Missing Observations and Evolutionary Spectrum for Random Fields

Rachid Sabre
AgroSup/Laboratoire Le2i, UMR CNRS 6306, Université de Bourgogne, 26, bd Docteur Petitjean 21000 Dijon
r.sabre@agrosupdijon.fr

Abstract

There are innumerable situations where the data observed from a non-stationary random field are collected with missing values. In this work a consistent estimate of the evolutionary spectral density is given where some observations are randomly missing

Keywords: Spectral density, non-stationary processes, periodogram, smoothing estimate, oscillatory process

1. Introduction

Spectral analysis for stationary processes has been extensively studied in recent years. However, in many applications the signals must be modeled as non-stationary processes. This has motivated several authors to study non-stationary processes assuming that they are locally stationary. Priestley [11, 19] established the theory of the evolutionary spectrum generalizing spectral analysis developed for stationary processes. The evolutionary spectrum is time-dependent and describes the local power-frequency distribution at each time-instant. Other studies based on the Wold-Cramér decomposition have contributed to the development of the evolutionary spectrum [8, 13, 12, 14]. The applications of the evolutionary spectrum cover various scientific fields: signal and image processing [3, 1], seismic [16], oceanography, music [21]. The estimation of the evolutionary spectral density is studied in [19, 8, 6, 15, 7].

Moreover, Jones [4] was the first to consider the problems of missing data problems in spectral analysis. More precisely he studied the case where a block of observations was periodically unobtainable. In parallel, the theory of amplitude-modulated stationary processes was developed by Parzen [9], he applied this theory to solve the problem of periodic missing data problems. Bloomfield [2] has considered stationary processes with randomly missing data. He has given an asymptotically unbiased estimator of the spectral density and shown under suitable conditions that its variance converges to zero. We cite in this paper a few works that have contributed to finding solutions to the problems of missing observations: [17, 10, 5].

The aim of the present paper is to consider the problem of the randomly missing data for the class of non-stationary oscillatory random fields. Using the same techniques introduced by Bloomfield [2] for stationary processes, we give a consistent estimate of the evolutionary spectral density. The paper is organized as follows. In Section 2, we give some notations, assumptions and the amplitude modulating function \( Y_{\omega_1, \omega_2} \). In Section 3, we construct a periodogram and we show that it is an asymptotically unbiased estimator. Since, we smooth the periodogram in the neighborhood of the time-instant \( t \) via a weight function and we show that it is a consistent estimate of the (weighted) average value of \( h_{\omega_1, \omega_2} (\omega_{01}, \omega_{02}) \) in the
neighborhood of the time-instant \((t_1, t_2)\). Section 4 is devoted to proving theorems. In Section 5, we study numerical results and simulation. The concluding comments are given in Section 6.

2. The Amplitude Modulating Function, \(Y_{t_1,t_2}\)

As in Priestley [11, 19], we consider a non-stationary centred oscillatory random field \(\{X_{t_1,t_2}, t_1, t_2 \in Z\}\) i.e.

\[
X_{t_1,t_2} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i(t_1\omega_1 + t_2\omega_2)} A_{t_1,t_2}(\omega_1, \omega_2) dZ_1(\omega_1, \omega_2); \quad t_1, t_2 \in Z,
\]

where the function \(A_{t_1,t_2}(\omega_1, \omega_2)\) is given by

\[
A_{t_1,t_2}(\omega_1, \omega_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i(\theta_1 \omega_1 + \theta_2 \omega_2)} dF_{\omega_1,\omega_2}(\theta_1, \theta_2),
\]

where \(F_{\omega_1,\omega_2}\) is a measure satisfying: \(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dF_{\omega_1,\omega_2}(\theta_1, \theta_2) = 1\) and \(Z_1\) is a process with orthogonal increments defined on the interval \([-\pi, +\pi]^2\) and \(E|dZ_1(\omega_1, \omega_2)|^2 = d\mu_1(\omega_1, \omega_2)\) where \(\mu_1\) is a positive measure. The evolutionary spectral measure is defined by Priestley [11, 19] at each \((t_1, t_2)\) by

\[
dH_{t_1,t_2}(\omega_1, \omega_2) = \left| A_{t_1,t_2}(\omega_1, \omega_2) \right|^2 d\mu_1(\omega_1, \omega_2).
\]

Our choice of oscillatory random field is motivated by the fact that it has a physical interpretation and the variance of the process is interpreted as a measure of the total power of the process at time \(t\), because \(Var(X(t_1, t_2)) = \int_{-\infty}^{+\infty} dH_{t_1,t_2}(\omega_1, \omega_2)\). The evolutionary spectral density of the process \(\{X(t_1, t_2)\}\) is given by \(h_{t_1,t_2}(\omega_1, \omega_2)\) and defined as follows:

\[
h_{t_1,t_2}(\omega_1, \omega_2) = \frac{dH_{t_1,t_2}(\omega_1, \omega_2)}{d\omega_1 d\omega_2}, \quad \omega_1, \omega_2 \in R.
\]

Assume that the process \(\{X_{t_1,t_2}\}\) is observed with randomly missing observations. As Bloomfield [2], we consider the process \(L_{t_1,t_2}\) defined as the product of the process \(\{X_{t_1,t_2}\}\) and another process \(\{Y_{t_1,t_2}\}\) defined as follows:

\[
L_{t_1,t_2} = X_{t_1,t_2} Y_{t_1,t_2} \quad \text{where} \quad Y_{t_1,t_2} = \begin{cases} 1 & \text{if } X_{t_1,t_2} \text{ is observed} \\ 0 & \text{if not} \end{cases}
\]

56
The process $L_{t_1,t_2}$ is equal to a modified version of the original process \{X_{t_1,t_2}\} by replacing the missing observations by $E(X_{t_1,t_2})$ with zero as mean value, since \{X_{t_1,t_2}\} is centred.

To simplify matters, we suppose, as Bloomfield [2], that \{Y_{t_1,t_2}\} is stationary, independent of $X_{t_1,t_2}$ and satisfying:

\[
P\left\{ Y_{t_1,t_2} = 1 \right\} = p > \frac{1}{2},
\]
\[
P\left\{ Y_{t_1,t_2} = 0 \right\} = 1 - p,
\]

The assumption of stationarity means that the statistical properties of the process $Y$ are not time-depend. This case is often encountered in practice especially when collecting data provided by devices that are partially defective. Set $\bar{Y} = \frac{1}{p} E(Y_{t_1,t_2} Y_{t_1+\eta_1,t_2+\eta_2})$ \( \bar{Y} \in s\). Following hypotheses:

\[
P\left\{ Y_{t_1,t_2} = 1 \right\} = p > \frac{1}{2},
\]
\[
P\left\{ Y_{t_1,t_2} = 0 \right\} = 1 - p,
\]

This implies that $\bar{Y} = \frac{1}{p}$ is symmetrical in $(r_1, r_2)$. In the remainder of this paper, we assume the following hypotheses:

$H_1$)

There exists a real number $\rho > 0$ such that

\[
\sum_{q=\infty}^\infty \left| \sum_{i=\infty}^\infty E(Y_{t_1,t_2} Y_{t_1+\eta_1,t_2+\eta_2}) \right| \leq \rho \sum_{i=\infty}^\infty E(Y_{t_1,t_2}) \leq \rho (s_1, s_2) + 1 < \infty,
\]

$H_2$)

\[
\bar{Y} > 0 \text{ and } p\bar{Y} > 2p - 1 > 0 \quad r_1, r_2 \in Z
\]

Remark 2.1

- The first hypothesis $H_1$) means that the sum,

\[
\sum_{q=\infty}^\infty Cov(Y_{t_1,t_2} Y_{t_1+\eta_1,t_2+\eta_2}) \leq \rho (s_1, s_2) + 1
\]

is bounded by a function proportional to $p^2 (s_1, s_2) + 1$.

- The second hypothesis $H_2$) implies for each $(t_1, t_2)$, the probability that $X_{t_1,t_2}$ is observed (i.e. not missing) is greater than $\frac{1}{2}$. 

57
3. Estimation of the Evolutionary Spectral Density

We begin by giving some definitions introduced by Priestley [11, 19]. Let $F$ the family of oscillatory functions $\left\{ A_{\tau_1,\tau_2}(\omega_1,\omega_2)e^{i(t_1\omega_1+t_2\omega_2)} \right\}$. For each family $F$, we define the function $B_F(\omega_1,\omega_2) = \sup_{(\theta_1,\theta_2)} \left| dF_{\omega_1,\omega_2}(\theta_1,\theta_2) \right|$. Let $C$ be in the class of families $F$ such that $B_F(\omega_1,\omega_2)$ is bounded for all $(\omega_1,\omega_2)$. For each family $F$ we define the following constant $B_F$ termed the characteristic width of $F$:

$$B_F = \left[ \sup_{(\omega_1,\omega_2)} B_F(\omega_1,\omega_2) \right]^{-1}$$

The characteristic width of the process $X_{t_1,t_2}$ is defined by $B_X = \sup_{F\in C} B_F$. For more details about the definitions see Priestley [11, 19].

In this section, we propose a periodogram constructed as follows:

$$I_{t,F}(\omega_{10},\omega_{20}) = \left[ \sum_{t_1+t_2} \sum_{t_1+t_2} g_{u_1,u_2} \frac{L_{1-t_1,2-t_2}^1}{S} e^{-i(a_1-a_2)(t_1-u_1)+i(t_2-u_2)h_2}} \right]^2,$$

where $S = \left( 2\pi \sum_{u_1,u_2} p_{g_{u_1,u_2}} \right)^{\frac{1}{2}}$, and $\left\{ g_{u_1,u_2} \right\}$, is a filter satisfying the following conditions:

$C_1$: $g_{u_1,u_2} \geq 0$, $g_{u_1,u_2} = g_{-u_1,-u_2}$,

$C_2$: $\sum_{u_1,u_2} p_{g_{u_1,u_2}} g_{u_1,u_2} g_{-u_1,-u_2} < \infty$, where $\xi$ is defined in (4)

$C_3$: $g_{u_1,u_2}$ has a finite "width" defined by:

$$B_g = \sum_{u_1,u_2} p_{g_{u_1,u_2}} \left\| (u_1,u_2) \right\| g_{u_1,u_2} g^{*}_{v_1,v_2} < \infty,$$

$C_4$: $B_g << B_F$,

$C_5$: For any real numbers $k_1, k_2$, we have

$$\left\| \Gamma(s_1,s_2)h_{t_1,t_2}(s_1+k_1,s_2+k_2)ds_1ds_2 - h_{t_1,t_2}((k_1,k_2)) \right\| \Gamma(s_1,s_2)ds_1ds_2 < \frac{B_{g}}{B_F},$$

where the function $\Gamma$ is defined by:

$$\Gamma(s,s') = \sum_{u_1,u_2,v_1,v_2} p_{g_{u_1,u_2}} g_{u_1,u_2} g^{*}_{v_1,v_2} e^{-i(a_1-a_2)(t_1-u_1)+i(t_2-u_2)h_2}}.$$ 

The function $\Gamma$ is highly concentrated with relation to the function $h_{t_1,t_2}$. 

58
When this condition is satisfied, we say as in Priestley ([19] page 829) that the function \( \Gamma \) is \( \delta \) -function with respect to \( h_{t_1,t_2} \) in order \( \left( \frac{B_x}{B_F} \right) \).

\[
C_0 : g_{u_1,u_2} = O\left(e^{-\frac{1}{2}u_1 u_2} \right)
\]

The following theorem shows that the periodogram \( I_{t,T}(\omega_{01}, \omega_{02}) \) is an asymptotically unbiased estimator of the evolutionary spectral density \( h_{t_1,t_2}(\omega_{01}, \omega_{02}) \).

**Theorem 3.1** Let \( t_1, t_2 \) be an integer numbers; \( \omega_{01}, \omega_{02} \) are real numbers and suppose that \( \frac{B_x}{B_F} < \varepsilon \), then

\[
E\left[I_{t,T}(\omega_{01}, \omega_{02})\right] = h_{t_1,t_2}(\omega_{01}, \omega_{02}) + O(\varepsilon).
\]

To prove the theorem 1, we have need the two following lemmas

**Lemma 3.1** For any \( t_1, t_2, t_1', t_2', \lambda_1, \lambda_2 \) real numbers, we have

\[
\left| e^{it_1 t_2 + t_1' t_2'} \right| - e^{-it_1 t_2 + t_1' t_2'} \Gamma(s + k, s' + k') dF_{\lambda_1, \lambda_2}(s_1, s_2) dF_{\lambda_1', \lambda_2'}(s_1', s_2') -
\Gamma(k, k') \left| e^{it_1 t_2 + t_1' t_2'} \right| - e^{-it_1 t_2 + t_1' t_2'} \Gamma(s + k, s' + k') dF_{\lambda_1, \lambda_2}(s_1, s_2) dF_{\lambda_1', \lambda_2'}(s_1', s_2') \right| < \frac{2B_x}{B_F}
\]

**Lemma 3.2** Let \( \theta_1, \theta_2, \lambda_1, \lambda_2, t_1, t_2 \) and \( t_1', t_2' \) be real numbers, we have

\[
A_{t_1,t_2}(\lambda_1, \lambda_2)A^+_{t_1',t_2'}(\lambda_1, \lambda_2) \left( \Gamma(\theta_1, \theta_2) - \Gamma(\theta, \theta) \right) \leq \frac{2B_x}{B_F}, \quad \text{where}
\]

\[
\Gamma(\theta_1, \theta_2) = \sum_{u_1, u_2, v_1, v_2} p_{u_1, u_2, v_1, v_2} g_{u_1, u_2} g^*_{v_1, v_2} \beta(u, v, \theta)
\]

where

\[
\beta(u, v, \theta) = A_{t_1 - u_1, t_2 - u_2}(\lambda_1, \lambda_2)A^+_{t_1' - v_1, t_2' - v_2}(\lambda_1, \lambda_2) e^{-i[(u_1 - v_1)\theta_1 + (u_2 - v_2)\theta_2]}
\]

In order to obtain a consistent estimate of \( \{ h_{t_1,t_2}(\omega_{01}, \omega_{02}) \} \), we smooth the periodogram in the neighborhood of the time-instant \( (t_1, t_2) \) via a weight function:

\[
\hat{h}_{t_1,t_2}(\omega_{01}, \omega_{02}) = \sum_{v_1, v_2 \in M} w_{t_1,t_2,v_1,v_2} I_{t_1 - v_1,t_2 - v_2}(\omega_{01}, \omega_{02})
\]

where \( w_{t_1,t_2,v_1,v_2} \) is a weight-function depending on the parameters \( T_1, T_2 \) and satisfying

a) \( w_{t_1,t_2,v_1,v_2} \geq 0 \), for all \( v_1, v_2, T_1, T_2 \)
b) \( w_{T_1, T_2\cdot v_1, v_2} = 0, \; v_1, v_2 \notin M \), where \( M \) is a set of integers surrounding zero.

c) \( w_{T_1, T_2\cdot v_1, v_2} = w_{T_1, T_2\cdot v_1, v_2} \cdot \frac{1}{2} \),

d) \( \sum_{v_1, v_2 \in M} w_{T_1, T_2\cdot v_1, v_2} = 1 \),

e) \( \sum_{v_1, v_2 \in M} w^2_{T_1, T_2\cdot v_1, v_2} < \infty \).

f) We assume that there exists a constant \( C \) such that

\[
\lim_{T_1, T_2 \to \infty} \frac{T_1, T_2}{\sum_{v_1, v_2 \in M} w_{T_1, T_2\cdot v_1, v_2}} = C, \text{ where }
\]

\[
w_{T_1, T_2\cdot v_1, v_2} = \sum_{v_1, v_2 \in M} e^{-i(v_1 + v_2)} w_{T_1, T_2\cdot v_1, v_2}.
\]

The following theorem shows that the estimator \( \hat{h}_{t_1, t_2} (\omega_1, \omega_2) \) is an asymptotically unbiased of the (weighted) average value of \( h_{t_1, t_2} (\omega_1, \omega_2) \) in the neighbourhood of \( (t_1, t_2) \).

**Theorem 3.2** Let \( (\omega_0, 0) \) be an element of \( (\pi, \pi)^2 \) and suppose that \( \frac{B^B}{B_X} < \epsilon \), then

\[
E \left[ \hat{h}_{t_1, t_2} (\omega_1, \omega_2) \right] = \bar{h}_{t_1, t_2} (\omega_1, \omega_2) + O(\epsilon)
\]

where \( \bar{h}_{t_1, t_2} (\omega_1, \omega_2) = \sum_{v_1, v_2 \in M} w_{T_1, T_2\cdot v_1, v_2} h_{t_1, t_2\cdot v_1, v_2} (\omega_1, \omega_2) \).

To show that the variance converges to zero, as in Priestley ([11]) and Mélard [8], we assume that the process \( L_{t_1, t_2} \) is Gaussian.

**Theorem 3.3** Let \( (\omega_0, 0) \) be an element of \( (\pi, \pi)^2 \) and suppose that the process \( L_{t_1, t_2} \) is Gaussian, then we have

\[
\text{Var} \left[ \hat{h}_{t_1, t_2} (\omega_1, \omega_2) \right] = O \left( \frac{1}{T_1, T_2} \right).
\]

4. Numerical Studies

As in Bloomfield [2], we suppose that our process \( \{X_{t, x}\}_{x \in Z} \) is observed at the successive instants \( (t_1, s_1), (t_2, s_2), \ldots, (t_n, s_n) \) where \( t_i = t_{i+1} - t_i \) and \( s_i = s_{i+1} - s_i \) are independent random variables, each with the probability distribution \( f_{t_i, s_i} = P(\tau = (t, s) = (r, r)) \), and the finite mean \( p^{-1} \). As in Feller ([18], pp. 282-283), we define a process \( \{Y_{t, x}\} \) which coincides with \( \{Y_{t, x}\} \) except at origin \( Y'_{0,0} = 1 \). The event "\( Y' = 1 \)" is termed persistent and recurrent event. Using (6) we obtain
\[ \xi_{t_1 \ldots t_2} = p^{-1}E\left\{ Y_{t_1 \ldots t_2}^\gamma Y_{t_1 \ldots t_2}^\gamma + 2\right\} = P\left\{ Y_{t_1 \ldots t_2}^\gamma = 1\right\} \]

Feller ([18], pp. 282-283) has shown that

\[ \xi_{t_1 \ldots t_2} = \sum_{s_1 \ldots s_2=1} f_{s_1 \ldots s_2} \xi_{t_1 \ldots t_2 - s_1 \ldots s_2}, t_1, s_1, 1, 2 \ldots \]

The process \( L_{t_1 \ldots t_2} \) was obtained from the process \( X \) by omitting certain observations with the renewal-type mechanism defined above with \( f_{1,1} = \frac{8}{9}, f_{2,2} = \frac{1}{9}, f_{t_1 \ldots t_2} = 0 \) otherwise.

**The simulation of the process \( X \):**

Using the same method as in [20] for the simulation of Markov Gauss random field, we simulate the Gaussian random field \( Y = \left\{ Y_{n_1 n_2} \right\}_{n_1, n_2 \in \mathbb{Z}} \) such that the covariance function is given by \( C_y(n_1, n_2) = e^{-\sqrt{(n_1 + n_2)}} \), and its spectral density is \( f_y(\lambda_1, \lambda_2) = \frac{1}{\pi(1 + \lambda_1^2 + \lambda_2^2)} \).

The random field \( X_{t,s}, t, s \in \mathbb{Z} \) is given by the following model

\[ X_{t,s} = c_{t,s} Y_{t,s}, \quad t, s \in \mathbb{Z} \]

where \( c_{t,s} = \exp\left\{ -\frac{(t+s-500)^2}{2 \times 200^2} \right\} \) and \( A_{t,s}(\omega_1, \omega_2) = c_{t,s} \) is independent of \( (\omega_1, \omega_2) \). With respect to the family \( F = \left\{ c_{t,s} e^{i(\omega_1 t + \omega_2 s)} \right\} \), \( X_{t,s} \) has an evolutionary spectral density function \( h_{t_1 \ldots t_2}(\omega_1, \omega_2) = c_{t_1 \ldots t_2} f_y(\omega_1, \omega_2) \). The curve of the estimator with 5000 observations (Figure 2) and that of the spectral density (Figure 1) are very similar. So the estimator is quite satisfactory. If we take more observations (around 10000), the estimator becomes much smoother and the curve much approaches the density.

![Figure 1. Density](image1.png)

![Figure 2. Estimator](image2.png)
5. Conclusion

We have proposed in this paper some results about the estimation of the evolutionary spectral density for non-stationary random fields where the data observed are collected with missing values. The approach is based on the technique used by Bloomfield [2] for stationary processes combining the estimates of evolutionary spectrum introduced by Priestley [11]. This work could be applied to several cases when the process is non-stationary as for example in:

- the segmentation of a sequence of images of a dynamic scene and the detection of weeds in a farm field.
- the study of geostatistical mapping of certain chemical factors in agricultural soil.

This work could be supplemented by the study of optimal smoothing parameters using cross validation methods that have been proven in the field. It will also be extended to non-Gaussian process by assuming some hypotheses as for example when the cumulates are finite.

Acknowledgements

I would like to thank the anonymous referees for their interest in this paper and their valuable comments and suggestions.

References


Authors

Rachid Sabre

Responsible of MAIS, AgroSup, Member of the Laboratory Le2i, UMR CNRS 6306, University of Bourgogne, France.