



**HAL**  
open science

# The vegetation cycle in West Africa from AVHRR-NDVI data: horizons of predictability versus spatial scales

Sylvain Mangiarotti, P. Mazzega, Pierre Hiernaux, Éric Mougin

## ► To cite this version:

Sylvain Mangiarotti, P. Mazzega, Pierre Hiernaux, Éric Mougin. The vegetation cycle in West Africa from AVHRR-NDVI data: horizons of predictability versus spatial scales. *Remote Sensing of Environment*, 2010, 114 (9), pp.2036-2047. 10.1016/j.rse.2010.04.010 . ird-01062699

**HAL Id: ird-01062699**

**<https://ird.hal.science/ird-01062699>**

Submitted on 10 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The Vegetation Cycle in West Africa from AVHRR-NDVI data: Horizons of Predictability *versus* Spatial Scales

Mangiarotti<sup>a,b,1</sup> S., Mazzega<sup>b</sup> P., Hiernaux<sup>a</sup> P. and Mougin<sup>a</sup> E.

<sup>a</sup> *Centre d'Études Spatiales de la Biosphère, CNES CNRS IRD Université P. Sabatier Toulouse III, OMP, 18 av. E. Belin, bpi 2801, 31401 Toulouse cedex 9 (France)*

<sup>b</sup> *Laboratoire des Mécanismes et Transferts en Géologie, CNRS IRD Université P. Sabatier Toulouse III, OMP, 14 avenue E. Belin, 31400 Toulouse (France)*

## Abstract

The predictability of the vegetation cycle is analyzed as a function of the spatial scale over West Africa during the period 1982-2004. The NDVI-AVHRR satellite data time series are spatially aggregated over windows covering a range of sizes from 8x8 km<sup>2</sup> to 1024x1024 km<sup>2</sup>. The time series are then embedded in a low-dimensional pseudo-phase space using a system of time delayed coordinates. The correlation dimension ( $D_c$ ) and entropy of the underlying vegetation dynamics, as well as the noise level ( $\sigma$ ) are extracted from a nonlinear analysis of the time series. The horizon of predictability ( $H_p$ ) of the vegetation cycle defined as the time interval required for an  $n\%$  RMS error on the vegetation state to double (i.e. reach  $2n\%$  RMS) is estimated from the entropy production. Compared to full resolution, the intermediate scales of aggregation (in the range 64x64 km<sup>2</sup> to 256x256 km<sup>2</sup>) provide time series with a slightly improved signal to noise ratio, longer horizon of predictability (about 2 to 5 decades) and preserve the most salient spatial patterns of the vegetation cycle. Insights on the best aggregation scale and on the expected vegetation cycle predictability over West Africa are provided by a set of maps of the correlation dimension ( $D_c$ ), the horizon of predictability ( $H_p$ ) and the level of noise ( $\sigma$ ).

*Keywords: vegetation cycle, West Africa, horizon of predictability, spatial scale, NDVI satellite data, nonlinear data analysis.*

---

<sup>1</sup> To whom correspondence should be sent: sylvain.mangiarotti@cesbio.cnes.fr

## 1. Introduction

The distribution of the vegetation is structuring large scale ecological patterns in West Africa. From about 20°N to the shores of the gulf of Guinea in the south, the biogeographical arrangement is dominantly zonal with the succession from Sahara to the Sahel and the Sudanian belts, and finally the Guinean forest (Auberville, 1949). The variations in timing of the primary production are also contrasted from the steppes of southern Sahara, in the north, to the grasslands in Sahel, the savannas in the Sudanian belt and the tropical moist forests to the south. Unsurprisingly, the spatial distribution and annual cycle of the vegetation are strongly related to the spatial and temporal patterns of precipitations. The interannual variability of the vegetation cycle is also linked to a set of climate signals (Jarlan et al., 2005).

From a societal point of view the natural vegetation is a resource to feed livestock, to manage soil fertility in cropland and provide forestall goods and services. Thus, the inter-annual variability of primary production impacts on a range of economical or social assets in rural areas and indirectly affects urban population. Dry spells or extreme events like regional droughts deeply alter primary production and may even durably change ecosystem components and functioning. These environmental perturbations in turn affect human activities in many ways including through crop and livestock productions, and it possibly affect lifestyles as observed during the major droughts of the 70's and 80's (Thébaud, 2002). For these reasons reliable predictions of the vegetation state, from local to regional scales, are of interest in conjunction with the operational policies aiming the remediation of the socio-economical impacts of environmental stresses or disturbances. Time series of AVHRR NDVI data have been analysed in many contexts with various aims such as characterizing land use, assessing crop and forage yields, or analysing intra- and interannual variations of vegetation attributes (*e.g.* Jakubauskas et al., 2002; Alhamad et al., 2007; Martínez & Gilabert 2009).

In the present study the AVHRR NDVI values are used as a proxy for the vegetation state. The intrinsic limitations set by the vegetation dynamics to the prediction of its variability are estimated at a range of spatial scales. The main objective is to produce maps of the horizon of predictability, *i.e.* the time window over which a reliable prediction can be done, as a function of the scale of spatial aggregation. In Section 2 the rationale for the approach and the data pre-processing are given, including the data aggregation scheme. In Section 3, the extraction of useful information from data time series is described and the derivation from these of an estimate of the predictability of the vegetation cycle is explained. The geographical distributions of the estimated horizons of predictability associated to a range of spatial scales are presented and commented in Section 4. The reliability of the estimates of invariant characteristics of the vegetation cycle, including the horizons of predictability, is discussed in Section 5. In the last Section conclusions are drawn, and future line of research dedicated to the effective prediction of vegetation cycles are proposed. A brief mathematical appendix is added providing more detailed descriptions of the theory and algorithms used in this study.

## 2. NDVI Data and the Spatial Aggregation Scheme

The Normalized Difference Vegetation Index (NDVI) calculated for reflectances in the visible and near infrared bands measured from satellite is used to map estimates of the primary production in West Africa. The estimation process either rely on empirical relationships established by regression between NDVI data and vegetation yield data measured on the field (Tucker et al., 1985), or else on satellite data assimilation coupling a model of the vegetation production and a model of the vegetation radiative transfer (see *e.g.* Jarlan et al., 2008, Mangiarotti et al., 2008). By provision, the capacity to convert

NDVI into biomass, at least through empirical regressions, is assumed. Hence, NDVI time series are analysed to evaluate the predictability of the amplitude and spatial distribution of the vegetation cycle.

However, NDVI data at high spatial resolution are likely associated with high level of noise relative to signal, and possibly with stochastic-like patterns of vegetation dynamics mainly driven by local environmental conditions (such as spatial redistribution of rainfall, soil depth and texture, cropping practices, grazing pressure, etc.). In order to give some objective basis to this hypothesis, a simple spatial aggregation scheme was developed to analyze how the predictability of the NDVI time series changes with scale aggregation, in search for the optimal compromise between predictability and spatial resolution.

### **a. The Data and their Preprocessing**

The NDVI data from the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) sensor are used here over the years 1982 to 2004 at the raw resolution [ $8 \times 8 \text{ km}^2$ ]. The source of the Global Modeling and Mapping Studies (GIMMS) data is the Global Land Cover Facility (Tucker et al., 2005). These data have been composited over 10-day periods using a maximum value NDVI technique for reducing the atmospheric and view angle effects (Holben, 1986). The data set accounts for sensor degradation through a technique based on stable desert targets. It also includes a calibration and some corrections for the geometry of the view and for the effects of the stratospheric aerosols associated with the eruptions of the El Chichon (April 1982) and of the Mt Pinatubo (June 1991). Onboard acquisition is 8 bits coded, which leads to a high quantization of the signal.

The nonlinear data analysis method (Sec. 3) requires long and complete time series, hence gaps in the data set were filled as follows: (1) when a single datum is lacking, a spline interpolation is applied combined with a simple average of the nearest neighbors; (2) when two successive data are lacking, a weighted running average of the 5 nearest neighbors is applied; (3) for a gap of 3 or more successive data, the series were filled using an average of the nearest spatial neighbors; if (4) no data are immediately in the spatial neighbor of a gap, a Gaussian weighted average of the data available in a 40 km radius (excluding the pixels over the ocean) is used instead.

### **b. The Aggregation Scheme**

After the correction and interpolation processes, maps with no gaps are available every ten days from 1982 to 2004, covering most of West Africa at an  $8 \times 8 \text{ km}^2$  resolution. The pixels over the oceanic areas are discarded. The spatial aggregation scheme is strait forward: a tessellation of any decadal map is generated with non overlapping square windows of  $N \times N$  pixels. NDVI data are averaged in each window and the mean is geographically attributed to the center of the window. For convenience, the window size is referred to by the aggregation index  $I_A$  that designates a window with  $[2^{(I_A-1)} \times 2^{(I_A-1)}]$  pixels ( $1 \times 1$  pixel for  $I_A = 1$ ,  $2 \times 2$  pixels for  $I_A = 2$ , etc.). The larger aggregation window used in this study has the index  $I_A = 8$  ( $128 \times 128$  pixels and a spatial size of  $1024 \times 1024 \text{ km}^2$ ). The scene of the whole region is referred to as the ‘West African window’. The number of averaged pixels, area and number of non overlapping windows as a function of the aggregation index  $I_A$  are given in Table 1. Only 11 (resp. 2) windows being formed over West Africa with aggregation index  $I_A = 7$  (resp.  $I_A = 8$ ), the characteristics extracted from the corresponding aggregated time series must be considered with much caution when a statistical interpretation is sought for.

As  $I_A$  is increasing the aggregation process mixes together time series corresponding to different rainfall regimes, topographical features, climate influences and of course vegetation and terrestrial ecosystem types. The stronger contrasts among the time series contributing to the same window average are thus found in the sampling of the larger windows. Moreover, the number of independent series rapidly decreases with the square of the aggregation index. However, this aggregation scheme has three advantages: firstly, it can be readily associated to a spatial scale or resolution; secondly, any aggregation window corresponds to a well defined geographical area; thirdly, it is designed without using *a priori* knowledge on precipitations, soils and vegetation dynamics. As will appear below, these properties are used all along the interpretation of the results. Although the analysis extends over the whole region of West Africa, one part of the study focuses on the Agoufou field site of the AMMA program (15.3°N, 1.5°W, see zone N; Mougin et al., 2009). The Agoufou site is located in the Sahel belt and is composed of fixed sand dunes with a relatively dense herbaceous and shrub coverage, and a small percentage of bare soil patches. Several other sites of particular interest are localized on the Figure 1: rivers, lakes and wet lands, tropical forests, mountains, etc.

### **c. Aggregation and Vegetation Cycles**

The effect of the aggregation process are illustrated (Figure 2) by comparing three time series centered on the Agoufou site (Gourma) with aggregation indexes  $I_A = 1$  ( $8 \times 8$  km<sup>2</sup>),  $I_A = 5$  ( $128 \times 128$  km<sup>2</sup>) and  $I_A = 8$  ( $1024 \times 1024$  km<sup>2</sup>). Spatial aggregation appears to reduce the amplitude of the vegetation cycle. It also smoothes the interannual variability. Indeed, in the large scale window (Fig.2c;  $I_A = 8$ ) the cycles are very regular from year to year. Even during the severe droughts of the years 1983-1987, very low NDVI values, corroborated by the low yields measured on the field, (Hiernaux et al., 2009) vegetation cycles observed at large scale hardly distinguish from those of normal years. Both, the arid steppes to the north and the sub-humid savannas to the south of the window are less affected by droughts than are the sahelian grasslands where is located the Agoufou site. Consequently, at large scale the averaging process over wide window dampers the sahelian signal associated to severe droughts. Similar arguments hold in 1999 and 2001, for which high NDVI values matching the high yields and cumulative rainfall measured on the field (Hiernaux et al., 2009) observed at Agoufou (Fig. 2a) and over the whole Gourma (Fig. 2b), are not detectable on the cycles of the large regional window (Fig. 2c). Some noise is also expected to affect the NDVI values at the [ $8 \times 8$  km<sup>2</sup>] resolution (a single AVHRR pixel; Fig. 2a). This noise results from several sources: the instrumental noise, the residuals of the applied corrections, local environmental conditions (local rainfall effect on the vegetation, clouds, etc.). This noise is supposed to be largely uncorrelated between nearby pixels so that spatial averages filter the noise and improve the signal to noise ratio in the aggregated time series. Some quantitative support is given in the next sections to these hypotheses.

## **3. Estimating the Predictability of the Vegetation Cycle**

The prediction of a perfectly sinusoidal annual cycle would be trivial: considering the value of NDVI at one decade of any year in the past would be enough to predict the value for the same decade of any new year. In such case the horizon of predictability would be infinite: any future value can be exactly predicted now, considering the past observations. Looking at the NDVI time series in Figure 2, the prediction does not appear so easy: indeed, the cycle is not regular, the amplitude, possibly phase and secondary maxima change from one year to the next. Years 1984 to 1987, or else 1994, 1995, 2000 and 2003,

present similar amplitudes and timing pattern of the annual dynamics of the vegetation, at least at the 8x8 km<sup>2</sup> scale in Agoufou (Fig. 2a). In the language of dynamical system, the system is experiencing similar states. In some abstract (geometrical) metric space the system trajectory is visiting neighbour states during these non-necessarily successive years. Thus the process of prediction at an early stage of the season for 1987 (resp. 2003) may benefit from the observed vegetation cycles in the years 1984 to 1986 (resp. 1994, 1995, 2000).

The nonlinear analysis of time series aims at assessing the intrinsic dynamical limitations to the predictability of NDVI cycles. The most expressive quantity to be determined is the horizon of predictability  $H_p$  that indicates the number of decades necessary for an error of  $n\%$  RMS on the system state to double and reach  $2n\%$  RMS. If the NDVI data of a given year and season are diverging from the usual values for that season, the prediction capacity for the following seasons will be limited (low values of  $H_p$ ). On the contrary, in an average year the predictability potential is high and the evolution of the NDVI is likely to be reliably predicted several decades in advance. We now present the minimum background required to give some operational virtue to these considerations, the cited references providing full theoretical details to the interested reader.

### ***a. Embedding of the NDVI Time Series***

The NDVI time variations being dominated by an irregular annual cycle, it is proposed to embed the time series in a pseudo-phase space of dimension  $D_e > 1$ . Such an embedding is illustrated in Figure 3 (following a method being explained below). In this space, the differences of the vegetation cycle from year to year correspond to local divergences in the trajectory of the system state. The geometrical shape associated to these differences – the attractor of the system – looks like a noise trajectory within a torus, possibly deformed.

These geometrical and statistical views on dynamical systems (Wiggins, 1990) are woven into the ergodic theory (Lasota & Mackey, 1994). The system attractor – when it actually exists – is the support of the probability density function of the system states in the proper multi-dimensional space. But in practice, one does not have measurements of all the system dependent variables that are required to fully specify the state of the vegetation. This state is only apprehended through data time series or spatiotemporal maps of observable variables, like the NDVI that are linked to bio-physical variables (e. g. leaf area index, vegetation yield...) of interest through intricate relationships or models.

Fortunately, Takens (1981) (see also Packard et al., 1980) has shown that under the ergodic assumption, some information can be retrieved about the underlying nonlinear system dynamics from the time series of a single variable. One basic condition for this theorem to apply is that this observation variable must be related to some state variables through a continuous and differentiable transformation (a diffeomorphism). This information is synthesized in the form of a set of scalars that characterizes the statistical and geometric properties of the underlying unobservable system attractor and dynamics. In particular the horizon of predictability of the system state is related to the rate of entropy production and to the dynamical development of instabilities. These scalars are invariant under the diffeomorphic transformations of the system state space. In summary, some fundamental properties of the vegetation dynamics can be retrieved from the nonlinear analysis of the NDVI data time series, as shown below.

## b. Nonlinear Data Analysis

The nonlinear data analysis consists in estimating various dimensions and scalar invariants associated to the NDVI time series. The main steps performed are: (step 1) embed the data time series in a  $D_e$ -dimensional ( $>1$ ) pseudo-phase space and reconstruct the local attractor of the vegetation dynamics (see e.g. Figure 3); (step 2) extract various geometric (fractal dimensions) and statistical (e.g. signal to noise ratio) estimates associated to this attractor; (step 3) estimate the horizon of predictability and additional information on the predictability of the vegetation cycle. The realization of these steps are illustrated by the analysis of the three NDVI time series centered in Agoufou with aggregation indexes  $I_A = 1$ ,  $I_A = 5$  and  $I_A = 8$  delineated in Figure 2. An overview of this method is given by Abarbanel (1996) and Diks (1999). Several difficulties related to the analysis of real, finite, noisy geophysical data, and ways to minimize the noise effects are presented by Frede and Mazzega (1999a,b).

The embedding of time series (step 1) is performed using a coordinate system of statistically independent delayed variables. In a one-to-one correspondence these variables associate to any time series  $x = [x_1, x_2, \dots, x_{nd}]$  a vector time series  $y = [y_1, y_2, \dots, y_{nd-De+1}]$ , each single vector is defined by:

$$y_j = [x_j, x_{j+\tau}, x_{j+2\tau}, \dots, x_{j+(De-1)\tau}] \quad (1)$$

In this expression the integer time delay  $\tau$  is estimated as the delay corresponding to the first local minimum of the averaged mutual information function  $AMI(\tau)$  (Figure 4, first column; see also the Appendix). An alternative approach would consist in choosing for  $\tau$  the delay associated to the decrease of the autocorrelation of the times series by a factor  $\exp(-1)$  (e-folding). However, it has been shown that the  $AMI$  function, being built from probability distributions of the data, is more adapted to time series issued from nonlinear systems (Fraser & Swinney, 1986) that often present correlated events unevenly distributed along the time axis. Nevertheless, nearly same values of the integer time delay  $\tau$  were observed with the AVHRR NDVI time series, at any aggregation scale, irrespective of the  $AMI$  or the autocorrelation functions. As an example, the time series of Figure 2a is embedded in a 3-dimensional space in Figure 3, using a time delay  $\tau = 10$  decades.

The embedding dimension  $D_e$  is determined with a global false neighbor algorithm (Kennel & Abarbanel 2002). The objective is to find the lowest dimension that allows the unfolding of the reconstructed attractor. The principle of the algorithm is as follows. Two distant points on a plane (two dimensional space) can be projected in a very close neighborhood on an axis (one dimensional space) if the projection is nearly parallel to the segment joining the two points. Conversely, the two neighboring points on the axis will no more be neighbors when adding the second dimension: they were false neighbors in dimension 1 because of the projection. Similarly, two nearby points in a 2-dimensional space may reveal to be projections of distant points in a 3-dimensional space, and so on. The global false neighbor algorithm counts the number of neighbor data vectors  $y_j$  that reveal to be false neighbors when adding new dimensions (that is new delayed data coordinates in eq.1). The embedding dimension is found when no more false neighbors are left: when the attractor is fully unfolded, two data vectors are neighbors because they correspond to close system states, not as a result of projection effects. One difficulty is occurring with real data because the noise component tends to increase the embedding dimension (noise is infinite dimensional).

In Figure 4 (third column) there is a good convergence of the algorithm to embedding dimensions  $D_e$  of 3 or 4 for the NDVI time series over the Gourma region. It should be

noted that both the power spectra and the averaged mutual information functions obtained for the time series with respective aggregation indexes  $I_A = 1$  and  $I_A = 5$  are very similar. The aggregation on a local set of  $16 \times 16$  pixels ( $I_A = 5$ ) is preserving most of the data signal. The lower embedding dimension ( $D_e = 3$ ) is found for the large aggregation window ( $I_A = 8$ ) which power spectrum presents no significant long period tail and a small semi-annual cycle (compare the panels of the 2<sup>nd</sup> column in Fig. 4). Over the whole West Africa window, the time delay  $\tau$  slightly varies from 6 to 10 decades as a function of the latitude (Figure 5) whereas almost no variation is observed as a function of the longitudes. These longitudinally averaged values for the delay are used in the operational processing of the data. Very similar patterns are obtained at different scales up to  $18^\circ\text{N}$ . The very low signal and high level of noise (especially at low aggregation scales) explain the observed differences when considering the arid area at higher latitudes. The averaged embedding dimension  $D_e$  is shown as a function of the latitude in Figure 6. Between  $5^\circ\text{N}$  and  $13.5^\circ\text{N}$ , the embedding dimension  $D_e$  equals three in most of the cases whatever the aggregation index is. Between  $14^\circ\text{N}$  and  $16^\circ\text{N}$ , both  $D_e = 3$  and  $D_e = 4$  are found (up to 50% each). For low levels of aggregation (Fig. 6a), a high percentage of  $D_e = 2$  is found at high latitude, a value that clearly results from the low signal to noise ratio. Indeed no irregular cycle can be embedded in a two dimensional space (otherwise the trajectory would cross itself). In summary, the value  $D_e = 4$  guarantees that the reconstructed attractors are unfolded at all latitudes and aggregation scales, and thus this value is used in the operational processing of the NDVI data.

### ***c. Entropy Production and Predictability of the NDVI***

The local divergence of several segments of the reconstructed attractor (see Fig. 3) indicates that some instability is developing in the pseudo-phase space. In other words, the vegetation cycle is reaching some states associated with a high rate of entropy production, thus reducing the intrinsic predictability of the system evolution. On the contrary, when the segments of the trajectory are locally converging, the predictability of the vegetation cycle is high. However, the system never experiences exactly the same values of all its state variables. The way the trajectory fills the  $D_e$ -dimensional phase space is characterized with the correlation dimension  $D_c$ . Strange attractors associated with nonlinear dynamics have fractal correlation dimensions: the segments of the attractor tend to fulfill the (hyper-) planes locally orthogonal to the system trajectory in an organized, self-similar way. The celebrated algorithm proposed by Grassberger and Procaccia (1983) allows for evaluating the correlation integral which is parameterized by the correlation dimension  $D_c$ . The measure associated to the local distribution of the segments of the attractor depends on the resolution of the analysis: at very large scale (in the embedding phase space) the diameter of the attractor is merely visible; at high resolution (small radius  $h$ ) the analysis reveals the organization of the trajectory in phase space and leads to accurate estimates of the correlation dimension. However, the increase in resolution is limited by the quantization of the NDVI data (8 bits coding).

The original algorithm (Grassberger and Procaccia, 1983) is quite sensitive to the data spoiling by noise and to the finite length of the time series (Ding et al., 1993). Diks (1996) has proposed to use Gaussian kernels for estimating the correlation integral: this approach is much less sensitive to the noise level. Under the assumption of an additive Gaussian noise, this formula of the correlation integral is parameterized not only by the correlation



dimension of the attractor  $D_c$  but also by the correlation entropy  $K$  and by the signal to noise ratio  $\sigma$  (estimated as a % RMS) (Diks, 1999; see the Appendix). A method has been designed that extracts  $D_c$ ,  $K$  and  $\sigma$  from the data time series via the estimation of the correlation integral (Yu et al., 2000) and is used here. The correlation entropy  $K$  is a measure of the rate of development of instabilities in the vegetation cycle and is thus related to the intrinsic predictability of the system evolution. So it is proposed to estimate the horizon of predictability with the following function:

$$H_p = \ln(2)/K \quad (2)$$

With this definition  $H_p$  (in decades) represents the time interval required for a small perturbation of the vegetation state to increase by a factor two. The horizon of predictability is itself a stochastic variable that changes with time. Only horizons averaged over the time series will be estimated.

The correlation integral has been computed for each NDVI time series and each aggregation scale. An example based on the time series (with different aggregation indexes) centred on the Agoufou site is represented in Figure 7. Thanks to the Gaussian kernel algorithm, the correlation integral is rather smoothed but the convergence of the curve when increasing the embedding dimension is partly hidden by the effect of noise. This convergence appears a bit more clearly for  $I_A = 8$  when comparing  $D_e = 4$  and  $D_e = 5$  which are quite close. Then, the estimates of the correlation dimension  $D_c$ , horizon of predictability  $H_p$ , and noise level  $\sigma$ , are plotted as a function of the embedding dimension  $D_e$  in Figure 8. Due to the finite information of the series (22 years length, 10-day sampling, 1/256 quantization and additive noise), the convergence of the results is not reached when increasing the embedding dimension. Longer time series are clearly needed. However, following Yu et al. (2000) more reliable estimates of  $D_c$ ,  $H_p$  and  $\sigma$  are extracted by averaging the results obtained with higher embedding dimensions (see Table 2). This was performed with embedding dimension  $D_e = 3$  to 5. The associated error is related to the range of the estimated values. This method is used in routine in the following of this study.

#### 4. Scale-dependent horizons of predictability in West Africa

First, the results at different aggregation scales over the Gourma are analysed. Then, are presented and discussed the maps of the horizons of predictability of the vegetation cycle over West Africa.

##### a. The Gourma region

System dynamics – and thus the invariants characterizing the involved active processes – usually varies with the scale of analysis (*e.g.* Pascual et al., 2001). This behaviour also applied to vegetation cycle in this analysis. The correlation dimension  $D_c$ , the horizon of predictability  $H_p$ , and the level of Gaussian noise  $\sigma$ , are plotted as a function of the aggregation index  $I_A$  in Figure 9. As explained above the preferred value for the embedding dimension is  $D_e = 4$ , but the results obtained with  $D_e = 3$  and  $D_e = 5$  are also figured for comparison. The corresponding averaged values (computed with embedding dimensions 3 to 5) are given in Table 2.

The vegetation cycle aggregated over the larger window ( $I_A = 8$ ) exhibits a simple pattern. The correlation dimension tends to low values (note that the *local* dimension of an

exactly periodic cycle is 1). As expected, the horizon of predictability is at least three times larger than for the other aggregation scales, remembering that the time series quite looks much like a regular annual cycle (see Fig. 2c). The estimated percentage of noise is quite high, as a result of mix of signals of many heterogeneous vegetation timing patterns aggregated over a north-south gradient of more than 1000 km long. At the pixel scale ( $I_A = 1$ ), the noise level is high as well; the correlation dimension is elevated and the horizon of predictability at his lowest level. These characteristics indicate a mix of stochastic and deterministic behaviors in the NDVI time series.

The intermediate scales ( $I_A$  between 3 and 5) are the most interesting: the noise is partly filtered and the high value of the correlation dimension seems to indicate that the deterministic nonlinearity of the vegetation cycle dynamics is preserved. The horizon of predictability is about 0.6 decades. It is worth making a few comments about this value. It means that at these spatial scales (approximate range of 1000 - 15000 km<sup>2</sup>) the uncertainty attached to the knowledge of the vegetation state (e.g. in terms of biomass) is doubling every 6 days (exponential growth). Another definition of the horizon of predictability is the time interval necessary for a 1% RMS error on a system state to grow to a 50% RMS error level. With this definition, the horizon is of about 34 days. Looking for the best compromise between predictability, spatial resolution and noise filtering, in predicting the vegetation cycle over the Gourma region from AVHRR NDVI would be to use an aggregation scale of about 4×4 pixels, or 1024 km<sup>2</sup> ( $I_A = 3$  ; with  $D_e = 4$  and  $\tau = 8$  decades).

### ***b. The West Africa window***

The geographical distribution of the correlation dimension  $D_c$ , horizon of predictability  $H_p$  and % of Gaussian noise  $\sigma$  over the whole West Africa are presented in Figure 10 and in the associated error maps in Figure 11. These maps are built from the non aggregated NDVI data time series ( $I_A = 1$ ). This choice of spatial resolution might not be optimal for prediction (we have seen that  $I_A = 3$  or 4 is a better choice for pixels centred on Agoufou) but it allows drawing guidelines for the prediction of the vegetation cycle.

Each variable mainly exhibits (irregular) zonal patterns that broadly follow the latitudinal bioclimatic arrangement of vegetation types given in Table 3: (i) Guinean forests from 4°N to 8°N; (ii) Guinean forest and savanna mosaic from 9°N to 10°N; (iii) Sudanian savannas between 11°N and 13°N; (iv) Soudano-Sahelian and Sahelian grasslands between 13°N and 16°N; (v) northern Sahelian grasslands between 16°N and 17°N, and (vi) Sahelo-Saharan steppes to the north of 17°N. The correlation dimension presents lower values both, in the Sahara and Sahel bands to the north, and in the Guinean forests to the south. This feature suggests that the data vectors are not densely distributed in phase-space but are rather scattered. Indeed, the noisy spatial pattern of the correlation dimension is correlated with the geographical distribution of the Gaussian noise level (Fig. 10, bottom): relatively low values of the correlation dimension correspond to high noise levels. The lower values of the horizons of predictability are found in the Sahel and Soudano-Sahelian strips between 13°N and 16°N, with quite homogeneous values in the range 0.5 - 2.0 decades. This low predictability level is interpreted as reflecting the fast dynamics of the vegetation dominated by annual grasslands. Indeed, except for shrubs and trees, the above-ground vegetation at these latitudes is somewhat reset to a standing straws and litter state every year at the end of the 2 to 4 months rainy season; and the cumulated productivity of these grasslands much depends on the precipitation of the current year. South of 13°N down to ~8°N the map of the horizons is much less coherent, with values changing over a large range (~1.0 to ~5.0 decades) between nearby pixels. These regions are covered with a mix

of savannas, cropland (mostly annual crops) and forests spatially distributed in a heterogeneous way. The results may reflect this spatial heterogeneity. The strip south of 8°N is very noisy (Fig. 10, bottom) because of the perturbations due to frequent cloud cover and poor ability of NDVI values close to saturation to monitor vegetation phenology at these latitudes. The horizon of predictability in these areas presents no spatial coherence in a range of low values (~0.0 - 2.5 decades).

The distribution of the error estimates associated with the correlation dimension, horizons and noise level show similar spatial patterns (Fig. 11). The estimates of the parameters ( $D_c$ ,  $H_p$ ,  $\sigma$ ) of the vegetation cycle are badly constrained in the northern and southern strips. They are relatively robust in the Sahel and Sudanian belts, the associated error level being relatively low. Note that most of the particular sites or structures designed by letters in Figure 1 also appear contrasting with the surroundings in the maps of the vegetation parameters (Fig. 10) and the maps of the associated errors (Fig. 11) (Senegal River (A), Mandingue plateau with the Tambaoura cliff western limit (E), lakes of Manantali (F), Kainji (O), and Volta (P), etc.).

The latitudinal patterns of the correlation dimension  $D_c$ , horizon  $H_p$  and % of noise  $\sigma$  are presented in Figure 12 for both scales  $I_A = 1$  and  $I_A = 5$ . Only the series with a noise level lower than 50% RMS are considered (the number of remaining time series is plotted in Fig. 12 - bottom). The resulting high values of  $D_c$  and low values of  $H_p$  obtained south of 8°N are interpreted as arising from a mix of noise contamination, complex vegetation dynamics and bimodal rainfall regime. Moreover, at the most southern and northern latitudes not enough time series can be exploited to conclude. In a latitudinal band extending from the guinean mosaic of forests and savannas to the Soudano-Sahelian regions (from ~8°N to 16°N) similar and coherent latitudinal patterns are obtained with the two aggregation indices  $I_A = 1$  and  $I_A = 5$ . At the larger scale ( $I_A = 5$ ) the correlation dimension is slightly smaller and the horizon of predictability is larger, by up to several decades at some latitudes (~12°N). These features indicate a simpler vegetation dynamics and a better predictive potential of the aggregated time series. As expected, the noise level is also generally lower at smaller resolution ( $I_A = 5$ ), the aggregation process partly filtering the noise. It should also be noted that the best predictive potential is found in the strip 10°N – 14°N irrespective of the aggregation scale.

The scale dependence of  $D_c$ ,  $H_p$  and  $\sigma$  is systematically considered in Figure 13. The % of Gaussian noise  $\sigma$  is gently decreasing as the aggregation scale increases (from 33% at  $I_A = 1$  to 28% at  $I_A = 6$ ). A decrease of the correlation dimension  $D_c$  (from ~1.57 to ~1.47) coupled with an increase of the horizon of predictability  $H_p$  (from ~1.5 to ~3.5 decades) clearly appear with increasing spatial scale from  $I_A = 1$  to  $I_A = 6$ . The values of the correlation dimension  $D_c$ , the horizon  $H_p$  and the noise  $\sigma$  thus follow trend behaviours across the aggregation scales at least from  $I_A = 1$  to  $I_A = 6$ . The values of  $D_c$ ,  $H_p$  and  $\sigma$  obtained at larger scales  $I_A = 7$  and  $I_A = 8$  are coherent in magnitude with the trends observed across lower aggregation scales, but no statistical significance can be attached to these values.

## 5. Discussion

### ***a. The estimates of the vegetation cycle parameters and noise***

The robustness of the results is limited by the length of the series, by the 10-day sampling rate and by the 8 bits quantization of the NDVI values. The qualitative impact of these observational factors was assessed by analysing resized, under-sampled and

quantized synthetic series derived from the Lorenz system under a chaotic regime. The performed tests show that these limitations tend to induce underestimations of the correlation dimension  $D_c$  and overestimations of the predictability horizon  $H_p$  and noise level  $\sigma$ .

The high level of noise is also affecting the results. The estimator of the correlation dimension is also sensitive to noise level. Indeed, the estimation of  $D_c$  depends on the slope of the correlation integral at small resolution radii (see the Appendix), and the determination of noise  $\sigma$  relies as well on the changes in slope of the same function. The higher the level of noise, the larger the radius at which the slope of the correlation integral is affected. Using various tests it was checked that the error associated with  $D_c$ ,  $H_p$  and  $\sigma$  are lower than 0.25, 1.5 decades and 10% RMS respectively for a noise below the 50% RMS level.

As a consequence, the large errors (up to 30%) on the parameters ( $D_c$ ,  $H_p$  and  $\sigma$ ) of the vegetation cycle reached over the Guinean forests, the North Sahel and Sahara coincide on the maps with the larger estimates of the noise level (Fig. 11). Elsewhere, the error is relatively low (in the range 2% to 10%). The slow decrease of the estimated noise level with increasing aggregation scale (Figure 13) suggests that the noise presents spatial correlations that are filtered only at the largest scale (1024x1024 km<sup>2</sup>). An indirect source of such noise in NDVI data might be the spatially coherent meteorological events at intermediate scales. It should be noted that, due to the limited length of time series, the estimator of the noise level is partly lumping the Gaussian noise and the non-Gaussian noise components together.

### ***b. The patterns of predictability***

In the Sudanian strip (11°N – 13°N), there is good spatial coherence of the correlation dimension  $D_c$  associated with large spatial heterogeneity of the horizon of predictability  $H_p$ . Low values of the correlation dimension associated to short horizon of predictability is a common characteristic to many low-dimensional chaotic systems (regime characterized by rapidly developing dynamical instabilities; see Sprott & Linz 2000). However, the contrasted spatial distribution of these two parameters is noticeable. The smooth spatial pattern of the correlation dimension  $D_c$  is attributed to the relatively high spatial homogeneity of the NDVI scenes along zonal strips. The heterogeneous spatial pattern of the horizon of predictability  $H_p$  is linked to the high sensitivity of the savannas vegetation cycles (dominantly C4 grassland) to the rainfall distribution which small fluctuations may lead to significant changes of the seasonal growth. Crop calendar and wild fire may also contribute to the contrasted spatial patterns.

The behaviour observed across the Sudano-Sahelian and Sahelian belts is explained by the lower predictability of the rainfall regime that largely controls vegetation production. Indeed, the distribution of rainfall can vary widely at short distances, and from one year to another (Ali et al., 2003; Le Barbé et al., 2002; Taupin, 2003), especially at the northern edge. The stochastic behaviour of the time-space rainfall distribution does not exclude a high contribution of deterministic processes due to the smooth response of the vegetation to the rainfall (Jarlan et al., 2008) and also to the positive feedback between land surface and rainfall (the rainfall patterns in subsequent events tend to persist and to reinforce soil moisture patterns, Taylor & Lebel, 1998). The variability in annual species composition (% of C3 and C4 plants, % of mono- and dicotyledons, Mougín et al., 2009) will also contribute to reduce system predictability. Indeed, the respective contributions of the various vegetation functional types significantly influence the averaged dynamics of the vegetation. As an example, dominant sahelian dicotyledons are characterized by a rapid germination and growth, while C4 annual grasses have high photosynthesis rate (Cissé,

1986). Locally, it is interesting to note that, much higher values of the horizon of predictability  $H_p$  are obtained in rice fields of the Office du Niger (J), in the Massina region (L) and in the Segou region (Bani valley) (K) due to the higher predictability of a cultivated area. At these sites, the uncertainty of the precipitation is partly compensated by the farming calendar organisation, the flood or field irrigation.

The main objective of this study is to find a scale that constitutes a good trade-off between high resolution and acceptable predictability of the vegetation dynamics. The statistical significance of the results is only guaranteed for the aggregation scales  $I_A = 1$  to  $I_A = 6$  (not enough time series being formed at larger aggregation scales). As a result of the filtering effect of the aggregation process the noise level is lower, although only marginally so, in the range  $I_A = [4-6]$  (see Fig. 13). The correlation dimension is following a similar pattern: it is lower in this range of scales so that corresponding time series embedded in the pseudo-phase space are arranged in a more regular, smoother way. This interpretation is in agreement with the observed increase of the horizon of predictability of the vegetation cycle (as measured by the AVHRR NDVI) in the same range of spatial scales. These intermediate spatial scales with indices  $I_A = [4-6]$  constitute thus for these data (and owing to the structure and magnitude of the noise) the sought scale of non trivial determinism that preserves the nonlinear fundamental properties of the vegetation cycle without smoothing too heavily its most significant spatial patterns.

This general conclusion should be tempered when the interest on vegetation dynamics is focused on a precise region (as seen for example with the cultivated areas). This question is not investigated in detail here but some relevant information on the appropriate scale to be used in a definite region can be found by first considering the maps of the correlation dimension, horizon of predictability and percentage of noise level drawn in Fig. 10. Although this map is produced with non aggregated data, it provides useful insights on the values of the vegetation parameters to be expected. The dependency on the latitude position of the vegetation parameters at different scales (Fig. 12) should be also considered. It is likely for example that the preferred scale of data aggregation to predict vegetation cycle will differ in Sudanian, Sudano-Sahelian and Sahelian bands among which large differences of correlation dimension  $D_c$  and horizon of predictability  $H_p$  are found.

## 6. Conclusion

A method is proposed to determine the spatial scale of data aggregation that constitutes the best trade-off between useful predictability of the nonlinear vegetation cycle and good representation of its spatial patterns. This method is based on a nonlinear characterization of the vegetation dynamics observed by remote sensing. The correlation dimension  $D_c$ , the horizon of predictability  $H_p$ , and the Gaussian noise level  $\sigma$ , are estimated from NDVI data of the NOAA-AVHRR satellite over West Africa during the period 1982 to 2004.  $D_c$  is an indicator of the complexity of the vegetation dynamics. The horizon  $H_p$  is related to the rate of instability of the vegetation dynamics and is deduced from the inverse of correlation entropy. It is statistically defined in this paper as the time interval necessary for an error on the vegetation system state to double.

The geographical distributions of the correlation dimension  $D_c$ , and of the horizon of predictability  $H_p$ , are broadly arranged in latitudinal bands following the bioclimatic zones of West Africa. The horizon  $H_p$  presents a range of values between merely one to four or five decades depending on the local stability of the vegetation dynamics. As expected, several low lands and areas under irrigated cropping show a better predictability (e.g. the Senegal delta and river, the Office du Niger, the Massina region or the Segou region in

Bani valley). The Gaussian noise level in the data time series is found to vary from about 15% to 90% RMS, with some interpretable spatial patterns.

The vegetation parameters of interest for prediction also present a smooth dependence on the spatial scale of aggregation. In most regions of West Africa the range of scale from  $64 \times 64 \text{ km}^2$  (aggregation index  $I_A = 4$ ) to  $256 \times 256 \text{ km}^2$  ( $I_A = 6$ ) is the best compromise for spatial aggregation. Indeed, the signal to noise ratio is relatively high in this range, the horizon of predictability is significantly higher (a few decades are needed to double the error) and the spatial resolution preserves the richness of nonlinear vegetation dynamics.

## Appendix

The Average Mutual Information function  $\bar{I}(\tau')$  is used to estimate the time delay for decorrelation between two time series issued from nonlinear processes. When applied to a single time series  $x = [x_1, x_2, \dots, x_{n_d}]$  its first minimum provides an estimate of the time after which the information of the system is lost. It is computed with the following relation (Fraser & Swinney, 1986):

$$\bar{I}(\tau') = \sum_{n=1}^{n_d - \tau' + 1} P[x_n, x_{n+\tau'}] \ln \left( \frac{P[x_n, x_{n+\tau'}]}{P[x_n]P[x_{n+\tau'}]} \right) \quad (\text{A1})$$

where  $P[x_n]$  is the  $x$ 's empirical probability density function (PDF) and  $P[x_n, x_{n+\tau'}]$  the joint PDF of the  $x$ 's and of the  $x$ 's with time delay  $\tau'$ .

The correlation integral  $T_m(h)$  associated to a time series with some Gaussian noise is given by (Diks 1996, 1999):

$$T_m(h) = \phi \left( \frac{h^2}{h^2 + \sigma^2} \right)^{m/2} e^{-K\tau m} \left( \frac{h^2 + \sigma^2}{m} \right)^{D_c/2} \text{ for } \sqrt{h^2 + \sigma^2} \rightarrow 0, m \rightarrow \infty \quad (\text{A2})$$

where  $h$  is the resolution radius,  $\phi$  a normalizing constant,  $\sigma$  the % of Gaussian noise,  $m$  a trial embedding dimension,  $\tau$  the time delay,  $K$  the correlation entropy and  $D_c$  the correlation dimension. The % of Gaussian noise  $\sigma$  is defined as:

$$\sigma = \sigma_n / \sigma_s = \sigma_n [\sigma_c^2 + \sigma_n^2]^{-1/2} \quad (\text{A3})$$

$\sigma_s$  being the standard deviation of the input noisy signal,  $\sigma_c$  the standard deviation of the underlying clean part of the signal and  $\sigma_n$  the Gaussian noise component. The parameters of interest for analyzing the predictability of the dynamics,  $D_c$ ,  $K$  and  $\sigma$ , are extracted through a fitting procedure applied to the empirical curves of the correlation integral  $T_m(h)$ .

## Acknowledgements

We are grateful to Prof. Cees Diks (Center for Nonlinear Dynamics in Economics and Finance, Dept. of Quantitative Economics, University of Amsterdam) for providing us with the algorithm for the Gaussian kernel estimate of the correlation integral. This work was conducted within the framework of the AMMA program. Based on a French initiative, AMMA was built by an international scientific group and is currently funded by a large number of agencies, especially from France, the UK, the USA and Africa. It has been the beneficiary of a major financial contribution from the European Community's Sixth Framework Research Programme (AMMA-EU). Detailed information on scientific coordination and funding is available on the AMMA International web site, <http://www.amma-international.org>.

## 7. References

- Abarbanel H.D.I., 1996. Analysis of Observed Chaotic Data, *Springer Verlag*, New York, 272 pp.
- Alhamad M.N., Stuth J. & Vannucci M., 2007. Biophysical modelling and NDVI time series to project near-term forage supply: spectral analysis aided by wavelet denoising and ARIMA modelling. *International Journal of Remote Sensing*, 28(11), 2513-2548, doi:10.10180/01431160600954670.
- Ali A., Lebel T. & Amani A., 2003. Invariance in the spatial structure of Sahelian rain fields at climatological scales. *Journal of Hydrometeorology*, 4(6), 996-1011.
- Aubreville A., 1949. Climats, Forêts et Désertification de l'Afrique Tropicale. *Société d'Éditions Géographiques, Maritimes et Coloniales*, Paris. 351pp.
- Cissé A.M. 1986. Dynamique de la strate herbacée des pâturages de la zone sud-sahélienne. Ph.D. dissertation. *Wageningen Agriculture University*, Wageningen. 211 pp.
- Diks C., 1996. Estimating invariants of noisy attractors, *Physical Review E*, 53(5), R4263-R4266.
- Diks C., 1999. Nonlinear time series analysis – methods and applications, *World Scientific*, Singapore, 209 pp.
- Ding M., Grebogi C., Ott E., Sauer T. & Yorke J.A., 1993. Estimating correlation dimension from a chaotic time series: when does plateau onset occur? *Physica D*, 69, 404-424.
- Fraser A.M. & Swinney H.L., 1986, Independent coordinates for strange attractors from mutual information, *Physical Review A*, 33, 1134-1140.
- Frede V. & Mazzega P., 1999a. Detectability of deterministic non-linear processes in Earth rotation time series – I. Embedding, *Geophysical Journal International*, 137, 551-564.
- Frede V. & Mazzega P., 1999b. Detectability of deterministic non-linear processes in Earth rotation time series – II. Dynamics, *Geophysical Journal International*, 137, 565-579.
- Grassberger P. & I. Procaccia, 1983. Characterization of strange attractors. *Physical Review Letters*, 50, 346-349.
- Hiernaux P., Mougin E., Diarra L., Soumaguel N., Lavenu F., Tracol Y. and Diawara M., 2009. Sahelian rangeland response to changes in rainfall over two decades in the Gourma region, Mali, *Journal of Hydrology*, 375 (1–2), 114–127.
- Holben B. N., 1986. Characteristics of maximum-value composite images for temporal AVHRR data. *International Journal of Remote Sensing*, 7, 1417-1434.
- Jakubauskas M.E., Legates D.R. & Kastens J.H., 2002. Crop identification using harmonic analysis of time-series AVHRR NDVI data. *Computers and electronics in agriculture*, 37, 127-139.
- Jarlan L., Tourre Y., Mougin E., Philippon N. & Mazzega P., 2005. Dominant patterns of AVHRR NDVI interannual variability over the Sahel and linkages to key climate signals (1982-2002). *Geophysical Research Letters*, vol. 32, L04701, doi:10.1029/2004GL021841.
- Jarlan L., Mangiarotti S., Mougin E., Mazzega P., Hiernaux P. & Le Dantec V., 2008. Assimilation of SPOT/VEGETATION NDVI data into a Sahelian vegetation growth model, *Remote Sensing of Environment*, 112, 1381-1394, doi:10.1016/j.rse.2007.02.041.
- Kennel M.B. & Abarbanel H.D.I., 2002. False neighbors and false strands: A reliable minimum embedding dimension algorithm. *Physical Review E*, 66, 1-18, doi:10.1103/PhysRevE.66.026209.
- Lasota A. & Mackey M.C., 1994. Chaos, Fractals and Noise, *Springer Verlag*, Berlin, 472pp.
- Le Barbé L., Lebel T. & Tapsoba D., 2002. Rainfall variability in West Africa during the years 1950-90. *Journal of Climate*, 15(2), 187-202.

- Martínez B. & Gilabert M.A., 2009. Vegetation dynamics from NDVI time series analysis using wavelet transform. *Remote Sensing of Environment*, 113, 1823-1842, doi:10.1016/j.rse.2009.04.016.
- Mangiarotti S., Mazzega P., Jarlan L. Mougin E., Baup F. & Demarty J., 2008. Evolutionary Bi-Objective Optimization of a semi-arid vegetation dynamics model with NDVI and  $\sigma^0$  Satellite data. *Remote Sensing of Environment*, 112, 1365-1380, doi:10.1016/j.rse.2007.03.030.
- Mougin E., Hiernaux P. and 34 co-authors, 2009; The AMMA Gourma observatory site in Mali: relating climatic variations to changes in vegetation surface, hydrology, fluxes and natural resources, *Journal of Hydrology*, 375, 14-33, doi:10.1016/j.jhydrol.2009.06.045.
- Packard N., Crutchfield J., Farmer D. & Shaw R., 1980. Geometry from a time series, *Physical Review Letters*, 45, 712-716.
- Pascual M., Mazzega P. & Levin S.A., 2001. Oscillatory dynamics and spatial scale: the role of noise and unresolved pattern. *Ecology*, 82(8), 2357-2369.
- Sprott J. C. & Linz S. J., 2000. Algebraically simple chaotic flows. *International Journal of Chaos Theory and Applications*. 5, 3-22.
- Takens F., 1981. Detecting strange attractor in turbulence, in *Lecture Notes in Mathematics*, 898, Rand D. and L.S. Young eds., *Springer Verlag*, Berlin, 366-381.
- Taupin J.-D., 2003. Précision de l'estimation des précipitations au Sahel selon la densité du réseau d'observation pluviométrique (Accuracy of the precipitation estimate in the Sahel depending on the rain-gauge network density). *Comptes Rendus Géoscience*, 335, 215-225.
- Taylor C.M. & Lebel T., 1998. Observational evidence of persistent convective-scale rainfall patterns. *Monthly Weather Review*. 126, 1597-1607.
- Thébaud B., 2002, *Foncier Pastoral et Gestion de l'Espace au Sahel*, Karthala, Paris, 318pp.
- Tucker C.J., Vanpra C.L., Sharman M.J. and Van Ittersum G., 1985. Satellite remote sensing of total herbaceous biomass production in the Senegalese Sahel: 1980-1984, *Remote Sensing of Environment*, 17, 223-249.
- Tucker C.J., Pinzon J.E., Brown M.E., Slayback D.A., Pak E.W., Mahoney R., Vermote E.F. & Saleous N.E., 2005. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26:20, 4485-4498, doi:10.1080/01431160500168686.
- Wiggins S., 1990. Introduction to Applied Nonlinear Dynamical Systems and Chaos, *Springer Verlag*, Berlin, 672 pp.
- Yu D., Small M., Harrison R.G. & Diks C., 2000. An efficient implementation of the Gaussian kernel algorithm in estimating invariants and noise level from noisy time series data, *Physical Review E*, 61(4), 3750-3756.



## 8. Tables

TABLE 1: Correspondence between the aggregation index  $I_A$ , the number of averaged pixels and the surfaces of the windows. The last column indicates the number of non overlapping windows over West Africa.

Aggregation Index $I_A$	# Averaged pixels	Surface of the window ( $\text{km}^2$ )	# Windows (non-overlapping)
1	$1 \times 1$	64 ( $=8^2$ )	64119
2	$2 \times 2$	256 ( $=16^2$ )	12290
3	$4 \times 4$	1 024 ( $=32^2$ )	3159
4	$8 \times 8$	4 096 ( $=64^2$ )	782
5	$16 \times 16$	16 384 ( $=128^2$ )	194
6	$32 \times 32$	65 536 ( $=256^2$ )	48
7	$64 \times 64$	262 144 ( $=512^2$ )	11
8	$128 \times 128$	1 048 576 ( $=1024^2$ )	2

TABLE 2: Estimated time delay  $\tau$ , embedding dimension  $D_e$ , correlation dimension  $D_c$ , horizon of predictability  $H_p$  and % RMS of Gaussian noise  $\sigma$  computed from the NDVI time series centred on the Gourma region at different spatial scales (aggregation index  $I_A$  varying from 1 to 8). The values given for the  $D_c$ ,  $H_p$  and  $\sigma$  are averages of the values obtained with the trial embedding dimension  $D_e = 3, 4$  and  $5$ .

$I_A$	$\tau$ (# decades)	$D_e$	$D_c$	$H_p$ (# decades)	$\sigma$ (% RMS)
1	8	4	1.59	0.5	26.0
2	8	4	1.55	0.8	23.7
3	8	4	1.67	0.7	21.7
4	8	4	1.68	0.6	22.5
5	8	4	1.73	0.6	22.6
6	8	4	1.68	0.7	23.4
7	9	4	1.59	1.0	23.0
8	9	3	1.33	2.4	24.5

TABLE 3: Large scale zonal regions and associated vegetation types over West Africa (region number and name, broad latitudinal range, dominant climate modes and vegetation types). For each region are also given the mean, range and index of qualitative heterogeneity ( $QH$ ) of the correlation dimension  $D_c$ , of the horizon of predictability  $H_p$  and of the estimated % RMS of Gaussian noise  $\sigma$  in the NDVI data time series. The stars associated with the heterogeneity index  $QH$  indicates a high (\*) or very high (\*\*) number of nonlinear parameters that could not be estimated because of the too low signal to noise ratio in the data time series.

Band #	Regions	Latitudes	Qualitative climate mode	Vegetation	$D_c$			$H_p$ (decades)			$\sigma$ (% RM)		
					Mean	Range	$QH$	Mean	Range	$QH$	Mean	Range	$QH$
(vi)	Sahel-Sahara	[17°N-19°N]	Monomodal	Almost no vegetation	1.4	[0.9-1.6]	VH**	0.8	[0-1]	L**	90	[60-90]	L*
(v)	North Sahel	[16°N-17°N]	Monomodal	Annual grassland (C3 & C4); few shrubs and trees	1.3	[0.9-1.6]	VH*	1.1	[0-6]	H	60	[40-70]	H
(iv)	Sahel-Soudano Sahelian	[13°N-16°N]	Monomodal	Grassland, shrubs, few trees	1.5	[1.3-1.6]	VL	0.9	[0-2]	VL	30	[15-35]	L
(iii)	Soudanian	[11°N-13°N]	Monomodal	Savannas (C4 grassland with widely spread small trees)	1.4	[1.1-1.5]	L	3.8	[1-8]	VH	35	[20-40]	VL
(ii)	Guinean Forest-Savannas	[8°N-11°N]	Monomodal	Alternatively Forest/Savannas/grassland	1.5	[1.3-1.6]	L	3.2	[0-8]	H	45	[30-50]	L
(i)	Guinean Forest	[4°N-8°N]	Bimodal	Tropical forest	1.3	[0.9-1.6]	VH**	1.6	[0-3]	VH*	60	[50-90]	L*

## 9. Figures

FIGURE 1: Map of the Standard Deviation  $\mu_2$  of the normalized AVHRR NDVI time series at  $8 \times 8 \text{ km}^2$  resolution computed over 1982-2004 (West Africa). The windows with aggregation indexes  $I_A = 4$  to  $I_A = 8$  centred on the Agoufou site (N) in the Sahelian Gourma (Mali) are represented. The salient feature of this map is the broadly north-south gradient of the NDVI signal that corresponds to the zonal distribution of the main vegetation types over West Africa (see Table 3 for details). Some interesting zones are designated by letters: the Senegal river and its delta (A), the limit of active vegetation (B), the fossil valley of Ferlo (C), the Gambia river (D), the Mandingue Plateau (E), the Manantali (F) and Selingué lakes (G), Bamako city (H), the Niger river (I), rice fields of the Office du Niger (J), the Bani valley (K), the Massina flood plain (or inner delta of the Niger River) (L), the North Seno region (no agriculture) (M), the Agoufou site (N), the Kainji lake (O), the Volta lake (Akosombo dam) (P), the Baoulé region (mix of savannah and forest) (Q), tropical forest in the South of Liberia and Côte d'Ivoire (R) or Sierra Leone (S).

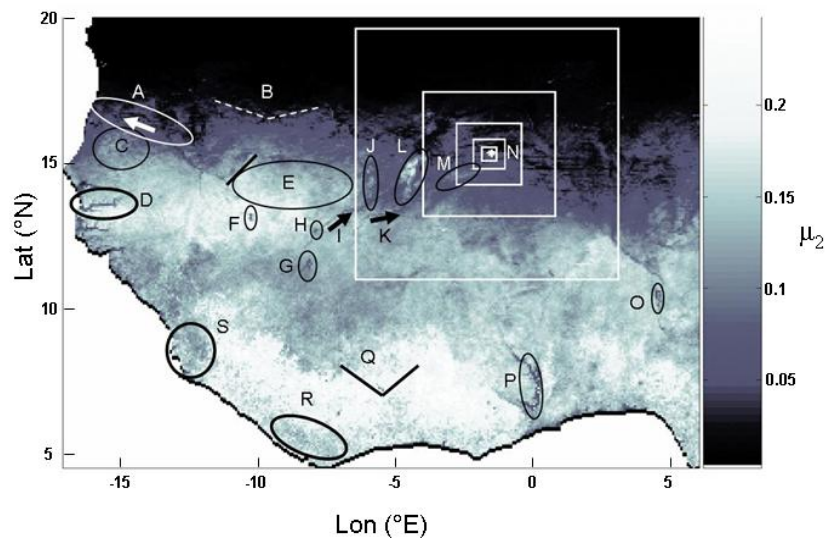


FIGURE 2: Aggregated NDVI time series in the windows centred on the Gourma region (Mali; see symbol (N) in Fig. 1). The size of the aggregation windows corresponds to (a)  $8 \times 8 \text{ km}^2$  (aggregation index  $I_A = 1$ ; see Table 1), (b)  $128 \times 128 \text{ km}^2$  ( $I_A = 5$ ), (c)  $1024 \times 1024 \text{ km}^2$  ( $I_A = 8$ ). The range of variation of the non aggregated NDVI time series (a) is reported has horizontal lines in the middle and bottom panels for comparison.

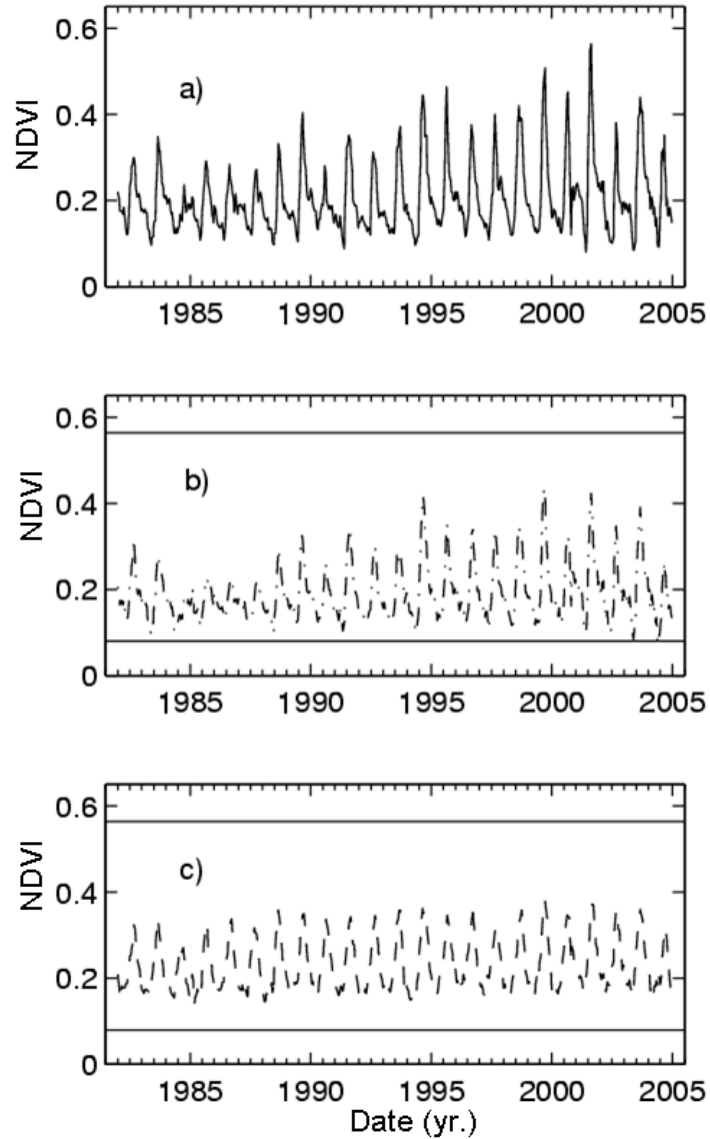


FIGURE 3: The NDVI data time series of the Agoufou site (aggregation index  $I_A = 1$ ; see Fig. 2a) embedded in a three dimensional space, using a coordinate system of delayed variables with a delay of eight decades (see Table 2). The vegetation dynamics being dominated by an annual cycle and its interannual variability, the trajectory is more or less established on a (deformed) torus. The dashed arrows show the sense of rotation of the system state on the attractor. The dry 1984 and rainy 2001 years can be easily identified: we here indicate the months of August that usually correspond to the maximum of the biomass production in that site. The non-smooth aspect of the trajectory mainly results from noise perturbations.

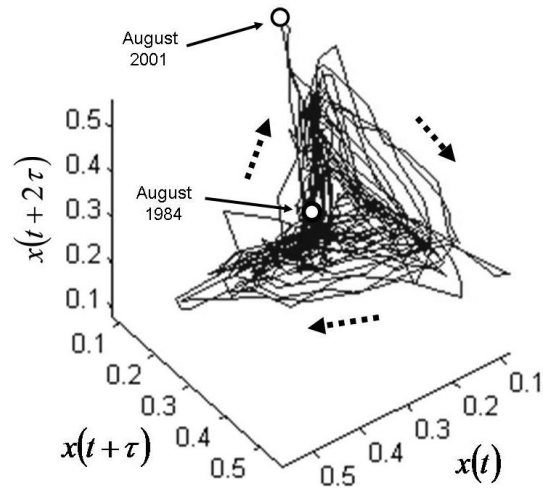


FIGURE 4: Average Mutual Information  $AMI$ , power spectra and % of global false neighbors as functions of the time delay  $\tau'$ , period  $T$  and trial embedding dimension  $D_e$  respectively. These results correspond to the analysis of NDVI time series centered on the Agoufou site, with respective aggregation index  $I_A = 1$  (top panels),  $I_A = 5$  (middle panels) and  $I_A = 8$  (bottom panels).

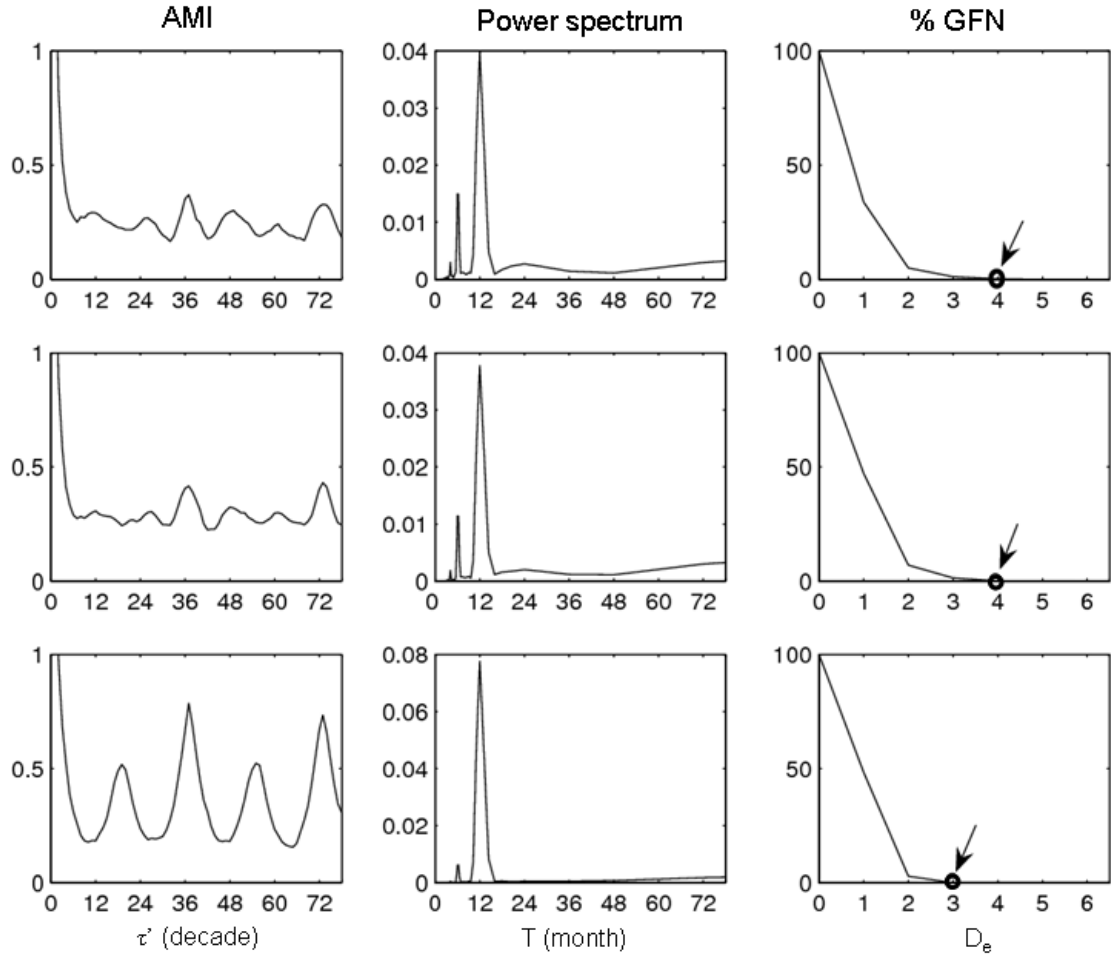




FIGURE 5: Median (over the longitude) of the distributions of the time delays  $\tau$  (in decades) as determined from an analysis of the average mutual information functions, plotted versus the latitude (in  $^{\circ}\text{N}$ ) for all the spatial windows with aggregation indexes  $I_A = 1$  (plain line) and  $I_A = 5$  (dashed-dotted line) respectively. Note that  $\tau = 9$  is obtained for  $I_A = 8$ .

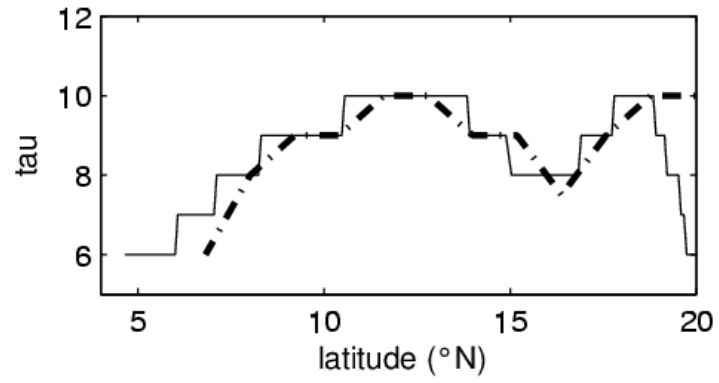


FIGURE 6: Percentage of estimated embedding dimension  $D_e$  as determined from the Global False Neighbors analysis, plotted versus the latitude (in  $^{\circ}\text{N}$ ) for all the spatial windows with aggregation indexes  $I_A = 1$  (a) and  $I_A = 5$  (b) respectively. Note that  $D_e = 3$  is obtained for  $I_A = 8$ .  $D_e = 2$  (plain line),  $D_e = 3$  (dotted line) and  $D_e = 4$  (dashed line) respectively.

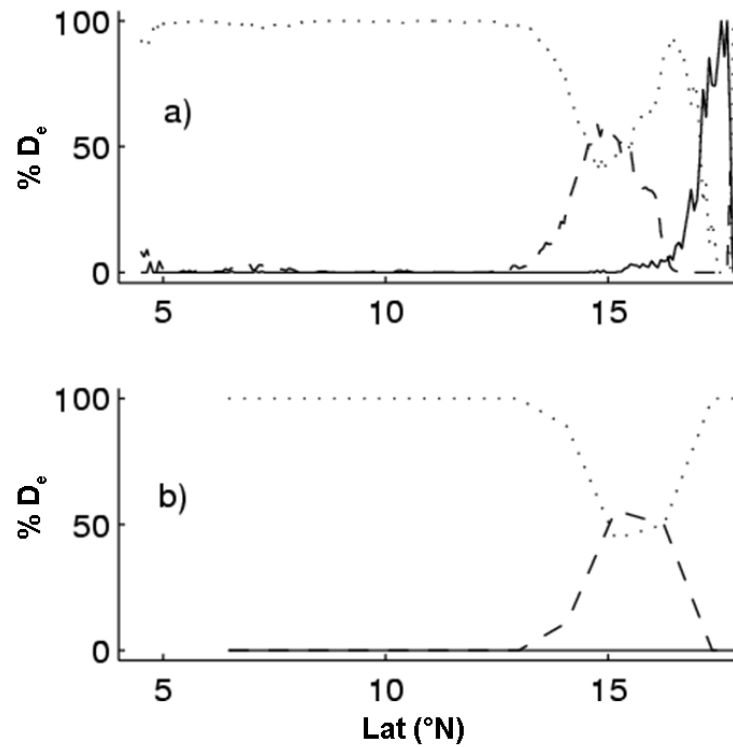


FIGURE 7: log-log representation of the correlation integral  $T_m$  as a function of the radius  $h$  for embedding dimension  $D_e = 2$  to  $D_e = 5$  (alternatively grey, black, grey and black lines). Three NDVI data time series centred on the Agoufou site are used here, with different aggregation indexes (see Table 1):  $I_A = 1$  (plain lines),  $I_A = 5$  (dashed dotted lines) and  $I_A = 8$  (dashed lines). Note that the quantization of the original NDVI time series (coded on 8 bits) prevents us from estimating the correlation integral  $T_m$  over a range of smaller radii  $h$ .

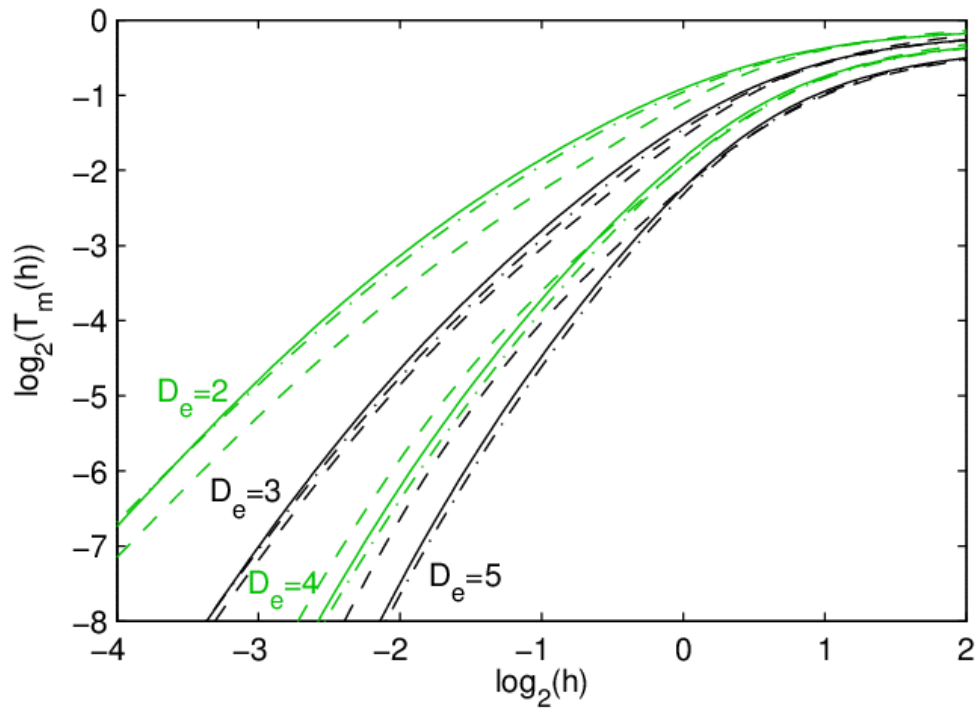


FIGURE 8: Correlation dimension  $D_c$  (top panel), horizon of predictability  $H_p$  (middle panel) and % RMS of Gaussian noise  $\sigma$  (bottom panel) as a function of the trial embedding dimension  $D_e$  (taken in the range [2-5]). The three NDVI data time series centred on the Agoufou site with different aggregation indexes (see Table 1):  $I_A = 1$  (plain lines),  $I_A = 5$  (dashed dotted lines) and  $I_A = 8$  (dashed lines) are used here.

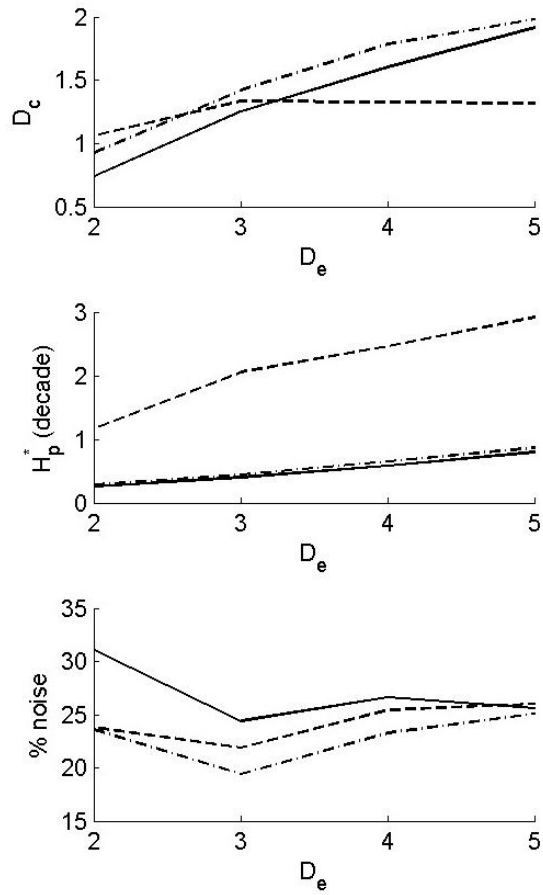


FIGURE 9: Correlation dimension  $D_c$  (top panel), horizon of predictability  $H_p$  (middle panel) and % of Gaussian noise  $\sigma$  (bottom panel) versus the aggregation index  $I_A$  of the data time series centred on the Agoufou site. For each panel:  $D_e = 3$  (plain line),  $D_e = 4$  (dashed dotted line) and  $D_e = 5$  (dashed line) are plotted.

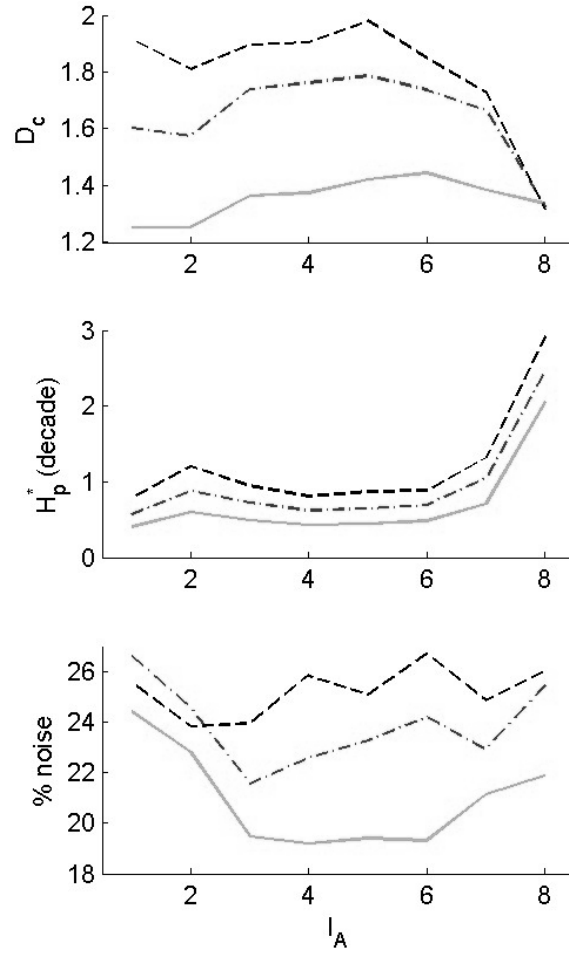


FIGURE 10: Geographical maps of the correlation dimension  $D_c$  (top panel), the horizon of predictability  $H_p$  (in decade) (middle panel) and the % of Gaussian noise  $\sigma$  (bottom panel) computed at the AVHRR pixel scale ( $I_A = 1$ ; no aggregation) over the West Africa window (values averaging estimates obtained with embedding dimensions  $D_e = 3$  to 5).

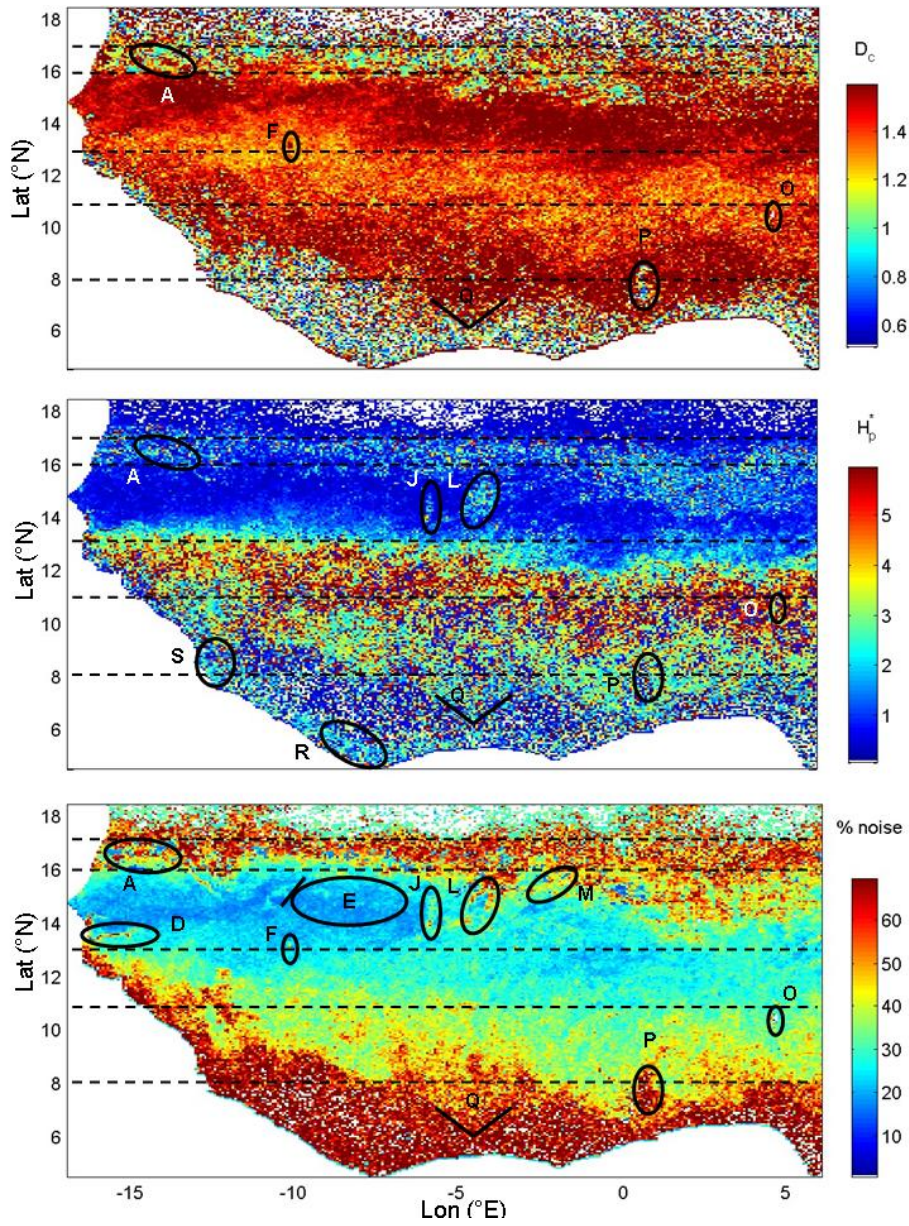


FIGURE 11: Error maps associated with the estimates of  $D_c$  (top panel),  $H_p$  (in decade) (middle panel) and % of Gaussian noise  $\sigma$  (bottom panel) (see text).

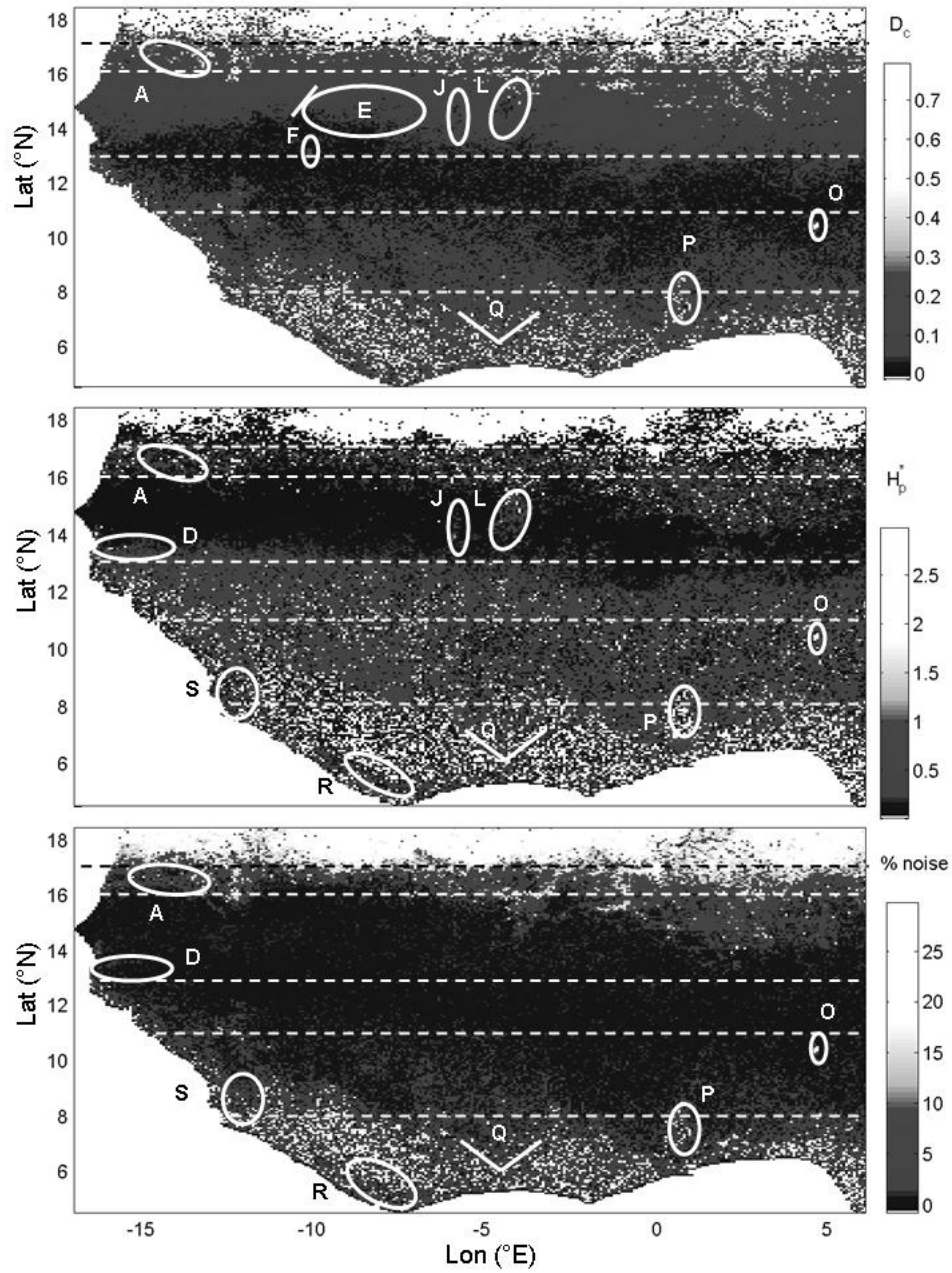


FIGURE 12: Spatial averaged values of (top panel) the correlation dimension  $D_c$ , (second panel) the horizon of predictability  $H_p$ , (third panel) the % of Gaussian noise  $\sigma$  and (bottom panel) the number of available time series plotted versus the latitude ( $^{\circ}$ N), for spatial aggregation indexes  $I_A = 1$  (plain and dashed lines) and  $I_A = 5$  (squares). Note that for  $I_A = 8$ , the following values  $D_c = 1.35$ ,  $H_p = 4.1$  decades and % of Gaussian noise  $\sigma = 24.0\%$  are obtained.

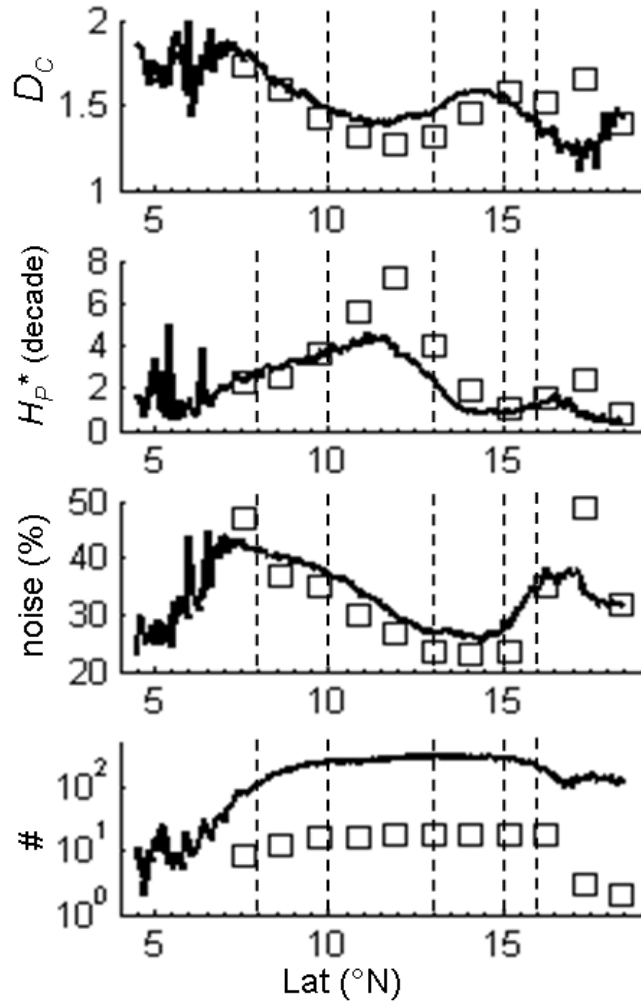




FIGURE 13: Spatially averaged values of (top panel) the correlation dimension  $D_c$ , (second panel) horizon of predictability  $H_p$ , (third panel) % of Gaussian noise  $\sigma$ , and (bottom panel) number of time series used for in the estimation process, versus the spatial aggregation index  $I_A$ .

