# Periodic and aperiodic traveling pulses in population dynamics: An example from the occurrence of epidemic infections

A. Vecchio, L. Primavera, and V. Carbone
*Dipartimento di Fisica, Università della Calabria, and Istituto Nazionale di Fisica della Materia, Unità di Cosenza, Ponte P. Bucci, Cubo 31C, 87030 Rende (CS), Italy*

(Received 25 May 2005; revised manuscript received 9 November 2005; published 14 March 2006)

The dynamics of the occurrence of the dengue hemorrhagic fever in the 72 provinces of Thailand is investigated by performing a proper orthogonal decomposition (POD) on spatiotemporal data. Using this technique, we are able to identify and select the contribution of different modes, selected according to the energy content, to the evolution of the epidemic during 14 years. We found that the phenomenon is characterized by periodic cycles of yearly occurrence characterized by spatial scales of about 420 km. Superimposed on this basic mode, POD analysis is able to reveal the presence of high-energetic aperiodic traveling pulses of the epidemic, which extend spatially for about 510 km from Bangkok.

DOI: 10.1103/PhysRevE.73.031913 PACS number(s): 87.23.Cc, 07.05.Kf, 87.18.Hf, 89.75.Fb

## I. INTRODUCTION

Spatial synchrony of population dynamics which is observed in ecology is yet not well understood and is currently an active area of research [1,2]. Apart from predator-prey systems that have an intrinsic tendency to cycle [3–6], ecologists find regular empirical patterns of population dynamics as a kind of self-organization within the system [7,8]. One key to progress in understanding pattern formations in complex systems is to recognize that many apparently random structures obey a symmetry as striking as that obeyed by regular structures. A development in this field is made possible by new statistical methods, extensively used on spatiotemporal data sets [9], and by massive computer resources to analyze huge data sets.

An important contribution to the debate on the cyclic behavior of population dynamics has been the identification of periodic traveling waves in the dynamics of cyclic field voles [10]. More generally traveling waves are represented by equally spaced peaks of a state field observed in population dynamics. A popular version of this phenomenon is the occurrence of the well known "Mexican wave" in sports stadiums. More dramatically the wave is often caused by the invasion of a virus within a population in a given spatial region, thus generating periodic infection. Recently this has been observed in the occurrence of dengue hemorrhagic fever (DHF) in Thailand [11]. Dengue fever [12] is a mosquito-borne virus that infects 50–100 million people each year. Few tools exist to control dengue virus infection and trasmission. In particular control is focused on controlling the mosquito responsible for the desease, *Aedes aegypti*, and on effective management of cases of infections [12]. The incidence spatiotemporal pattern presents nonstationary and nonlinear features due to the complex effects and interaction of several factors. The predominant factors influencing the incidence data include environmental and climate factors [13,14], predator-prey dynamics between the pathogen and the host population [15,16], and viral factors [17,18]. As a consequence the incidence data show strong seasonality and multiyear and intrayear oscillations.

For many families of patterns occurring in a wide range of different physical situations, it is possible to obtain a useful systematic characterization of the spatiotemporal dynamics. Often, the main motivation is that the family is low dimensional, that is, each member of the family might be represented by a small number of parameters. In these cases we can use the proper orthogonal decomposition (POD), also called the Karhunen-Loéve decomposition [19], to obtain a finite set of discrete modes that optimally describe the problem at hand. In the present paper we report the results obtained by applying the POD to the DHF data set. We show that POD is a powerful tool to investigate a complex two-population dynamics, characterized by the presence of both aperiodic and periodic components. POD has been extensively used in the natural context of analyzing turbulent fields [19,20]; it has been successfully applied also in astronomy [21–23]. In our present work the POD is used, to our knowledge for the first time, to analyze spatiotemporal fields in a biomedical context.

## II. THE DHF INCIDENCE DATA SET AND ITS FOURIER SPECTRUM

The data set we used contains the DHF incidence in Thailand $I(r_n,t)$, where $r_n$ $(n=1,\ldots,72)$ indicates the distance from Bangkok of different provinces of Thailand. The various $r_n$ define a spatial grid which is not equally spaced. The field describes 850 000 infections during the period 1983–1987, recorded on a monthly frequency with a total of $T=168$ time steps. Figure 1 shows the contour plot of monthly incidence of DHF for every Thailand province in the time-distance plane. The distances from Bangkok are expressed in kilometers. At first sight a kind of temporal synchrony across the countries is present. To recover this synchrony we will make a Fourier analysis of the field, namely, we use the following decomposition of the field:

$$I(r_n,t)=\sum_{k_n,\omega} a_{k_n,\omega}\exp[i(k_n r_n-\omega t)]$$

where $k_n$ is the wave vector associated with the spatial grid and $\omega$ is the frequency. The Fourier analysis applied to

 

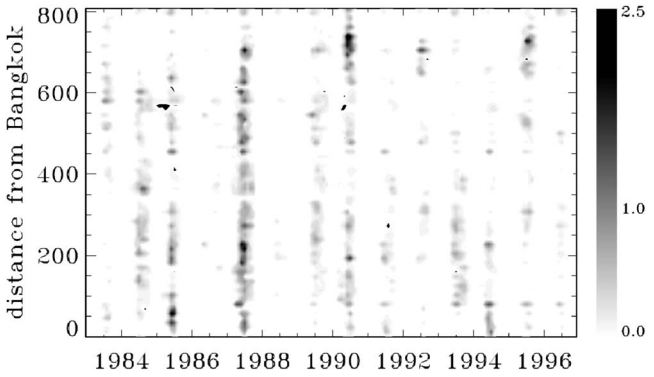


FIG. 1. Contours of monthly DHF incidence for the 72 provinces of Thailand in the space-time plane. The distance $r_n$ corresponding to the $n$th province represents the distance from Bangkok in kilometers.

$I(r_n,t)$ furnishes the set of coefficients $a_{k_n,\omega}$. A plot of the contours of the spectral power $|a_{k_n,\omega}|^2$ in the $(k_n,\omega)$ plane is reported in Fig. 2. Two main features can be noted: (1) a high-power region along the line corresponding to a frequency $\sim 1/1$ yr; (2) the presence of some power randomly distributed at lower frequencies.

It seems that lower frequencies are excited, but the Fourier analysis is unable to recover these frequencies or some typical spatial pattern corresponding to these frequencies. This is perhaps due to some troubles when the Fourier analysis is applied to real data. Two major difficulties are the fact that often Fourier modes are far from being eigenfunctions of the phenomenon at hand, and that boundary conditions are not periodic. In this last case Fourier modes are mixed together in order to build up a solution that corresponds to the fictitious periodic boundary conditions imposed by the Fourier analysis. Here we would like to apply the POD analysis to the DHF incidence data set. The main advantages of the applicability of POD to data sets coming from spatiotemporal complex systems are (1) the fact that POD modes are selected according to the energy content of structures, namely, POD provides an optimal basis to describe the spatiotemporal field at hand; (2) the fact that POD eigenfunctions are not fixed *a priori*, rather they are selected by maximizing the average projection of the field onto the eigenfunctions themselves. In the next section we will briefly introduce the POD technique.

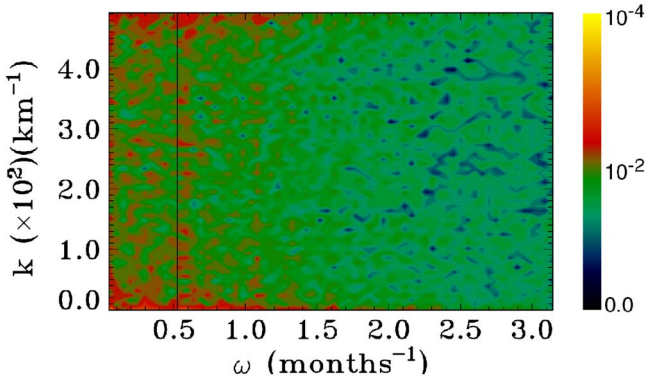


FIG. 2. (Color online) Contours of spectral power $|a_{k_n,\omega}|^2$, in the plane $(k_n,\omega)$, obtained by Fourier trans- forming the monthly DHF incidence for the 72 provinces of Thailand. The vertical line indicates the frequency corresponding to 1 yr.

## III. THE PROPER ORTHOGONAL DECOMPOSITION

Introduced in the framework of turbulence [19], the POD decomposes a spatiotemporal field $u(x,t)$ as

$$u(x,t) = \sum_{j=0}^{\infty} a_j(t)\Psi_j(x) \tag{1}$$

the eigenfunctions $\Psi_j$ being constructed by maximizing the projection of the field onto $\Psi_j$ averaged in time and constrained to the unitary norm. This procedure leads to an optimization problem that can be cast as

$$\int_{\Omega} \langle u(x,t)u(x',t)\rangle \Psi(x')dx' = \lambda\Psi(x) \tag{2}$$

where $\Omega$ represents the spatial domain and the angular brackets represent time averages. The integral Eq. (2) provides the eigenfunctions $\Psi_j$ and a countable, infinite set of ordered eigenvalues $\lambda_j \geqslant \lambda_{j+1}$. When $u(x,t)$ is the velocity field of a turbulent flow, each $\lambda_j$ represents twice the kinetic energy of the $j$th mode. Thus, POD builds up the basis functions, which are not given *a priori*, but rather obtained from observations. The time coefficients $a_j(t)$ are computed from the projection of the data on the corresponding basis functions $\Psi_j(x)$,

$$\int_{\Omega} u(x,t)\Psi_j(x) = a_j(t). \tag{3}$$

Being extracted directly from experiments, the POD eigenfunctions $\Psi_j(x)$ can assume the proper functional shape of the phenomenon, and the associated temporal part $a_j(t)$ represents the time evolution of the $j$th mode associated with that eigenfunction. The POD is an optimal expansion of fields; thus a truncated POD expansion like (4) contains the largest possible energy with respect to any other linear decomposition of the same truncation order [19]:

$$u_N(x,t) = \sum_{j=0}^{N} a_j(t)\Psi_j(x). \tag{4}$$

This means that when we are dealing with a periodic field, the empirical eigenfunction is just the Fourier base $\Psi_j(x) = \exp(ik_jx)$ [19]. The POD basis functions are optimal in turbulence studies with respect to the classical Fourier analysis, where the basis functions are not proper eigenfunctions of the signal.

Once the Fredholm equation has been solved and eigenvalues and eigenfunctions have been obtained, a reconstruction of the original field by selectively choosing a finite number $N$ of the most energetic modes is possible. In this way we form a subspace spanned by the first $N$ eigenfunctions. These eigenfunctions can be used either to build a reduced model

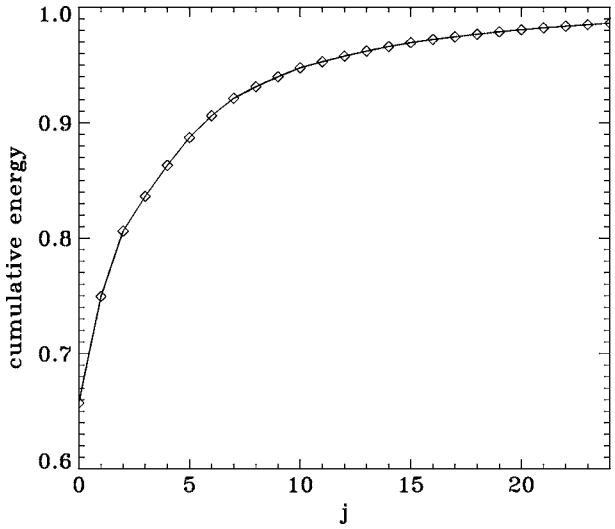


FIG. 3. Cumulated energy of the first $j \leqslant 20$ POD modes as a function of the index $j$.

of the dynamical behavior of the system or in an indirect way to quantify the number of relevant modes taking part in the dynamics. This method is particularly appropriate when analyzing complex physical systems, where different dynamical regimes coexist. POD allows one to identify these regimes and to characterize their energetics and their spatial structure. It must be realized that POD codes need massive computer resources to solve the Fredholm integral equation. It is thus difficult to work with big data sets.

## IV. DATA ANALYSIS

The POD analysis applied on the DHF incidence gives a set of eigenfunctions $\Psi_j(r_n)$ and coefficients $a_j(t)$, as well as the sequence of eigenvalues $\lambda_j$ ($j=0$, 1,…, 167), sorted by decreasing *energetic* content. In the usual turbulence analysis, where the field to be analyzed is represented by the velocity of the fluid flows, $\lambda_j$ represents the energy content of fluctuations associated with the $j$th POD mode. In the present context $\lambda_j$ represents the square of the fluctuating DHF incidence associated with the $j$th POD mode. Then it is naturally related to the *strength* of the infection, namely, the higher $\lambda_j$ the stronger the infection.

In Fig. 3 we report the cumulative "energy" for each POD mode, defined as $E_{cum}(j)=\Sigma_{k=0}^{j}(\lambda_k/E_{tot})$ where $E_{tot}$ is the sum of $\lambda_j$ for the 167 POD modes. The mode $j=0$, being the average contribution, contains the majority of energy, about 65%, while the first seven modes $j=0,\ldots,j=6$ contain about 90% of the total energy. Further modes add little contribution; 99% of total energy resides in the first $j \leqslant 24$ POD modes. In contrast, in laboratory turbulent flows, where large scale energetic coherent structures are present, POD confines almost 99% of the total energy in the modes $j \leqslant 2$ [20]. This difference indicates the presence of some complex dynamics in the DHF incidence related to nonlinear interactions among different modes at all scales. A reconstruction $I_N(r_n,t)$ of the field, as $N$ is varied, has been performed according to Eq. (4). Four reconstructed fields are shown in Figs. 4–7. The field $I_0(r_n,t)$ reconstructed using only the first $j=0$ POD mode gives us the spatiotemporal evolution of the average DHF incidence. Further POD modes add stochasticity to the system, the basic temporal periodicity being perturbed.

In order to investigate the spatial synchrony of the various incidence functions, we analyze the cross correlation of reconstructed DHF incidence at different truncation order $N$ and different lag times $\tau$,

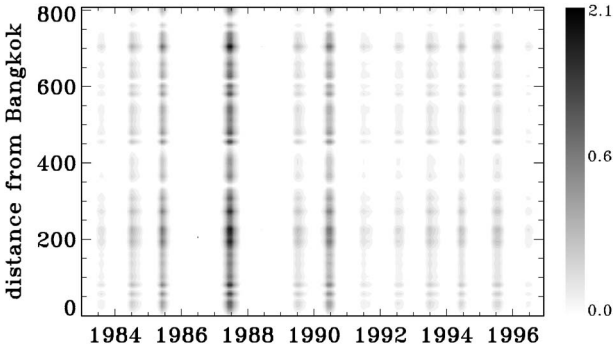


FIG. 4. Contours of the reconstruction of the incidence field by using $N=0$ in the time-space plane. The reconstructions have been made by using relation (4).

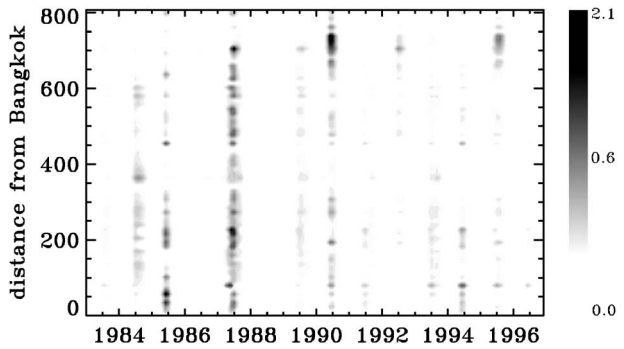


FIG. 5. Contours of the reconstruction of the incidence field by using $N=4$ in the time-space plane. The reconstructions have been made by using relation (4).

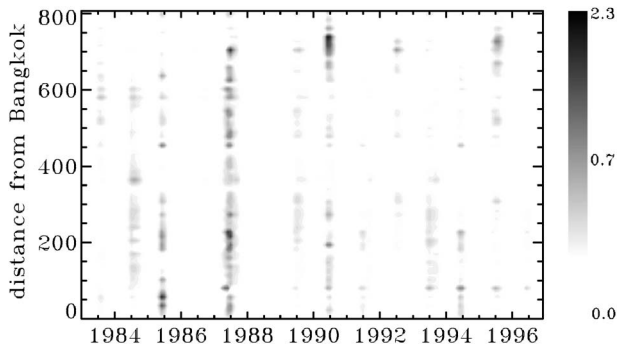


FIG. 6. Contours of the reconstruction of the incidence field by using $N=6$ in the time-space plane. The reconstructions have been made by using relation (4).

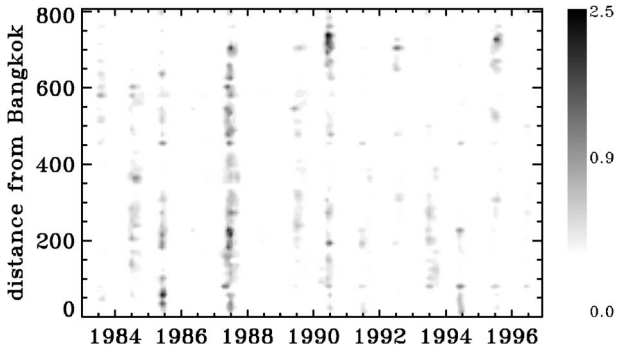


FIG. 7. Contours of the reconstruction of the incidence field by using $N=24$ in the time-space plane. The reconstructions have been made by using relation (4).

$$C_N(r_n,\tau)=\frac{\langle F_N(0,t)F_N(r_n,t+\tau)\rangle}{\sqrt{\langle F_N(0,t)^2\rangle}\sqrt{\langle F_N(r_n,t+\tau)^2\rangle}} \quad (5)$$

(the angular brackets mean time averages). The cross correlation performed using only the first POD mode $N=0, F_N(r_n,t)=I_0(r_n,t)$ (see Fig. 8 upper panel), shows a temporal periodicity of one year but does not show spatial differences among the different provinces. This means that, on average, almost periodic traveling pulses of epidemic are felt at the same time across all provinces. This is the basic POD mode. By averaging the cross correlation $C_N(r_n,\tau)$ over all provinces $r_n$ we get the function $C_N(\tau)$ reported in the lower panel of Fig. 8. Even though the yearly periodicity is well visible, the average cross correlation is greater than 0.6 for $\tau=3$ and 5 yr. This last result is perhaps very interesting because it suggests looking for the presence of aperiodic pulses with different energy content.

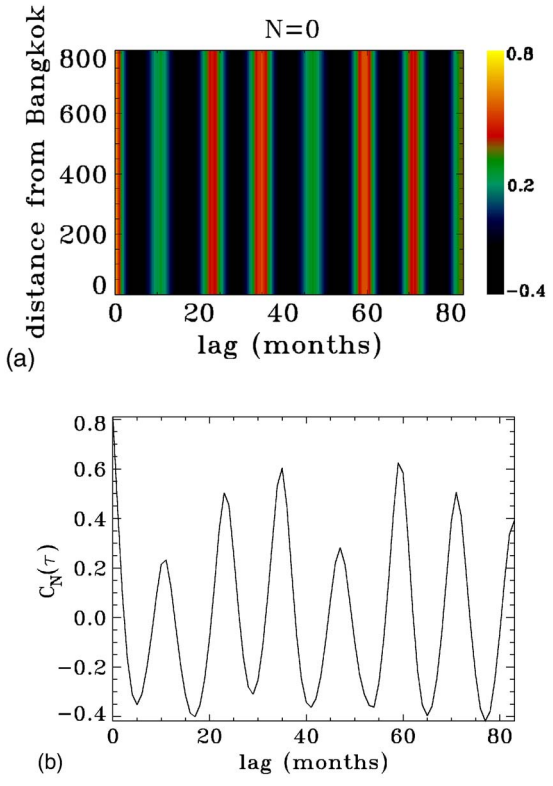


FIG. 8. (Color online) Contours of the cross-correlation coefficients $C_0$, calculated using only the field $I_0(r_n,t)$, in the $(r_n,\tau)$ plane (see text) between Bangkok and the other provinces of Thailand (upper panel). Averaged cross correlation $C_N(\tau)$ as a function of the lag time $\tau$ (lower panel).

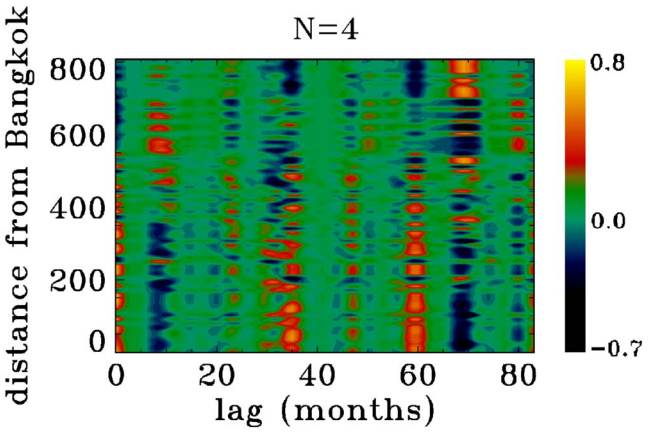


FIG. 9. (Color online) Contours of the cross-correlation coefficients $C_4(r_n,\tau)$ in the $(r_n,\tau)$ plane (see text) between Bangkok and the other provinces of Thailand. The coefficients have been calculated by using the fluctuating fields $F_4(r_n,t)$.

In order to get information about the presence of pulses with different temporal and spatial properties, we use the fluctuation POD modes $F_N(r_n,t)=I_N(r_n,t)-I_0(r_n,t)$ for different truncation order. In Figs. 9–11 the cross-correlation functions $C_N(r_n,\tau)$ for three different truncation orders $N$ are shown. For each $N$ the function $C_N(r_n,\tau)$ presents a maximum for more than one time lag. This indicates the presence of energetic fluctuations of the DHF epidemic that, starting from Bangkok, invades the farther provinces [11]. The $N=4$ truncation of fluctuation modes is different from the other, because it shows two maxima for lag times of about $\tau\simeq 3$ and 5 yr; the basic yearly periodicity seems to be lost. When the truncation order is increased and further POD modes are included in the reconstruction (4), equispaced maxima for the cross correlation appear. The yearly periodicity becames again the dominant trend.

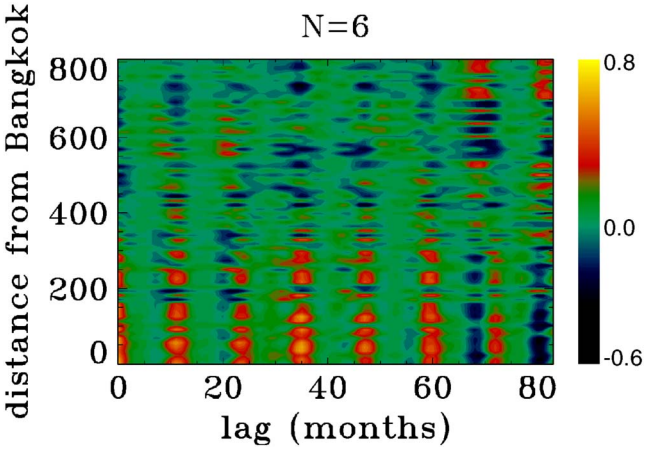


FIG. 10. (Color online) Contours of the cross-correlation coefficients $C_6(r_n,\tau)$ in the $(r_n,\tau)$ plane (see text) between Bangkok and the other provinces of Thailand. The coefficients have been calculated by using the fluctuating fields $F_6(r_n,t)$.

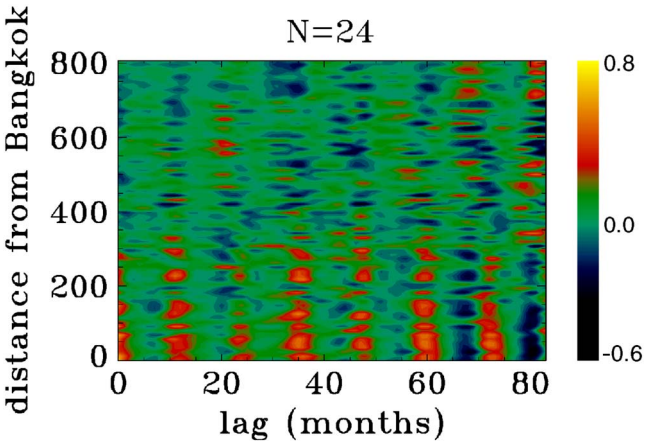


FIG. 11. (Color online) Contours of the cross-correlation coefficients $C_{24}(r_n,\tau)$ in the $(r_n,\tau)$ plane (see text) between Bangkok and the other provinces of Thailand. The coefficients have been calculated by using the fluctuating fields $F_{24}(r_n,t)$.

Looking at Figs. 9–11 we can realize that the spatial extension of the cross correlation of fluctuations depends on the truncation order $N$. In particular, by comparing the $N=4$ and 6 truncations (Figs. 9 and 10, respectively), one gets the feeling that the spatial coherence of the pulses identified by the $N=4$ truncation is greater than that identified by the $N=6$ truncation. In other words the spatial extension from Bangkok of aperiodic epidemic pulses looks to be greater than the spatial extension of the periodic waves obtained with truncation of higher-order $N$. To quantify the spatial extension of traveling waves we define a spatial coherence length $R_N$ as the average

$$D(r_j) = \left( \frac{1}{j} \sum_{n=0}^{j} \int d\tau\, C_N(r_n,\tau) \right).$$

This function, when plotted against $r_j$, has a parabolic shape. The distance $r_j^{(max)}$, corresponding to the maximum of the curve $D(r_j)$, measures the averaged spatial extension of the cross correlation $R_N = r_j^{(max)}$. Thus the spatial coherence is

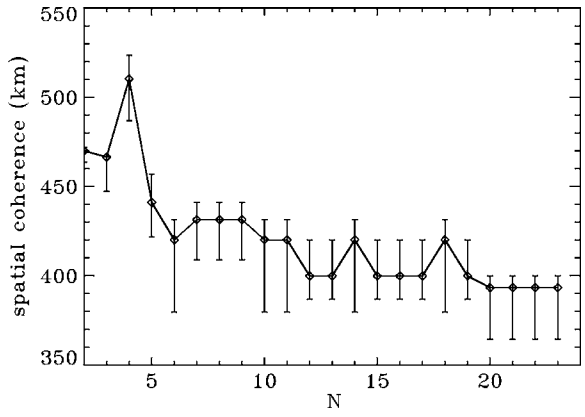


FIG. 12. Spatial coherence as a function of the number of fluctuating modes used in the reconstruction of the fluctuation field. The error bars correspond to the nearest provinces from the maximum (see text).

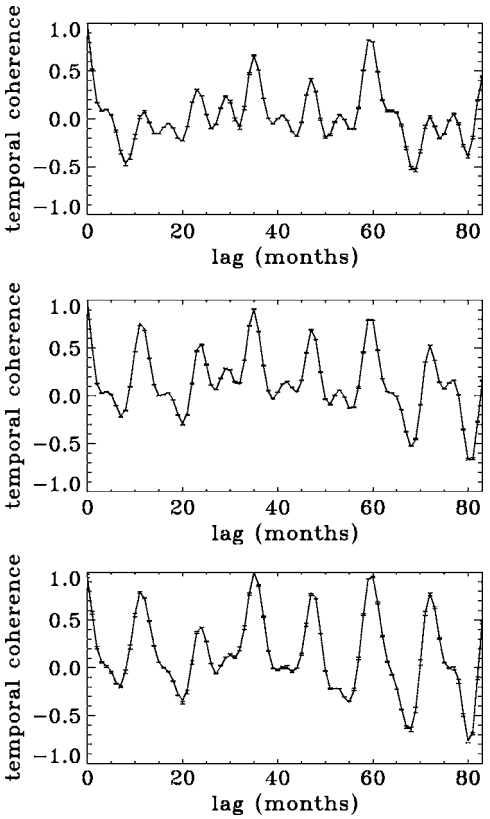


FIG. 13. Time coherence for different approximated fields, namely, $N=4$ (top panel), 6 (middle panel), and 24 (bottom panel). The error bars are calculated as in Fig. 12 (see text).

defined as the distance $r$ of the province corresponding to the maximum of the time averaged cross correlation. In Fig. 12 the spatial extention $R_N$ as a function of the number of fluctuating modes used in the reconstruction is shown. For each $N$ both the two nearest provinces to the maximum define the error bars $\Delta r_j^{\pm}$, namely, both $\Delta r_j^{+} = r_{j+1} - r_j^{(max)}$ and $\Delta r_j^{-} = r_j^{(max)} - r_{j-1}$. The larger spatial extension of about 510 km, corresponds to the fluctuation field reconstructed by using $N=4$ modes. Reconstruction involving $N>4$ fluctuation modes shows a lower spatial extension of about 420 km. Then aperiodic pulses, characterized by a stronger degree of infection, start from Bangkok at some time and move away coherently for a large distance.

To measure the characteristic times corresponding to the DHF occurrences, we define a kind of temporal coherence $T_N$ built up by averaging the cross correlation in space up to $R_N$, namely,

$$T_N(\tau) = \sum_{r_n \leq R_N} C_N(r_n,\tau).$$

In Fig. 13 we report the temporal coherence calculated for $N=4$, 6, and 24, the error bars being defined by summing up to $R_N \pm \Delta r_j^{\pm}$, respectively. For $N=4$ two peaks corresponding to about 3 and 5 yr dominate the other. On the contrary, when $N=6$ or 24, the one-year periodicity becames the dominant feature.

## V. CONCLUSIONS

In conclusion we applied POD analysis within a biomedical context to spatiotemporal data of DHF occurrence in Thailand. We showed that, at variance with the Fourier analysis, POD is able to capture the main energetic and spatial features of the spatiotemporal evolution of the epidemic field. A reconstruction of the stochastic part of the epidemic field is obtained by varying the number of POD modes, and spatiotemporal information about the propagation of epidemic pulses from Bangkok toward the farther provinces has been extracted from the data set. Apart from the basic yearly periodicity, we found that a reconstruction using $N=4$ POD modes mainly describes aperiodic pulses, peaked at 3 and 5 yr, of DHF occurrence propagating away from Bangkok to the outer provinces. These pulses, which are superimposed on the basic yearly cycle, remains coherent up to 510 km from Bangkok. When the number of POD modes used in the reconstruction is increased, the level of fluctuation increases as well and the yearly periodicity, probably related to seasonality (e.g., the monsoon), became the dominant feature of the epidemic. In this case the spatial extension is about 420 km. While the periodic occurrence of the epidemic is well known and in some sense expected, large and unpredictable epidemics of DHF [24] must be considered as a purely occasional increase of DHF. However, POD analysis shows that these episodic pulses have a coherence of 3 or 5 yr. Up to now the data set is too poor to firmly establish a recurrence for high-energetic pulses, or to establish their stochastic nature.


[1] P. Turchin, *Complex Population Dynamics: A Theoretical/Empirical Synthesis* (Princeton University Press, Princeton, NJ, 2003).

[2] M. Kot, *Elements of Mathematical Ecology* (Cambridge University Press, Cambridge, U.K., 2001).

[3] A. J. Lotka, *Elements of Physical Biology* (Williams and Wilkins, Baltimore, 1925).

[4] V. Volterra, Atti Accad. Naz. Lincei, Cl. Sci. Fis., Mat. Nat., Rend., Mat. Nat., Rend. **2**, 13 (1926) (Italian), translated in R. N. Chapman, *Animal Ecology* (McGraw-Hill, New York, 1931), p. 409.

[5] G. J. Ackland and I. D. Gallagher, Phys. Rev. Lett. **93**, 158701 (2004).

[6] J. A. Sherratt, Ecol. Lett. **4**, 30 (2001).

[7] E. Gilad, J. von Hardenberg, A. Provenzale, M. Shachak, and E. Meron, Phys. Rev. Lett. **93**, 098105 (2004).

[8] C. W. Eurich, A. Thiel, and L. Fahse, Phys. Rev. Lett. **94**, 158104 (2005).

[9] O. N. Bjørnstad, R. A. Ims, and X. Lambin, Trends Ecol. Evol. **14**, 427 (1999).

[10] E. Ranta and V. Kaitala, Nature (London) **390**, 456 (1997).

[11] D. A. T. Cummings, R. A. Irizarry, N. E. Huang, T. P. Endy, A. Nisalak, K. Ungchusak, and D. S. Burke, Nature (London) **427**, 344 (2004).

[12] D. J. Gubler, Clin. Microbiol. Rev. **11**, 480 (1998).

[13] D. M. Watts, D. S. Burke, B. A. Harrison, R. E. Whitmire, and A. Nisalak, Am. J. Trop. Med. Hyg. **36**, 143 (1987).

[14] W. Tun–Lin, T. R. Burkot, and B. H. Kay, Med. Vet. Entomol. **14**, 31 (2000).

[15] N. Ferguson, R. Anderson, and S. Gupta, Proc. Natl. Acad. Sci. U.S.A. **96**, 790 (1999).

[16] S. I. Hay *et al.*, Proc. Natl. Acad. Sci. U.S.A. **97**, 9335 (2000).

[17] D. J. Gubler and L. J. Rosen, J. Med. Entomol. **13**, 469 (1977).

[18] L. Rosen, L. E. Roseboom, D. J. Gubler, J. C. Lien, and B. N. Chaniotis, Am. J. Trop. Med. Hyg. **34**, 603 (1985).

[19] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge University Press, Cambridge, U.K., 1996).

[20] G. Alfonsi and L. Primavera, J. Flow Visualization Image Process. **9**, 89 (2002).

[21] V. Carbone, F. Lepreti, L. Primavera, E. Pietropaolo, F. Berrilli, G. Consolini, G. Alfonsi, B. Bavassano, R. Bruno, A. Vecchio, and P. Veltri, Astron. Astrophys. **381**, 265 (2002).

[22] P. D. Mininni, D. O. Gómez, and G. B. Mindlin, Phys. Rev. Lett. **89**, 061101 (2002).

[23] A. Vecchio, V. Carbone, F. Lepreti, L. Primavera, L. Sorriso–Valvo, P. Veltri, G. Alfonsi, and T. Straus, Phys. Rev. Lett. **95**, 061102 (2005).

[24] D. DeRoeck, J. Deen, and J. D. Clemens, Vaccine **22**, 121 (2003).