Estimation & Dependence in Space & Time
Lecture 2: Non-Stationary Time Series

Sofia Olhede

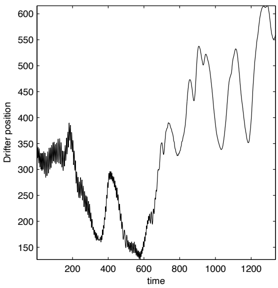
EPFL

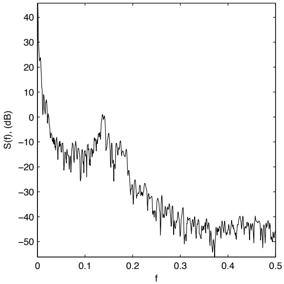
February 2, 2024

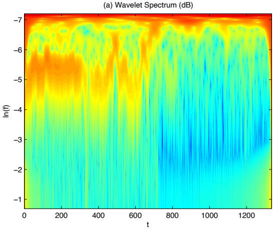
Models of Non-Stationary Processes

Life is Non-Stationary

- Many observed processes are NOT stationary.
- Data collected over time has natural windows of stability,
- Over a given window there are regions of a degree of smoothness etc,
- Non-Stationarity may arise in the mean, i.e. μ_t or $\Sigma(t_1, t_2)$ cannot be assumed to take a simpler form, even if over a region local stationarity is observed.
- Our intrinsic understanding of the **non-stationarity** depends on if the description of the stationary time series.







Representation in time or frequency. Do first or second order properties look stationary?

"Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of undergraduates."

D. Thomson, Jackknifing Multiple Window Spectra, 1994.

How do we formulate non-stationary models?

- We start from **stationary representations** of the process.
- a) **Time domain representations** of the process, i.e. $\mu \to \widetilde{\mu}_t$, and $s_\tau \to \widetilde{s}_{t,\tau}$.
- b) **Spectral Representation of the process**, i.e. $S(f) \to S(t,f)$.
- c) **Wold Representation**, MA or AR representation, becomes a time-varying MA or AR:

$$Z_t - \mu = \sum_{j=0}^{p} h_\tau \epsilon_{n-\tau} \to Z_t - \mu_t = \sum_{j=0}^{p} h_{\tau,t}\epsilon_{t-\tau} \tag{1}$$

- Only in special cases will the auto-covariance, spectrum and Wold representation admit interpretable objects, and never can you easily swap between the representations.

"Time is connected. And functions of time reflect this fact in their structure, not only in the tendency towards continuity shown by individual time functions, but even more obviously in the associated probability structures."

J. Tukey, Discussion, 1961.

What is time?
Time is **Global**, Time is **Relative**, and Structured behaviour in time is dependent or smooth. Thus

- Firstly define
$$\widetilde{s}_{t_1,t_2} = \mathbb{C}\text{ov}\{Z_{t_1}, Z_{t_2}\} \quad (2)$$
- This must be a valid auto-covariance sequence (positive semi-definite).
- Define global time $t = \frac{1}{2}(t_1 + t_2)$ and local time $\tau = t_2 - t_1$
- To make sense as a local function defined in τ for fixed t, it must vary (decay) more rapidly in τ than change in t.
- Thus consider:
$$s_{t,\tau} = \mathbb{C}\text{ov}\left\{Z_{t-\frac{1}{2}\tau}, Z_{t+\frac{1}{2}\tau}\right\} = \widetilde{s}_{t-\frac{1}{2}\tau, t+\frac{1}{2}\tau}.$$

Locally Stationary (a là Mallat) Processes [33]

- $s_{t,\tau}$ is only a function of τ if the process is stationary.
- Locally $s_{t,\tau}$ is not changing with t. Let $l(t')$ be the local interval of approximate stationarity. Assume for all $t_1 \in [t' - l(t'), t' + l(t')]$, $\widetilde{s}_{t_1,t_2} \approx s_{t',t_1-t_2}$ (rate of change of correlation).
- Let $d(t)$ be the local decorrelation length. For $t_1 \in [t' - l(t'), t' + l(t')]$, $\widetilde{s}_{t_1,t_2} \approx 0$ if $|t_1 - t_2| \geq d(t')$.

Definition (Locally Stationary Process [33, 5])

Locally stationary processes are such for which $l(t')$ and $d(t')$ can be defined to satisfy above eqns with

$$d(t') < \frac{l(t')}{2}. \tag{3}$$

Going from Time-Varying Models to Spectral Models [42, 46]:

- Assuming that ϵ_t is **stationary** it possesses a representation of:

$$\epsilon_t = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi ft} d\mathcal{Z}_\epsilon(f), \quad \mathbb{V}\mathrm{ar}(d\mathcal{Z}_\epsilon(f)) = S_\epsilon(f)\, df.$$

- We can then use (1) to understand the output, it follows that with $\varsigma_t(f) = H_t(f)e^{2i\pi ft}$:

$$\begin{aligned} Z_t - \mu_t &= \sum_{j=0}^{p} h_{\tau,t}\epsilon_{t-\tau} = \sum_{j=0}^{p} h_{\tau,t} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2i\pi f(t-\tau)} d\mathcal{Z}_\epsilon(f) \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} H_t(f)e^{2i\pi ft} d\mathcal{Z}_\epsilon(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \varsigma_t(f) d\mathcal{Z}_\epsilon(f). \end{aligned} \tag{4}$$

Locally Stationary Processes

- Referring to eqn (1), we state

$$\mathbb{V}\text{ar}(Z_t) = \int_{-\frac{1}{2}}^{\frac{1}{2}} |H_t(f)|^2 S_\epsilon(f)\, df. \tag{5}$$

- This seems to imply that the time-varying spectrum of $\{Z_t\}$ is

$$S_t(f) = |H_t(f)|^2 S_\epsilon(f), \tag{6}$$

  which at time t has "energy" $|H_t(f)|^2 S_\epsilon(f)$ associated with frequency f.
- **In no way can you make this statement, without further assumptions.**

Locally Stationary Processes II

- The problem arises because f **need not** correspond to "frequency" content.
- Revert back to eqn (4). It seems to be saying that $e^{2i\pi ft}$ is given a random weighting of $H_t(f)d\mathcal{Z}_\epsilon(f)$.
- What if $H_t(f) = e^{2i\pi f_o t}\widetilde{H}_t(f)$?
- Thus in general the interpretation of eqn (4) is suspect.
- We need to formulate when it is permitted to interpret eqn (6).

Oscillatory Processes

Oscillatory Processes

- We need to understand when eqn (4) is **interpretable**.
- We represent $H_t(f)$ by

$$H_t(f) = \int_{-\infty}^{\infty} e^{2i\pi\nu t}\, dK_f(\nu). \tag{7}$$

- We denote $\varsigma_t(f)$ an **oscillatory function** if uniformly in f

$$\arg_\nu \max |dK_f(\nu)| = 0 \tag{8}$$

- This ensures that $dK_f(\nu)$ is sufficiently smooth, and that $\varsigma_t(f)$ is associated with frequency f. In this case eqn (6) defines the **evolutionary spectral density** at time t wrt the family of functions $\{\varsigma_t(f)\}$.

Families of Oscillatory functions

- If Z_t is stationary and $\mathcal{F}$ is the family of complex exponentials then $S_t(f)$ is the sdf.
- It is convenient to adopt a standardization of $H_t(f) = 1$.
- In general there is a wealth of possible choices for families of $\varsigma_t(f)$, denoted $\mathcal{F}$.
- Would like to chose a suitable family $\mathcal{F}$: say to give the least variable $H_t(f)$.

Semi-Stationary Processes [42]

- Define $B_{\mathcal{F}}(f)$ by:

$$B_{\mathcal{F}}(f) = \int_{-\infty}^{\infty} |\nu||dK_f(\nu)|. \quad (9)$$

Definition (Semi-Stationary Process)

A family $\mathcal{F}$ is called semi-stationary if $B_{\mathcal{F}}(f)$ is bounded for all f. In this case the **characteristic width** is given by:

$$B_{\mathcal{F}} = \sup_f [B_{\mathcal{F}}(f)]^{-1} \quad (10)$$

A process $\{Z_t\}$ is semi-stationary if there exists a semi-stationary family $\mathcal{F}$ for which eqn (4) holds.

Locally Stationary (a là Dahlhaus) [14]

- Dahlhaus [14, 15] has developed theory for locally stationary processes.

Definition (Locally Stationary Process)

A sequence of stationary processes $Z_{t,N}$ is called locally stationary with transfer function $A^0(\cdot)$ and trend $\mu(\cdot)$ if there exists a representation

$$Z_{t,N} = \mu\left(\frac{t}{N}\right) + \int_{-\frac{1}{2}}^{\frac{1}{2}} A^0_{t,N}(f) e^{2i\pi ft}\, d\mathcal{Z}(f), \quad t \in [1, N), \tag{11}$$

where the following conditions are met:

- (a) $d\mathcal{Z}(f)$ is a stochastic process satisfying a set of cumulant conditions,
  (b) $A^0_{t,N}(f)$ approximates a continuous function in increasing N, $A(u,f)$ and $\mu(\cdot)$ is a continuous function.
- The smoothness of $A(u,f)$ guarantees the locally stationary behaviour of the process.
- Various models fit such as uniformly modulated and time-varying MA's, as well as, AR's.
- This introduces the important concept of **rescaled** time, $z = t/N$.

Wavelet Models

Wavelet Models

- Local stationarity was based on time-domain understanding. Semi-stationarity was based on time-frequency. What about time-scale?
- Recall that the 1-D MODWT and IMODWT decomposes a vector $\mathbf{Z} = [Z_{t_1}, \ldots, Z_{t_N}]^T$. as:

$$\begin{aligned} \widetilde{W}_{j,t}^{(Z)} &= \sum_{l=0}^{L_j-1} \widetilde{h}_{j,t-l}^{\circ} Z_{l \mod N}, \quad \widetilde{V}_{j,t}^{(Z)} = \sum_{l=0}^{L_j-1} \widetilde{g}_{j,t-l}^{\circ} Z_{l \mod N} \\ Z_t &= \sum_{j=1}^{J_0} \sum_{t=0}^{N} \widetilde{h}_{j,t-l}^{\circ} \widetilde{W}_{j,l}^{(Z)} + \sum_{t=0}^{N} \widetilde{g}_{j,t-l}^{\circ} \widetilde{V}_{j,t}^{(Z)}. \end{aligned} \tag{12}$$

Locally Stationary Wavelet Processes

- Can we define a stochastic process starting from eqn (12)?
- In a sense eqn (12) is like a IDFT.
- The representation of eqn (1) permits all frequencies.

Definition (Locally Stationary Wavelet Process [36])

The LSW processes are a sequence of doubly indexed stochastic processes $\{Z_{t,N}\}$ where $N = 2^J > 1$ having the representation in the mean-square sense:

$$Z_{t,N} = \sum_{j=1}^{J} \sum_{k=0}^{N} \widetilde{h}^{\circ}_{j,t-k} \alpha_{jk} \xi_{jk}, \qquad (13)$$

where α_{jk} and ξ_{jk} are independent random variables.

LSW II

The objects in the theorem satisfy:

- $\mathbb{E}(\xi_{jk}) = 0$,
- $\mathbb{C}\text{ov}(\xi_{jk}, \xi_{j'k'}) = \delta_{jj'}\delta_{kk'}$,
- $\forall j \geq 1, \quad \exists W_j(z)$ that for $z \in (0,1)$ is Lipschitz continuous and satisfies the following properties
- $\sum_{j=1}^{\infty} |W_j(z)|^2 < \infty$ uniformly in z,
- the Lipschitz constants L_j are uniformly bounded and satisfy $\sum_{j=1}^{\infty} 2^{-j} L_j < \infty$,
- $\exists C_j$ so that for any fixed T $\sup_{0 \leq k \leq N-1} |\alpha_{jk} - W_j(k/N)| \leq C_j/N$ and $\sum_{j=1}^{\infty} C_j \infty$.
- $z = k/N$ is **rescaled time**.

LSW III:

- We wish to measure the local energy of a LSW process.

Definition (Evolutionary Wavelet Spectrum [36])

The Evolutionary Wavelet Spectrum wrt $\widetilde{h}_{jk}$ of the sequence $\{Z_{t,N}\}$ is defined by:

$$S_j(z) = |W_j(z)|^2, \quad j = 1, \ldots, J, \quad z \in (0, 1) \tag{14}$$

- Gives a time-scale decomposition of $Z_{t,N}$ that can be deemed analogous to eqn (5).
- Let $\Psi_j(\tau) = \sum \widetilde{h}^{\circ}_{jk} \widetilde{h}^{\circ}_{jk-\tau}$ for $j = 1, 2, \ldots$ and $\tau \in \mathbb{Z}$.

LSW III:

- Using the autocorrelation wavelets $\Psi_j(\tau)$ we can construct valid autocovariance functions of quasi-stationary processes.

Definition (The Local Auto-Covariance [36])

The Local Auto-Covariance $c(z,\tau)$ of an LSW process with EWS $[S_j(z)]$ is defined as:

$$c(z,\tau)=\sum_{j=1}^{\infty} S_j(z)\Psi_j(\tau), \quad \tau\in\mathbb{Z}, \quad z\in(0,1). \tag{15}$$

- For stationary processes the dependence on z is redundant.
- The EWS is uniquely defined given the LSW process,
- $S_j(z)=\sum A_{jl}^{-1}\sum c(z,\tau)\Psi_l(\tau)$. ($A_{jl}$ is defined from $\Psi_l(\tau)$.

Classes of Nonstationary Processes

Karhunen-Loève Expansion

- We started from the linear representation of eqn (1), and noted the representation of Z_t in terms of uncorrelated elements ϵ_t.
- We can relax other aspects of the representation. This yields new classes of stochastic processes.

Definition (Karhunen Processes [25])

A process $Z(t)$ is a Karhunen process if it admits a representation of:

$$Z(t) = \int_{\mathbb{R}} d\mathcal{Z}(f)\psi_f(t), \tag{16}$$

but where the covariance structure of the $\{d\mathcal{Z}(f)\}$ process is given by

$$\mathrm{cov}(d\mathcal{Z}(f), d\mathcal{Z}^*(f')) = dS_K^{(I)}(f)\delta(f - f'). \tag{17}$$

Harmonizable Processes:

- So we could start from using the complex exponentials still.

Definition (Harmonizable Process [31])

A second-order process $Z(t)$ is (weakly) Harmonizable if it admits the representation:

$$Z(t) = \int_{\mathbb{R}} d\mathcal{Z}(f) e^{i2\pi ft}, \tag{18}$$

where the covariance structure of the $\{d\mathcal{Z}(f)\}$ process is given by

$$\mathrm{cov}(d\mathcal{Z}(f), d\mathcal{Z}^*(f')) = d^2 S_H^{(I)}(f, f').$$

- Eqn (18) provides a decomposition of the process $\{Z_t\}$ in complex exponentials. **Correlated** weights may be used in the expansion.

Equivalence of Karhunen and Harmonizable Classes of Processes.

- It transpires that despite their superficial dissimilarity the harmonizable class is a subset of the Karhunen class.
- For Karhunen Processes eqns (16) corresponds to the Karhunen-Loève expansion of Z_t.
- The functions $\{\psi_f(t)\}$ are determined from the eigenequation of the process.
- The covariance function can be retrieved from the eigensystem, this is known as Mercer's theorem.

Cyclostationary Models

- How do we reduce this (sparsity!)
- Classical sparse models include [18, 19]) for $C \in \mathbb{N}$

$$S(f_1, f_2) = \sum_{c=-C}^{C} S_c(f_2)\delta\left(f_1 - f_2 - \frac{c}{D}\right), \tag{19}$$

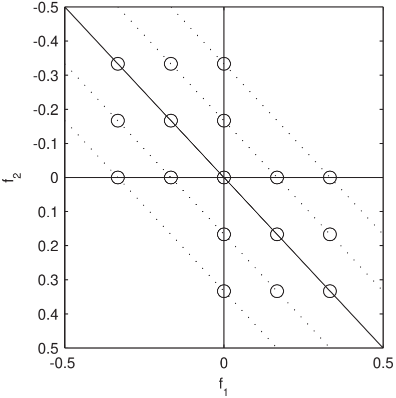
- This is the definition of a cyclostationary process.
- Identical to covariance

$$\gamma(t, l) \quad = \quad \sum_{c=-C}^{C} \gamma_c(l) e^{2i\pi \frac{c}{D} t}. \tag{20}$$

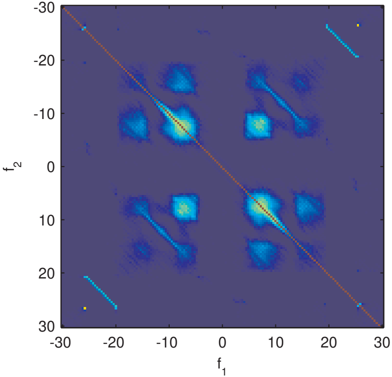
- It has been generalized to lines of arbitrary slope and location, [30] from

$$f_1 = f_2 + \frac{c}{D}.$$

Geometry of Sparse models



Not diverse/distributed enough



How change?

- We [38] start from the form of $s(t, l)$ and relax the strong form to

$$\gamma(t, l) = \sum_{c=-C}^{C} \varsigma_c(t, l) = \sum_{c=-C}^{C} a_c(t)\gamma_c(l)e^{2i\pi tc/D}. \quad (21)$$

- The amplitude $a_c(t)$ modifies the weighting assigned to each harmonic across the time course of observation.
- With $S(f+\nu, f) = \sum_c S_c(f+\nu, f)$ produces a spread of frequency:

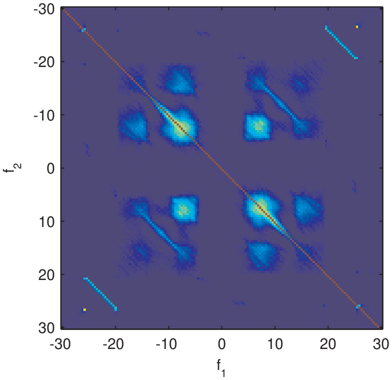
$$S_c(f+\nu, f) = \sum_{t,l} \varsigma_c(t, l)e^{2i\pi(t\nu + lf)} = S_c(f)A_c(\nu - \frac{c}{D}).$$

- We have taken $\nu = f_1 - f_2$ and $f = f_2$ to rotate the structure into the "right" frame of reference.

Islands of support

- We now have little "islands" of support, where the support is governed by the function $a_c(t)$.
- Seems arbitrarily simple to write out different functions when defining $S(f_1, f_2)$.
- Not all such functions are allowed: again we face validity as a covariance structure.

Distribution of frequencies



Larger sets of families

- Even more general classes may be defined:

Definition (Cramér Processes [11])

A process $Z(t)$ is a Cramer process if it admits a representation of:

$$Z(t) = \int_{\mathbb{R}} d\mathcal{Z}(f)\psi_f(t), \tag{22}$$

but where the covariance structure of the $\{d\mathcal{Z}(f)\}$ process is given by

$$\mathrm{cov}(d\mathcal{Z}(f), d\mathcal{Z}^*(f')) = dS_K^{(I)}(f)\delta(f - f'). \tag{23}$$

- The KF or 'asymptotically stationary' class, see Kampé de Feriet and Frenkiel [24] as well as Parzen [40] and Rozanov [43],
- the Cramér-Hida class of processes, see [12, 23, 8]

$$\begin{aligned} \text{Stationary} &\subset \text{Strongly Harmonizable} && (24) \\ &\subset \text{Weakly Harmonizable} && (25) \\ &\subset \text{Karhunen class} && (26) \\ &\subset \text{Cramer class.} && (27) \end{aligned}$$

More and more complicated models...

Why not useful?

"Given an arbitrary non-stationary process, the spectrum associated with the Wold-Cramer decomposition will often be meaningless."

G. Melard and A. Herteleer-De Schutter, Contributions to Evolutionary Spectral Theory, 1988.

More Recent Introductions:

- Deformation stationary processes were introduced by Sampson and Guttorp [44] for multi-dimensional processes.
- We assume that:
$$\widetilde{s}_{t_1,t_2} s(|g(t_1) - g(t_2)|), \tag{28}$$
where $g(t)$ is a non-linear mapping, $g: \mathbb{R} \Longrightarrow \mathbb{R}$.
- Thus a **warping** of time makes the process stationary.
- Such ideas have also been explored by [6].

Tempering

- The notion of **tempering** was introduced by Pintore and Holmes [41].
- Assume we have a parametric spectral density function $S(f|\boldsymbol{\theta})$ depending on the set of parameters $\boldsymbol{\theta}$.
- Define the localised sdf at position t by $S(f|\boldsymbol{\theta}) \mapsto [S(f|\boldsymbol{\theta})]^{\eta(t)}$.
- This **defines** a local covariance function

$$\widetilde{s}(t_1, t_2) = \int e^{2i\pi f(t_1 - t_2)} [S(f|\boldsymbol{\theta})]^{\eta(t_1)/2} [S(f|\boldsymbol{\theta})]^{\eta(t_2)/2}\, df. \qquad (29)$$

- This **defines** a valid autocovariance sequence [41].

Inference Methods for Non-Stationary Processes

Priestley's Inference approach

- Having introduced the class of non-stationary models, it is of course necessary to propose suitable inference procedures.
- Naturally the optimal inference method for evolutionary spectra **change** as we select the given non-stationary model.
- This corresponds to choosing a given semi-stationary family $\mathcal{F}$ for analysis.
- For given classes of models, specialized inference methods were proposed, i.e. see [21, 22, 13]. Naturally more general methods are necessary.

- Typical models include **piecewise stationary processes** [4], i.e. for some fixed change points $0 = u_0 < u_1 < \cdots < u_{m+1}$

$$Z_t = \sum_{k=1}^{m} \mathcal{I}(u_k \leq t/N < u_{k+1}) Z_t^{(k)}$$

where $Z_t^{(k)}$ are independent stationary processes,

- **blended stationary process** [39]

$$Z_t = \sum_{k=1}^{m} \mathcal{W}_k \left(\frac{\frac{t}{N} - u_k}{u_{k+1} - u_k} \right) Z_t^{(k)}.$$

- **uniformly modulated stationary processes** [3][p. 150] where

$$Z_t = A\left(\frac{t}{N}\right)\tilde{Z}_t,$$

- **autoregressive processes with time-varying coefficients**

$$\sum_{j=0}^{p} a_j\left(\frac{t}{N}\right) Z_{t-j} = \sigma\left(\frac{t}{N}\right)\varepsilon_t,$$

$a_0(u) \equiv 1$ and ε_t are independent rvs with mean 0 and variance 1.

Other approaches

- It is highly unsatisfactory to need to develop new methods for each proposed model.
- Adak [4] proposed a segmentation procedure that automatically identified lengths of the series that could be estimated as stationary.
- Adak's method can be applied to the models mentioned, and estimated the time-dependent spectrum. In the procedure the segmentation used CART [1] and smoothed the estimated spectrum.
- Kayhan *et al.* [26, 27] estimated the evolutionary periodogram. They modelled

$$Z_t = A_t(f)e^{2i\pi ft} + Y_t, \quad (30)$$

  where Y_t are all the contributions not associated with frequency f.
- $A_t(f)$ is estimated using a basis expansion.

Smooth Localized Complex EXponential (SLEX)

- Adak's method suffers from the usage of a non-orthogonal transform.
- Hence it is not possible to prove consistency, see [39].
- Furthermore, the segmentation method per se, may introduce blockiness in the estimated time varying sdf.
- A SLEX **basis function** is given by:

$$\phi_f(u) = \Psi_+(u)e^{2i\pi fu} + \Psi_-(u)e^{-2i\pi fu}.$$

- This is a generalization of the windowed Fourier basis.
- The SLEX basis vectors can be directly defined from this, see [39], and are denoted $\phi_{S,f_k,t}$.

SLEX continued

- The SLEX transform smoothly divides the time series dyadically.

$$d_{j,b}^{(X)}(f_k) = \frac{1}{\sqrt{M_j}} \sum_t Z_t \phi^*_{j,b,f_k,t}, \quad \phi_{j,b,f_k,t} = \phi_{S(j,b),f_k,t}. \tag{31}$$

- SLEX transform forms a library of orthonormal transforms. Hence can use the Best Basis Algorithm of Coifman and Wickerhauser [9] (see Day X) to determine the best basis selection.
- This estimates the time-varying spectrum.

Dahlhaus & Estimation

- Dahlhaus [14] defined the data window $h(x)$ with $h(x) = 0$ if $x \notin [0, 1]$ and

$$d_M(u, f) = \sum_{s=0}^{M-1} h\left(\frac{s}{M}\right) X_{[uN]-M/2+s+1,N} e^{-2i\pi fs}, \quad (32)$$

$$H_{k,M}(f) = \sum_{s=0}^{M-1} h^k\left(\frac{s}{M}\right) e^{-2i\pi fs}, \quad (33)$$

$$I_M(u, f) = \frac{\left|d_M(u, f)\right|^2}{2\pi H_{k,M}(0)}. \quad (34)$$

- Dahlhaus [14] then defined the Local Whittle likelihood Approximation, for shift q with $t_j = q(j-1) + M/2$ and $j = 1, \ldots, P$ where $N = q(P-1) + M$ or $u_j = t_j/N$ by:

Definition (Local Whittle likelihood Approximation)

From a given time series of length N, denoted $X_{t,N}$ we define the local Whittle Likelihood approximation by:

$$\mathcal{L}_N(\boldsymbol{\theta}) = \frac{1}{4\pi}\frac{1}{P}\sum_{j=1}^{P}\int_{-1/2}^{1/2}\left\{\log(S(u_j, f|\boldsymbol{\theta})) + \frac{I_M(u_j, f)}{S(u_j, f|\boldsymbol{\theta})}\right\} df. \tag{35}$$

Properties of $\mathcal{L}_N(\boldsymbol{\theta})$.

- The limit as $N \to \infty$ of $\mathbb{E}\big(N^{-1}\mathcal{L}_N(\boldsymbol{\theta})\big)$ as the asymptotic Kullback-Leibler [14].
- We define
$$\widehat{\boldsymbol{\theta}}_N = \arg_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \min \mathcal{L}_N(\boldsymbol{\theta}), \tag{36}$$
as the maximum Whittle likelihood estimator.
- [14] also discusses fitting time-varying ARs.
- [15] extended the methods to multiple dimensions.

Dahlhaus

- To be able to develop asymptotics for non-stationary time series some appropriate "large parameter" needs to be introduced.
- Just increasing the sample length is not sufficient, as the non-stationarity makes the effective number of parameters grow as the sample length grows.
- To this purpose Dahlhaus [14] introduced the concept of 'rescaled time'.
- Assume you collect data for $t = 0, \ldots, N-1$, then all theory is developed for $u = t/N$.
- This permits us to develop large sample theory.

Adaptive Covariance Estimation [33]

- Defining a locally stationary process by eqn. (3), Mallat *et al.* note that sde's often generate processes that are locally stationary.
- For locally stationary processes, one can construct local cosine vectors that approximately diagonalise the covariance matrix.
- It was assumed that there were multiple (R) replicates of the data.
- The covariance was estimated using 'best basis', see [9].

Macrotiles

- Donoho *et al.* [17] use macrotiles introduced by [33] to estimate the covariance of a non-stationary process.
- We still model $Z_t^{(r)}$ for $r = 1, \ldots, R$, as a zero-mean random process.
- A matrix $\boldsymbol{S} = [\widetilde{s}(t_1, t_2)]_{t_1, t_2}$ is estimated by $\widehat{\boldsymbol{S}} = [\widehat{s}(t_1, t_2)]_{t_1, t_2}$ by defining

$$\widehat{s}(t_1, t_2) = \frac{1}{R} \sum_{r=1}^{R} Z_{t_1}^{(r)} Z_{t_2}^{(r)}. \tag{37}$$

- This estimator is unbiased but has a large mean square error.
- Improving the estimator starts from $\widehat{\boldsymbol{S}} = \widetilde{\boldsymbol{S}} + \varepsilon$.

- The macrotile method considers a predefined family $\mathcal{F}$ of subspaces of vectors in $\mathbb{C}^{N^2}$.
- One constructs spaces $\mathcal{M}$ from a a dictionary $\mathcal{D} = \{\mathcal{B}_b\} = \left\{g_m^b\right\}_{m=1}^N$, where $\mathcal{B}_b$ is an orthogonal basis.
- We start by segmenting $[0, N^2]$ into sets $S = \{I_k\}$.
- The segmentation associated each I_k with a 1-D macrotile space $\mathbf{W}_k$, generated by $\sum_{m \in I_k} g_m^b$.
- $\mathcal{M} = \oplus_{k=1}^K \mathbf{W}_k$ defines a macrotile model.
- Donoho *et al* computes the best empirical model $\widehat{\mathcal{M}}$ from $Z_t^{(r)}$ and this is equivalent to finding the best basis in $\mathcal{D}$.
- Local Cosine basis used for applications in voice analysis [17].

Time-Varying Models and Wavelets

- Additionally multiscale methods have been combined with parametric models.
- Tsatsanis and Giannakis as well as Zheng [47, 51] model the observations Z_t by:

$$Z_t = \sum_{k=1}^{p} a_{t,k} Z_{t-k} + \varepsilon_t. \quad (38)$$

- They select a basis for representing $a_{t,k}$. Wavelet bases are proposed, and statistical procedure for selection are given.
- This is extended into time-varying ARMA, and MA [48].

Discussion

- This lecture introduced ways to relax the assumptions of stationarity.
- Temporal imhomogeneity, difference stationarity and frequency coupling are different ways to relax stationarity.
- General frameworks existed of stationary processes, locally stationary processes, locally stationary wavelet processes, Karhunen processes, harmonizable processes, Cramér processes etc.
- New estimation procedures are possible to design for these settings. Interpretation is harder.

References I

- L. Brieman, J. Friedman, R. Olshen and C. Stone,
  Classification and Regression Trees,
  Wadsworth, CA, 1984.

- D. Brillinger,
  Time Series: Data Analysis and Theory,
  SIAM, 2001.

- M. B. Priestley,
  Non-Linear and Non-Stationary Time Series,
  Academic Press, London, 1991.

References II

S. Adak,
Time-Dependent Spectral Analysis of Nonstationary Time Series,
J. Am. Stat. Assoc., vol. 93, pp. 1488–1501.

M. G. Amin,
Time and Lag Window Selection in Wigner-Ville Distribution,
in *Proc. IEEE Int. Conf. on Acoust. Speech and Signal Proc*, pp. 1529–1532, 1987.

R. G. Baraniuk and D. L. Jones,
Shear Madness – New Orthonormal Bases and Frames using Chirp Functions,
IEEE Transactions on Signal Processing, vol. 41, pp. 3543–3549, 1993.

M. Casdagli,
Chaos and Deterministic versus Stochastic Non-Linear Modelling,
J. Roy. Stat. Soc. B, vol. 54, pp. 303–328, 1992.

References III

G. Y. H. Chi,
Multiplicity and Representation Theory of Generalized Random Processes,
J. of Multivariate Analysis, vol. 1, pp. 412–432, 1971.

R. Coifman and M. Wickerhauser,
Entropy based Algorithms for Best Basis Selection,
IEEE Transactions on Information Theory, vol. 32, pp. 712–718, 1992.

J. Cooley, P. Lewis, and P. Welch,
The finite Fourier transform,
IEEE Trans. Audio Electroacoustics, vol. 17, pp. 77–85, 1969.

H. Cramér,
A contribution to the theory of stochastic process,
In *Proc. Second Berkeley Symp. Math. Statist. and Probability*, pp. 228–352, University of California, Berkeley, CA.

References IV

H. Cramér,
Structural and Statistical Problems for a class of Stochastic Processes
S. S. Wilks memorial lecture, Princeton University Press, Princeton, N. J.

G. R. Dargahi-Noubary, P. J. Laycock and T. Subba Rao,
Non-linear stochastic models for seismic events with applications in event identification,
Geophys. J. Roy. Astron. Soc., vol. 55, pp. 655–668, 1978.

R. Dahlhaus,
Fitting Time Series Models to Nonstationary Processes,
The Annals of Statistics, vol. 25, pp. 1–37, 1997.

R. Dahlhaus,
A likelihood approximation for locally stationary processes,
The Annals of Statistics, vol. 28, pp. 1762–1794, 2000.

References V

I. Daubechies,
Orthonormal Bases of Compactly Supported Wavelets,
Communications on Pure and Applied Mathematics, vol. 41, pp. 909–996, 1988.

D. L. Donoho, S. Mallat, R. von Sachs and Y. Samuelides,
Locally Stationary Covariance and Signal Estimation with Macrotiles,
IEEE Trans. Signal Proc., vol. 51, pp. 614–627, 2003.

E. G. Gladyshev.
Periodically correlated random sequences.
Soviet Math, 137, 385–388, 1961.

E. G. Gladyshev.
Periodically and almost periodically correlated random processes with continuous time parameter.
Theory Probab. Appl., vol. 8, 137â177, 1963.

References VI

Y. Grenier,
Time-Dependent ARMA modeling of Nonstationary Signals,
IEEE Trans. Acoust., Speech and Signal Proc., vol. 31, pp. 899–911, 1983.

J. K. Hammond,
On the response of single and multi degree-of-freedom systems to non-stationary random excitations,
J. Sound & Vibr., vol. 7, pp. 393–416, 1968.

J. K. Hammond,
Evolutionary Spectra in Random Vibrations,
J. Roy. Stat. Soc. B, vol. 35, pp. 167–188, 1973.

T. Hida,
Canonical Representations of Gaussian Processes and their applications,
Mem. Coll. Sci. Kyoto Univ. Sec. 4, vol. 32, pp. 109–155, 1960.

References VII

J. Kampé de Feriet and F. N. Frenkiel,
Correlation and spectra of nonstationary random functions,
Math. Comp., vol. 10, pp. 1–21, 1962.

K. Karhunen,
Über lineare Methoden in der Wahrscheinlichkeitsrechnung,
Ann. Acad. Sci. Fenn Ser A, I. Math., vol. 37, pp. 3–79, 1947.

A. S. Kayhan, A. El-Jaroudi, L. F. Chaparro,
Evolutionary Periodogram for Nonstationary Signals,
IEEE Transactions on Signal Processing, vol. 42, pp. 1527–1536 1994.

A. S. Kayhan, A. El-Jaroudi, L. F. Chaparro,
Data Adaptive Evolutionary Spectral Estimation,
IEEE Transactions on Signal Processing, vol. 43, pp. 204–213, 1995.

T.-H. Li and H.-S. Oh,
Wavelet spectrum and its characterization property for random processes,
IEEE Transactions on Information Theory, vol. 48, pp. 2922–2937, 2002.

References VIII

K.-S. Lii and M. Rosenblatt,
Line spectral analysis for harmonizable processes,
Proceedings of the National Academy of Sciences of the United States of America, 95, pp. 4800–4803, 1998.

K.-S. Lii and M. Rosenblatt,
Spectral analysis for harmonizable processes,
The Annals of Statistics, 30, 258–297, 2002.

M. Loève,
Fonctions aléatoires du secondre ordre,
In *Processes Stochastiques et Movement Browien*, by P. Lévy, pp. 228–352, Gauthier-Villars, Paris.

R. M. Loynes,
On the Concept of the Spectrum for Non-Stationary Processes,
J. Roy. Stat. Soc. B, vol. 30, pp. 1–30, 1968.

References IX

S. Mallat, G. Papanicolau and Z. Zhang,
Adaptive Covariance estimation for locally stationary processes,
Annals of Statistics, vol. 26, pp. 1–47, 1998.

W. Martin and P. Flandrin,
Wigner-Ville Spectral-Ananlysis of Non-Stationary Processes,
IEEE Trans. Acoust. Speech Signal Proc., vol. 33, pp. 1461–1470, 1985.

K. S. Miller,
Complex Gaussian Processes,
SIAM Review, vol. 11, pp. 544–567, 1969.

G. P. Nason, R. von Sachs and G. Kroisandt,
Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum,
J. Roy. Stat. Soc. B, vol. 62, pp. 271–292, 2000.

References X

G. P. Nason and R. von Sachs,
Wavelets in Time-Series Analysis,
Philosophical Transactions: Mathematical, Physical and Engineering Sciences, vol. 357, pp. 2511-2526, 1999.

S. C. Olhede and H. Ombao,
Modelling and Estimation of Covariance of Replicated Modulated Cyclical Time Series,
IEEE Transactions on Signal Processing, vol 61, pp. 1944â1957, 2013.

H. C. Ombao, J. A. Raz, R. von Sachs and B. A. Malow,
Automatic Statistical Analysis of Bivariate Nonstationary Time Series,
J. Am. Stat. Assoc., vol. 96, pp. 543–560, 2001.

E. Parzen,
Spectral Analysis of asymptotically stationary time series,
Bull. Internat. Statist. Inst., vol. 39, pp. 87–103, 1962.

References XI

A. Pintore and C. C. Holmes,
Non-stationary covariance functions via spatially adaptive spectra,
Imperial College London Statistics Section Technical Report, 2004.

M. B. Priestley.
Evolutionary Spectra and Non-Stationary Processes,
Journal of the Royal Statistical Society, B, vol. 27, pp. 204–237, 1965.

Y. A. Rozanov,
Spectral Analysis of Abstract functions,
Theory Probab. Appl., vol. 4, pp. 271–287, 1959.

P. D. Sampson and P. Guttorp,
Nonparametric Estimation of Nonstationary Spatial Covariance Structure,
J. Am. Stat. Assoc., vol. 87, pp. 108–119, 1992.

References XII

Arthur Schuster,
On the Periodicities of Sunspots,
Philosophical Transactions of the Royal Society of London. Series A, Vol. 206, pp. 69-100, 1906.

D. Tjøstheim,
Spectral Generating Operators for Non-Stationary Operators,
Advances in Applied Probability, vol. 8, pp. 831–846, 1976.

M. K. Tsatsanis and G. B Giannakis,
Time-varying system identification and model validation using wavelets,
IEEE Transactions on Signal Processing, vol. 41, pp. 3512–3523, 1993.

M. K. Tsatsanis and G. B Giannakis,
Subspace methods for blind estimation of time-varying FIR channels,
IEEE Transactions on Signal Processing, vol. 45, pp. 3084–3093, 1997.

References XIII

P. Whittle,
Estimation and information in stationary time series,
Arkiv för Matematik, vol. 2, pp. 423–434, 1952.

P. Whittle,
Gaussian Estimation in Stationary Time Series,
Bulletin of the International Statistical Institute, vol. 39, pp. 105–129, 1962.

Y. Zheng, Z. Lin and D. B. H. Tay,
Time-varying parametric system multiresolution identification by wavelets,
Int. J. of Syst. Sci., vol. 32, pp. 775–793, 2001.