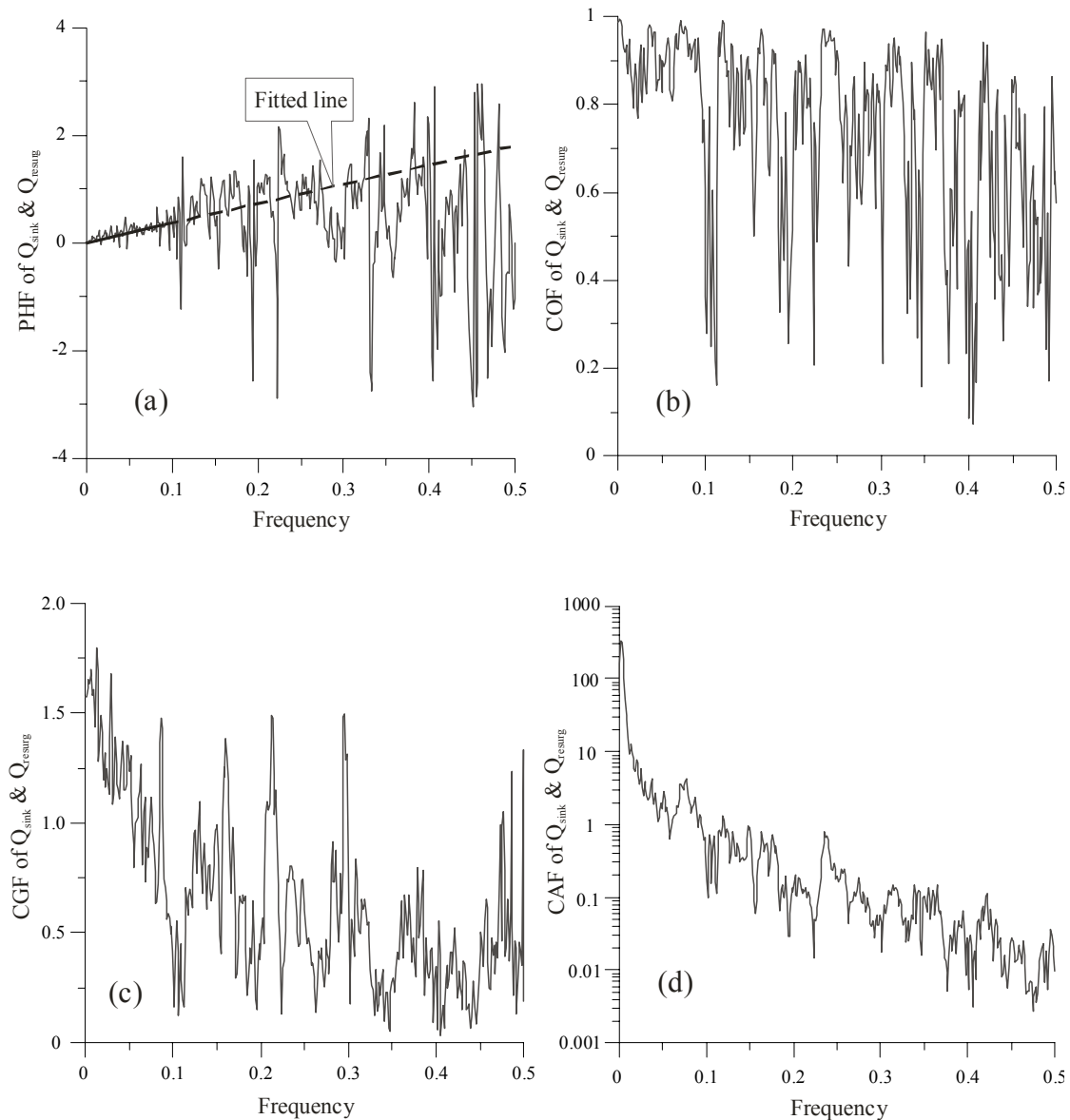
the sinkhole (CCF peaks at lag time of 1 day, Fig. 5(c)) and 2 days for the resurgence (Fig. 5(b)). The delay difference of 1 day shows the average duration of the streamflow translation through the conduit system. This can also be observed in Fig. 5(d), where the CCF of sinkhole–resurgence discharge relationship shows a much smoother curve with a maximum value of 0.91 at a lag time of 1 day. The CCF descends to zero at a time lag of 75 days, which indicates that an event entering the system can have a long-term influence on the system output.

The cross-correlation and cross-spectrum analyses were also carried out to evaluate the direct influence of net precipitation on the conduit storage. The net precipitation infiltrating into the cavern conduit system over the surface drainage area between the Cao Pha sinkhole and the resurgence was taken as input and the change



**Fig. 6** Analyses of the net precipitation and change rate of storage relationship: (a) cross-correlation function, (b) cross-gain function, (c) cross-phase function, and (d) cross-amplitude function.

rate of the conduit storage,  $dS/dt$ , as output of the analysis. The aim here is to explore the filtering function of the soil mantle and epikarst zone, and also the significance of the infiltration to the conduit storage. The result is shown in Fig. 6 and proves that the infiltration has an insignificant contribution to the conduit storage. This can be observed by the low values and chaotic behaviour of the CCF curve (Fig. 6(a)) and almost no amplification of the input over the output in the CGF curve ( $CGF < 1$ , Fig. 6(b)). The maximum value of the CCF at a time lag of 21 days and the mean delay time of 22.3 days, computed with the slope of the fitted straight line shown in Fig. 6(c), gives an estimation of the duration that the net precipitation needs to reach the conduit system. This shows that precipitation enters the system mainly by infiltration through the epikarst zone and the surrounding rock fracture matrix. It is also noted that only components with an annual periodicity (frequency  $< 0.02$ ) exhibit



**Fig. 7** Cross-spectrum analysis of sinkhole and resurgence discharge: (a) cross-phase function, (b) cross-coherence function, (c) cross-gain function, and (d) cross-amplitude function.

an apparent covariance (CAF in Fig. 6(d)) and dephasing (alignment of PHF in Fig. 6(c)) between the input and output series. High frequency components do not exhibit such patterns, probably because during flood periods the infiltration is insignificant compared to the massive storage and input from the sinkholes. Hence, exclusion of net precipitation over the surface drainage area between the Cao Pha sinkhole and the resurgence as an additional input component to the cavern conduit is not crucial and has no influence on the conceptual method.

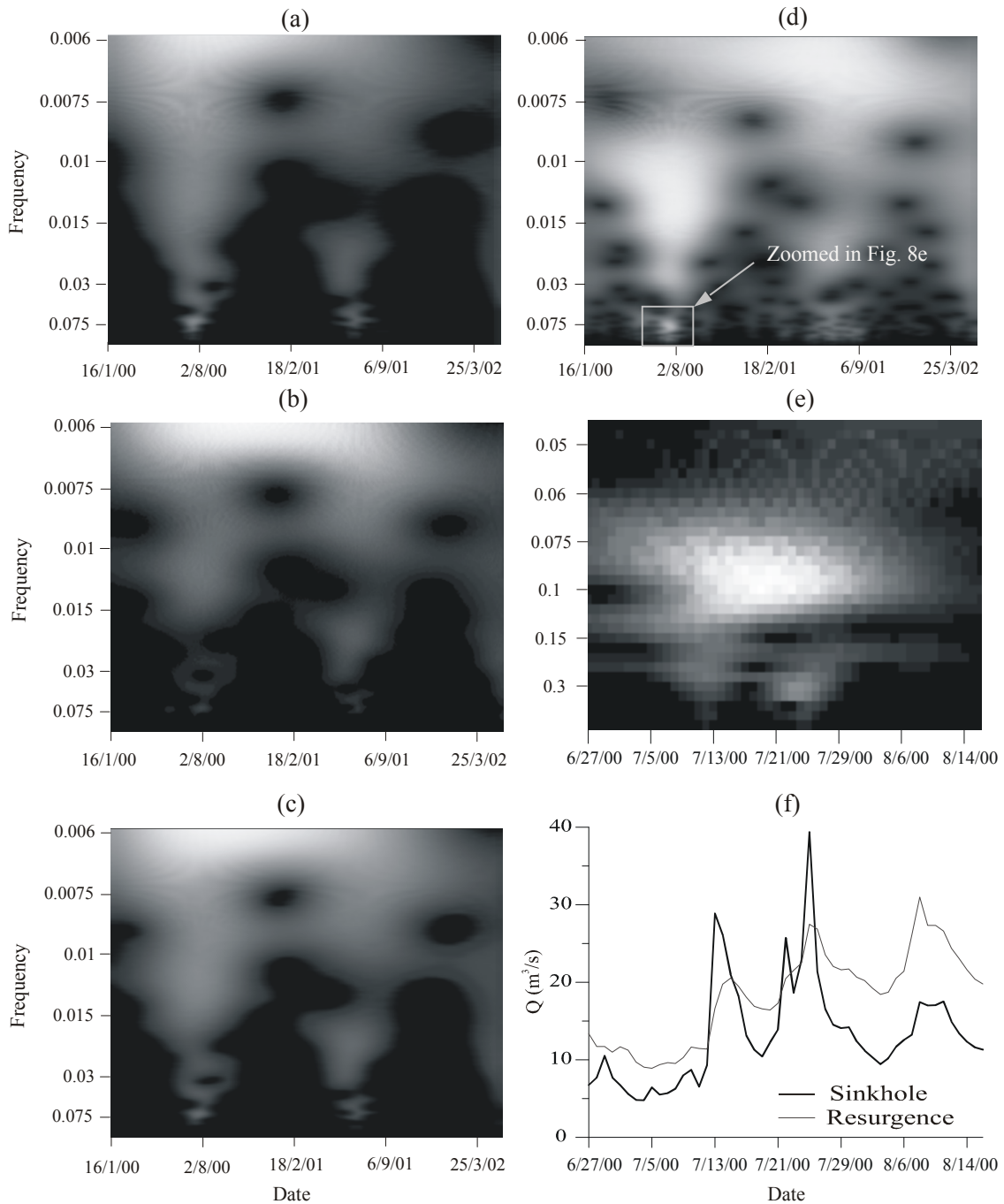
The cross-spectrum analyses of sinkhole and resurgence discharges are shown in Fig. 7. The mean delay, or mean input–output transition time, computed with the slope of the fitted line of the sinkhole–resurgence discharge PHF (Fig. 7(a)), is 0.83 day, which is comparable with the earlier estimate using the CCF. It is noted that the PHF is

virtually aligned only for frequencies lower than 0.075. Also, for this range of frequencies, the input–output relationship is linear (the sinkhole–resurgence discharge COF is nearly equal to 1, Fig. 7(b)), and the system shows an amplification of the outflow over the inflow, i.e. the sinkhole–resurgence discharge CGF is larger than 1 (Fig. 7(c)). This means that during recession periods the river water throughput to the sinkholes emerges at the resurgence after approximately 1 day. This linear transfer is no longer observed for higher frequencies. The reason is that during extensive rainy periods the river flow sometimes exceeds the conveying capacity of the cavern conduit, creating pressurized conditions in the conduit. In this case, only a part of the input flows through the system; the surplus quantity of water remains in the conduit in chambers and sumps as internal storage and also as feedback to the surrounding fractured matrix, and possibly emerges and temporarily stores in the Nong Lua doline. When the streamflow recedes and becomes less than the conveying capacity of the system, the pressurized state finishes and the linearity of the input–output transform resumes.

The cross-amplitude function (Fig. 7(d)) also shows a strong covariance between the recession flow components (frequency  $< 0.075$ ) of the sinkhole and resurgence streamflow. The function is probably not suitable for a nonlinear input–output system, especially in cases when sharply changing and small-amplitude components (high frequency) enter the system and are lumped there as internal storage for a certain time before flushing out of the system. Therefore, the system response to high-frequency flow components cannot be observed in the CAF. It must be emphasized that switching from a free-surface mode to a pressurized mode (and *vice versa*) is a very complicated hydraulic process that cannot be fully described by a simplified approach as used in this study. The purpose here is to show virtually if the phenomenon exists in the cavern conduit system and how it can be detected with the classical cross-correlation and cross-spectrum and wavelet analyses presented hereafter.

## APPLICATION OF WAVELET SPECTRUM AND CROSS-WAVELET SPECTRUM ANALYSES

The ability of wavelet analysis to represent localized and transient phenomena in the time domain allows clarification of what was discovered by the cross-correlation and cross-spectrum analyses. Continuous wavelet spectrum and cross-spectrum analyses were applied to the sinkhole and resurgence discharge time series based on a complex Morlet wavelet. Often, the result of a wavelet spectral analysis is a family of wavelet coefficients,  $C(a,\tau)$ , and is presented in a two-dimensional diagram (wavelet spectrum), where the scale  $a$  is in the vertical axis and the time  $\tau$  in the horizontal axis, and where the higher values of the wavelet coefficients correspond to the lighter grey scale and *vice versa*. For easy comparison with the results obtained with the Fourier analyses, scales in the vertical axis were converted to pseudo-frequencies. Because of software limitations, the scale was limited to a value of 256, i.e. components of frequencies lower than 0.0059 could not be investigated. This limitation is not a serious obstacle considering that the focus of this analysis is transient events of high frequencies. Also, note that the sampling unit is one day. Therefore, events of low frequencies near the above-mentioned limit were considered as having an annual



**Fig. 8** Continuous Morlet wavelet spectrum of (a) sinkhole discharge, (b) resurgence discharge, (c) cross-correlation of sinkhole–resurgence discharge, (d) change rate of conduit storage  $dS/dt$ , (e) zoomed view of the wavelet spectrum of  $dS/dt$ , and (f) stream discharge at the sinkhole and resurgence for the flood period 27 June–16 August 2000.

periodicity. Discrete wavelet transform was also performed based on the fourth-order Daubechies wavelet, to identify long-term evolution, discontinuities and breakdown points, and self-similarity of the time series.

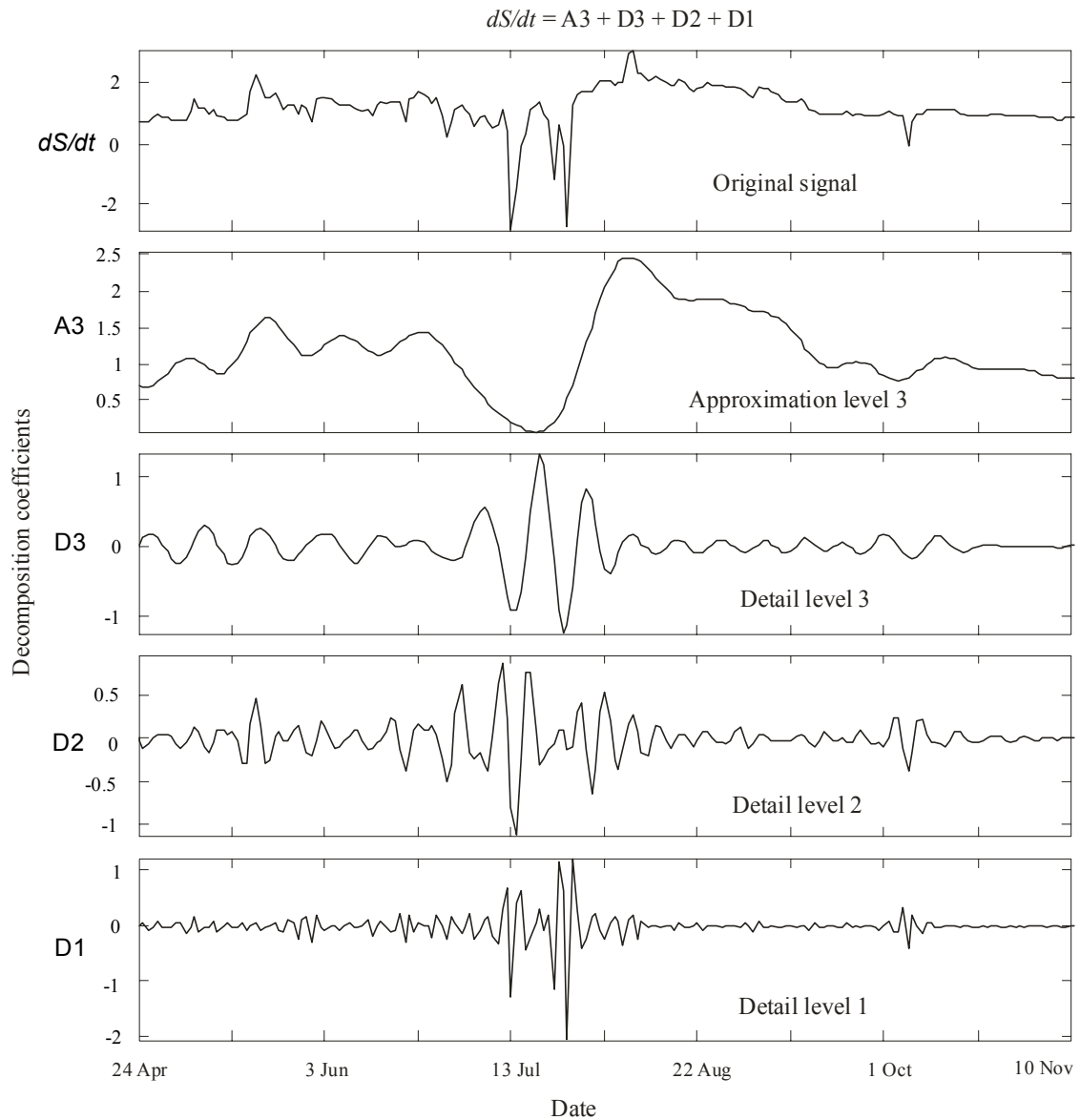
The results of the Morlet wavelet analysis are shown in Fig. 8. The wavelet spectrum coefficients of the sinkhole discharge series (Fig. 8(a)) display the presence

of flow components of all frequencies, from abruptly varying flow occurring during flood periods to smoothly varying flow during recession periods. It is also observed that the wavelet coefficients gradually decrease as the frequency increases. In contrast, the wavelet spectrum coefficients of the resurgence discharge series (Fig. 8(b)) greatly enhance the low frequency ( $<0.0067$ ) components while blurring other higher frequencies. This implies that the cavern conduit system acts as a low-pass filter that strongly attenuates high frequency components of the input flow.

The wavelet cross-spectrum of the sinkhole–resurgence discharge (Fig. 8(c)) again ascertains that the linear input–output relationship virtually exists for flow components of frequencies lower than 0.075, which agrees with the results of the Fourier analyses. In combination with Fig. 5(d), one may deduce that the cross-correlation coefficient reaches a maximum value of 0.91 for low-frequency flow components during recession periods and rapidly drops with high-frequency flow components during flood periods.

The spotted pattern (i.e. many black hollows of very low coefficients surrounded by higher coefficients) in the high-frequency bands of the wavelet spectrum (Fig. 8(d)) shows that the change rate of storage is highly variable in time in response to high-frequency flow components. Three areas of high wavelet coefficients are noticed: (a) the area of low frequencies ( $<0.0075$ ) distributed throughout the year, (b) the area of intermediate frequencies (0.02–0.01) distributed in the rainy season, and (c) the area of high frequency occurring during the flood period 12–26 July 2000. A zoomed view of the latter reveals that, at the frequency of 0.094, the wavelet coefficient attains the highest value (Fig. 8(e)). This corresponds to the period when the stream discharge abruptly varies and the input of the cavern system exceeds the output (Fig. 8(f)). It is also noticed that this phenomenon occurs outside the range of frequencies where the linearity of the input–output relationship is observed. It is suspected that the exceeding water volume is temporarily stored in the fractures surrounding the cavern conduit and also possibly in the Nong Lua doline; hence there is no longer a linear response as for the lower frequency components. Based on Fig. 8(e)–(f), the conveying capacity is estimated at around 20 to 25  $\text{m}^3 \text{s}^{-1}$ . It is worth noting that the conveying capacity of a cavern conduit system can vary slightly depending upon the hydraulic heads imposed on the two ends of the system.

For a better interpretation of the transition from a linear to a nonlinear mode of the input–output relationship, multilevel decomposition of the change rate of storage for the summer of 2000 was carried out based on the Daubechies discrete wavelet transform (Fig. 9). The figure shows the original signal  $dS/dt$ , approximation A3, and three details D1–D3. Detail D3 reveals a variation that is composed of three oscillations with duration of about 32 days (3 June–4 August). This variation of change in storage, at frequency  $1/(32/3) = 0.094$ , is indeed present in the continuous wavelet spectrum (Fig. 8(d) and (e)). Detail D3 also shows the presence of components with an approximate frequency of 0.083 before and after the component of frequency 0.094. This can be associated with the filling-up of (and release from) the surrounding fractures of the cavern conduit system. The purpose here is to define a threshold that can be used to identify periods within the entire time series when the conduit flow possibly comes under pressure. The magnitude of the decomposition coefficient makes components of certain frequency distinguishable from others in the wavelet spectrum presentation and this leads to the identification of the threshold for the cavern conduit



**Fig. 9** Multilevel decomposition of signal  $dS/dt$  for the period 24 April–10 November 2000 based on fourth-order Daubechies discrete wavelet transform.

system. The system is defined by a threshold frequency of 0.094 and decomposition coefficient of 1.0; any event above these thresholds can be associated with pressurized flow. With this identified threshold, one can see that some events in the rainy season of 2001 (light spots near the bottom of Fig. 8(d)) are not considered as pressurized flows since their decomposition coefficient is not high enough, although they have a frequency higher than 0.094.

## CONCLUSIONS

In this study, cross-correlation, cross-spectrum and wavelet spectrum analyses were carried out to investigate the hydrodynamic functioning of a karstic conduit system in

the Nam La basin, northwest Vietnam. The system is considered as a filter that transforms, retains or eliminates the input signal into an output signal. Input to the system consists of the net precipitation and streamflow at the sinkholes, and the output is the streamflow at the resurgence. In the analyses, the time series of input and output are considered as signals that can be broken into components of different frequencies. The low frequencies can be associated with slowly and smoothly changing events such as streamflow during recession periods, while the high frequencies relate to abruptly changing events such as floods. By analysing the power distribution (i.e. spectrum) of the respective frequency components, one can identify the hydrodynamic behaviour of the conduit system for the different events. This approach yields hydrodynamic properties such as system memory, response time, and mean delay between the input and the output. The study also shows that the input–output relationship is linear only for low flows; during high-flow periods, the flow in the conduit locally comes under pressure and the water is temporarily stored in the fractured rocks around the conduit system. Fourier and wavelet analyses have shown to be effective study tools in the field of hydrology. However, applications of these techniques have been focused mostly on the scale of an entire basin. In this study, the power of these techniques is shown with an application on a smaller scale, i.e. a cavern conduit system. In the example presented here, the time series analyses provide information that is directly compatible with tracer tests. Moreover, they yield an insight into the hydraulic properties of the conduit flow; that is the transition from a linear to a nonlinear mode, which cannot be obtained with classical tracer tests. The results obtained can also be used as essential guidelines for the design of a numerical simulation model.

As shown in this study, the classical applications of cross-correlation, cross-spectrum and wavelet analyses are complementary. That is, the result of cross-correlation analysis paves the way for applying Fourier analysis, which in turn gives information for the application of the wavelet analysis. The central point of this study is the determination of a range of frequencies for which flow components are considered linear in the input–output relationship. Events beyond that frequency range feature transient pressurized flow, verified with overflow conditions occurring during a flood period. There is a clear coincidence of a nonlinear input–output relationship with temporal storage effects as shown in this study, and it is a likely indication of the conduit flow being under pressure. These phenomena have not been verified with other observations, and cannot be ascertained given the great complexity in morphology of the system.

**Acknowledgements** This work was carried out within the project A3210 “Rural development in the mountain karst area of NW Vietnam by sustainable water and land management and social learning: its conditions and facilitation (VIBEKAP)” funded by the Flemish University Council (VLIR). The authors are grateful to all VIBEKAP’s participants for their contributions. Special thanks are due to the Project Coordinators, Dr Do Tuyet and Dr Koen Van Keer for their warm encouragement and support during this study.

## REFERENCES

- Amraoui, F., Razack, M. & Bouchaou, L. (2003) Turbidity dynamics in karstic systems. Example of Ribaa and Bittit springs in the Middle Atlas (Morocco). *Hydrol. Sci. J.* **48**(6), 971–984.
- Bayazit, M., Onoz, B. & Aksoy, H. (2001) Nonparametric streamflow simulation by wavelet or Fourier analysis. *Hydrol. Sci. J.* **46**(4), 623–634.
- Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis Forecasting and Control*. Holden Day, San Francisco, California, USA.
- Chui, C. K. (1992) *Wavelets: a Tutorial in Theory and Applications*. Academic Press, Boston, Massachusetts, USA.
- Chui, C. K. & Montefusco, L. (1994) *Wavelets: Theory, Algorithms, and Applications*. Academic Press, San Diego, California, USA.
- Coessens, V., Deblaere, C., Lagrou, D., Masschelein, J., Tien, P. C. & Tuyet, D. (1996) Son La 1995–1996. Cave investigation, a start for research on sustainable development. Belgian-Vietnamese Speleological expedition, BVKCA-SPEKUL-RIGMR-report (unpublished).
- Coifman, R. R., Meyer, Y. & Wickerhauser, M. V. (1992) Wavelet analysis and signal processing. In: *Wavelets and Their Applications* (ed. by M. B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer & L. Raphael), 153–178. Jones and Bartlett, Boston, Massachusetts, USA.
- Daubechies, I. (1992) Ten lectures on wavelets. *CSBM-NSF Series Appl. Math. no. 61*, SIAM Publication, Philadelphia, Pennsylvania, USA.
- Delleur, J. W. & Rao, R. A. (1971) Linear systems analysis in hydrology: the transform approach, kernel oscillations and the effect of noise. In: *Systems Approach to Hydrology* (ed. by V. Yevjevich), 116–142. Water Resources Publications, Fort Collins, Colorado, USA.
- Dreiss, S. J. (1983) Linear unit-response functions as indicators of recharge areas for large karst spring. *J. Hydrol.* **61**, 31–44.
- Dusar, M., Masschelein, J., Tien, P. C. & Tuyet, D. (1994) Belgian-Vietnamese Speleological expedition Son La 1993. *Belg. Geol. Survey, Prof. Paper 1994/4 no. 271*. Brussels, Belgium.
- Erskine, A. D. & Papaioannou, A. (1997) The use of aquifer response rate in the assessment of groundwater resources. *J. Hydrol.* **202**, 373–391.
- Estrela, T. & Sahuquillo, A. (1997) Modeling the response of karstic spring at Arteta Aquifer in Spain. *Ground Water* **35**(1), 18–24.
- Ford, D. C. & Williams, P. W. (1989) *Karst Geomorphology and Hydrology*. Unwin Hyman, London, UK.
- Jenkins, G. M. & Watts, D. G. (1968) *Spectral Analysis and its Applications*. Holden Day, San Francisco, California, USA.
- Kaiser, G. (1994) *A Friendly Guide to Wavelets*. Birkhäuser, Boston, Massachusetts, USA.
- Labat, D., Ababou, R. & Mangin, A. (2000a) Rainfall–runoff relations for karstic springs, Part I: convolution and spectral analyses. *J. Hydrol.* **238**, 123–148.
- Labat, D., Ababou, R. & Mangin, A. (2000b) Rainfall–runoff relations for karstic springs, Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *J. Hydrol.* **238**, 149–178.
- Lagrou, D., Masschelein, J., Philips, P. & Tuyet, D. (2002) Belgian-Vietnamese Speleological expedition. Son La 2000–2001. Report of the fifth expedition in the provinces of Son La and Lai Chau, SPEKUL-BVKCA-RIGMR report, Leuven/Hanoi.
- Larocque, M., Banton, O. & Razack, M. (2000) Transient-state history matching of a karst aquifer groundwater flow model. *Ground Water* **38**(6), 939–946.
- Larocque, M., Mangin, A., Razack, M. & Banton, O. (1998) Contribution of correlation and spectral analyses to the regional study of a large karst aquifer (Charente, France). *J. Hydrol.* **205**, 217–231.
- Liu, P. C. (1995) Wavelet spectrum analysis and ocean wind waves. In: *Wavelets in Geophysics* (ed. by E. Foufoula-Georgiou & P. Kumar), 151–166. Academic Press, New York, USA.
- Long, A. J. & Derickson, R. G. (1999) Linear systems analysis in a karst aquifer. *J. Hydrol.* **219**, 206–217.
- Mallat, S. (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Pattern Anal. and Machine Intell.* **11**, 674–693.
- Mangin, A. (1984) Pour une meilleure connaissance des systèmes hydrologiques à partir des analyses corrélatoire et spectrale. *J. Hydrol.* **67**, 25–43.
- Molénat, J., Davy, P., Gascuel-Oudoux, C. & Durand, P. (1999) Study of three subsurface hydrologic systems based on spectral and cross-spectral analysis of time series. *J. Hydrol.* **222**, 152–164.
- November, J. (1999) Karst geological investigation in the area of Son La and Thuan Chau (NW-Vietnam). Lic. Dissertation, KU Leuven (in Dutch).
- Oppenheim, A. V. & Schaffer, R. W. (1975) *Digital Signal Processing*. Prentice Hall, New Jersey, USA.
- Padilla, A. & Pulido-Bosch, A. (1995) Study of hydrographs of karstic aquifers by means of correlation and cross-spectral analysis. *J. Hydrol.* **168**, 73–89.
- Palmer, A. N. (1991) Origin and morphology of limestone caves. *Geol. Soc. Am. Bull.* **103**, 1–21.
- Priestley, M. B. (1981) *Spectral Analysis and Time Series*. Academic Press, New York, USA.
- Priestley, M. B. (1988) *Non-linear and Non-stationary Time Series Analysis*. Academic Press, New York, USA.
- Smart, P. L. & Hobbs, S. L. (1986) Characterization of carbonate aquifers: a conceptual base. In: *Proc. Conf. on the Environmental Problems in Karst Terranes and Their Solutions* (Bowling Green, Kentucky, 28–30 October 1986), Proc. National Water Well Association, Dublin, Ohio, USA, 1–14.
- Tam, V. T., Vu, T. M. N. & Batelaan, O. (2001) Hydrogeological characteristics of a karst mountainous catchment in the northwest of Vietnam. *Acta Geol. Sinica* (English edn), **75**(3), 260–268.
- Torrence, C. & Compo, G. P. (1998) A practical guide to wavelet analysis. *Bull. Am. Met. Soc.* **79**, 61–78.
- White, W. B. (1969) Conceptual models of carbonate aquifers. *Ground Water* **7**(3), 15–21.

- Xuyen, C. X. (1998) Report on hydrogeological mapping scaled 1:200 000, Dien Bien Yen Bai Area. Geological Survey of Vietnam, (in Vietnamese).
- Yevjevich, V. (1972) *Stochastic Processes in Hydrology*. Water Resources Publications, Fort Collins, Colorado, USA.

**Received 30 September 2003; accepted 24 June 2004**