



# Local stationarity of graph signals: insights and experiments

Benjamin Girault, Shrikanth S Narayanan, Antonio Ortega

## ► To cite this version:

Benjamin Girault, Shrikanth S Narayanan, Antonio Ortega. Local stationarity of graph signals: insights and experiments. Wavelets and Sparsity XVII, Aug 2017, San Diego, United States. pp.1-17, 10.1117/12.2274584 . hal-01574215

**HAL Id: hal-01574215**

**<https://inria.hal.science/hal-01574215>**

Submitted on 11 Aug 2017

**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.

# Local stationarity of graph signals: insights and experiments

Benjamin Girault, Shrikanth S. Narayanan, and Antonio Ortega

University of Southern California, Los Angeles, CA 90089, USA

## ABSTRACT

In this paper, we look at one of the most crucial ingredient to graph signal processing: the graph. By taking a step back on the conventional approach using Gaussian weights, we are able to obtain a better spectral representation of a stochastic graph signal. Our approach focuses on learning the weights of the graphs, thus enabling better richness in the structure by incorporating both the distance and the local structure into the weights. Our results show that the graph power spectrum we obtain is closer to what we expect, and stationarity is better preserved when going from a continuous signal to its sampled counterpart on the graph. We further validate the approach on a real weather dataset.

**Keywords:** Graph signal processing, stationary graph signals, graph learning

## 1. INTRODUCTION

Graph Signal Processing (GSP) is thriving as the tool of choice for studying data lying on arbitrary irregular structures.<sup>1,2</sup> By exploiting these structures, we have been able to provide reasonable extensions to classical definitions of the signal processing toolbox, including decompositions such as the graph Fourier transform, wavelet decompositions,<sup>3</sup> filterbanks,<sup>4</sup> or graph empirical mode decomposition,<sup>5</sup> but also tools to study the properties of said transforms such as the uncertainty principle,<sup>6</sup> or more recently our proposed model for stochastic graph signals: graph stationarity.<sup>7,8</sup>

In most of the literature, the graph structure is supposed to be known. Yet, the question of choosing the right graph is of utmost importance and often overlooked. Since graph signal processing is the study of signals in the context of a given graph, then working on a graph that is not adequate with respect to the actual structure of the data leads to a study whose conclusions may not be relevant. In this paper we tackle this problem with a data driven approach, and focus on providing evidence that some conventional approaches for selecting graph structures may not be ideal, even though they are widely used in the literature.

More specifically, we refer here to graphs constructed to represent data obtained by sampling a Euclidean space. The conventional approach uses a Gaussian kernel of the Euclidean distance to assign weights to the edges. Since working with complete graphs is expensive, the next step is often to sparsify the graphs: Those weights that correspond to distant nodes are removed to have reasonable algorithm complexity. This leads to a sparse graph whose vertices have a small number of neighbors compared to the total number of vertices.

This setting corresponds for example to sensor network applications that appear frequently in the graph signal processing literature, and is a well studied problem from a mathematical perspective in the context of manifold sampling.<sup>9</sup> Yet, theoretical results only guarantee that graph signal smoothness is related to the underlying continuous signal smoothness: The fact that the continuous Laplacian can be approximated by the graph Laplacian when applied to a signal does not guarantee that the graph Fourier transform can approximate the continuous Fourier transform, or that graph frequencies correspond to continuous frequencies. Consequently, sampling a continuous stationary signal leads to a non-stationary graph signal hinted by spectral correlations.

The further step of approximating the graph by removing those edges with small weights to make it sparse does not help either with regards to extending the Fourier transform. Observations show that this step can

---

Further author information: send correspondence to B. Girault

Matlab code: <http://www.benjamin-girault.com/>

E-mails: benjamin.girault@usc.edu, shri@sipi.usc.edu and antonio.ortega@sipi.usc.edu

actually change the properties of the graph quite drastically. Ultimately, such a method to define the graph relies on trade-offs between sparsity and accuracy.

We propose here an alternative approach involving learning a graph structure from data to overcome issues arising with these conventional methods. We study this approach on two typical cases where the common question is that of finding a graph that gives a relevant spectral model for the data. The first case is that of approximating the continuous Fourier modes: Can we model continuous frequencies on a Euclidean space using an irregular sampling of said space? The second case involves studying real data, for which distances are only one of the explanation to correlations, and we answer the question of learning a relevant graph to capture a good statistical model for the data.

Several methods exist in the literature to learn a graph structure from data.<sup>10–12</sup> We choose here the approach of Egilmez et al.<sup>10</sup> for its ability to constraint the connectivity of the graph to any set of edges and learn their weights. Doing so, we show that the learnt graphs lead to smoother graph Fourier modes in the sense of the Euclidean domain, sparse graphs, and, more importantly, better stochastic models for the data in the sense of smoother graph power spectrum and more interpretable spectral components.

The rest of this paper is organized as follows. [Section 2](#) introduces the necessary background on graph signal processing. [Section 3](#) recalls the classical definition of stationarity and our definitions of (global) graph stationarity<sup>7</sup> and local graph stationarity.<sup>13</sup> [Section 4](#) describes our approach and that of the literature, and both [Sec. 5](#) and [Sec. 6](#) applies our approach to an ideal case and a real case.

## 2. GRAPH SIGNAL PROCESSING

Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E}, w)$  be a graph with  $\mathcal{V}$  its set of vertices,  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  its set of edges, and  $w : \mathcal{E} \rightarrow \mathbb{R}_+$  its edges' weight function. We assume here that the graph is undirected, *i.e.*, whenever  $ij$  is an edge from vertex  $i$  to  $j$ ,  $ji$  is also an edge and  $w(ij) = w(ji)$ . Let  $A$  be the adjacency (or weight) matrix defined as  $A_{ij} = w(ij)$  whenever  $ij$  is an edge, and  $A_{ij} = 0$  otherwise. Note that since the graph is symmetric,  $A$  is also symmetric.

At the core of Graph Signal Processing (GSP), lies harmonic analysis of graph signals. Graph signals are functions  $x : \mathcal{V} \rightarrow \mathbb{R}$  or  $\mathbb{C}$ . Assuming a numbering of the vertices from 1 to  $N = |\mathcal{V}|$ ,  $x(i)$  will be denoted  $x_i$  such that  $x$  can be represented by the vector  $x = [x_1, \dots, x_N]^T$ . Everything that follows is independent of the numbering chosen. The most important tool for harmonic analysis is the so-called *Graph Fourier Transform* (GFT) defined as the projection onto a basis of graph signals called the *graph Fourier modes*. Definition of these modes varies. In this paper, the only requirement is to have a unitary transform. Let  $F$  be the matrix of the GFT. Then  $F^{-1} = [u_0 \dots u_{N-1}] = F^*$  is the inverse GFT, where the last equality follows from unitarity and  $\cdot^*$  is the complex conjugate operator. We obtain the following operations on a signal  $x$ :

$$\hat{x}(l) = \langle x, u_l \rangle = u_l^* x = \sum_{i=1}^N u_l^*(i) x_i \quad x_i = \sum_{l=0}^{N-1} \hat{x}(l) u_l(i). \quad (1)$$

The second requirement is to have a set of *graph frequencies* associated to the set of graph Fourier modes such that the lowest frequency is 0, and their magnitude is upper bounded by 1. We denote  $\lambda_l$  those frequencies, with  $\lambda_0 = 0$  associated to the lowest frequency  $u_0$ . Let  $\Lambda = \text{diag}(\lambda_0, \dots, \lambda_{N-1})$  be the diagonal matrix of graph frequencies. From now on, we denote  $Q = F^* \Lambda F$  the fundamental matrix of the graph<sup>\*</sup>.

One example of modes is the set of eigenvectors of the Laplacian matrix  $L = D - A$ , where  $D$  is the diagonal degree matrix verifying  $D_{ii} = \sum_j A_{ij}$ , associated to the eigenvalues of  $L$ , normalized (the exact normalization is not a sensitive parameter here and the interested reader can find one our previous work<sup>14</sup>), as graph frequencies.<sup>1</sup> We obtain then  $Q = \frac{1}{\alpha} L$ , with  $\alpha$  this normalization factor. Other definitions of graph Fourier modes exist,<sup>2</sup> however, the rest of this paper is not dependent on any specific GFT.

---

<sup>\*</sup>This matrix  $Q$  is sometimes written  $S$  and called the *shift matrix* of the graph. Usual choice of  $Q$  include the adjacency matrix and the Laplacian matrix. These matrices can be identified as first order differential operator (gradient) or second order ones. *Shift* however refers only to gradients while theory applies to both, hence our naming choice of *fundamental matrix* instead of *shift matrix*.

The second harmonic analysis tool we are using in this paper is a localized decomposition. Let  $\{g_{i,m}\}_{i,m}$  be a family of atoms (*i.e.*, graph signals) such that  $g_{i,m}$  is localized on vertex  $i$ , and  $m$  indexes the span of the atom. Additionally, we suppose these atoms to be of unit norm:  $\|g_{i,m}\|_2 = 1$ . The decomposition is then defined as a projection onto this family similarly to the GFT:  $W_x(i, m) = \langle x, g_{i,m} \rangle$ . One such family is the *Spectral Graph Wavelet Transform*<sup>3</sup> (SGWT) defined using the spectral properties of the atoms. More precisely,  $g_{i,m}$  is the localization around vertex  $i$  of  $m^{\text{th}}$  scale kernel  $g_m$  defining these spectral properties:

$$\widehat{g_m}(l) = \exp(-\kappa s_m \lambda_l) \quad g_{i,m} = c_{i,m} T_i g_m, \quad (2)$$

where  $\{s_m\}_m$  is a set of scales and  $T_i$  is the localization operator verifying  $\widehat{T_i x}(l) = \widehat{\delta_i}(l) \widehat{x}(l)$ . This operator has been shown<sup>3</sup> to effectively localize a smooth kernel such as  $g_m$  above. In this particular case of an exponential kernel, it is shown that there is also an exponential decay of the energy of an atom  $g_{i,m}$  with the distance (on the graph) to vertex  $i$ , validating the concept of localized atom, and decomposition.

### 3. STATIONARITY

In this section we recall a classical model for stochastic signals, the *stationarity* model. We do so both with a Euclidean setting, and with a graph setting.

#### 3.1 Stationarity for Euclidean Spaces

In the rest of this paper, we denote with bold character stochastic signals. Let  $\mathbf{s}$  be a stochastic signal. *Wide Sense Stationarity* (WSS) is then defined for Euclidean spaces as:<sup>15</sup>

$$\forall x, \quad \mathbb{E}[\mathbf{s}(x)] = \mu_{\mathbf{s}} \quad \forall x_1, x_2, \quad \mathbb{E}[\mathbf{s}(x_1) \mathbf{s}^*(x_2)] = \gamma_{\mathbf{s}}(x_1 - x_2). \quad (3)$$

In other words, the first moment is constant, and the second moment depends only on the coordinate difference  $x_2 - x_1$ . An equivalent definition involves the statistical invariance through translation:

$$\forall x, dx, \quad \mathbb{E}[\mathbf{s}(x)] = \mathbb{E}[\mathbf{s}(x + dx)] \quad \forall x_1, x_2, dx, \quad \mathbb{E}[\mathbf{s}(x_1) \mathbf{s}^*(x_2)] = \mathbb{E}[\mathbf{s}(x_1 + dx) \mathbf{s}^*(x_2 + dx)]. \quad (4)$$

Although equivalent for Euclidean spaces, they are not in a graph setting as we will be discussed in [Sec. 3.2](#): Only [Eq. \(4\)](#) can be directly transposed in a graph setting while [Eq. \(3\)](#) is more complex to transpose due to the coordinate difference  $x_1 - x_2$ .

WSS is a simplification of the *Strict Sense Stationarity* (SSS) definition where a signal is SSS if and only if it is statistically the same after translation. In other words, all the statistical moments of a given signal are equal, making it a particularly difficult, if not impossible, property to verify.

Intuitively, stationarity is the property that the signal does not depend on the origin of the system of coordinates: any changes of this origin leaves the signal unaffected. Conversely, a non-stationary signal is one that exhibit local properties whose location would be translated if the origin is modified.

#### 3.2 A Global Definition for Graphs

Global stationarity, or stationarity in short, is obtained by directly extending the concept of statistical invariance to the graph setting:

**DEFINITION 3.1 (WEAK STATISTICAL INVARIANCE (WSI)).** *A stochastic (graph or Euclidean) signal  $\mathbf{s}$  is said WSI with respect to the operator  $H$  if and only if its first two statistical moments are invariant through  $H$ :*

$$\mu_{\mathbf{s}} = \mu_{H\mathbf{s}} \quad R_{\mathbf{s}} = R_{H\mathbf{s}},$$

with  $\mu_{\mathbf{s}} = \mathbb{E}[\mathbf{s}]$  the first moment (mean) and  $R_{\mathbf{s}} = \mathbb{E}[\mathbf{s}\mathbf{s}^*]$  the second moment (correlation matrix).

In a Euclidean setting,  $H$  is a shifting operator, *e.g.* the time shift  $T\{s\}(t) = s(t-1)$  for temporal signals or a combination of the spatial shifts  $T_x\{s\}(x, y) = s(x-1, y)$  and  $T_y\{s\}(x, y) = s(x, y-1)$  for a 2D signal such as

in image<sup>†</sup>. Unfortunately, the notion of translation on a graph is not well defined, since there is no immediate system of coordinates, or any composition law on the set of vertices (if  $i$  and  $j$  are vertices, such a law would map them to a new vertex  $i \circ j$ , similarly to  $x + y$  being itself a new coordinate), for which a translation of the signal is equivalent to a translation of the system of coordinates.

We proposed recently a new operator<sup>8,14</sup> on graph signals whose behavior can be compared to that of a diffusion operator,<sup>16</sup> the *Graph Translation*:

$$T_G = \exp(-\imath\pi Q). \quad (5)$$

Contrary to  $Q$ , this operator is unitary: It does not alter the energy of the signal. This property is a natural assumption for an invariance operator since the second moment cannot be invariant if the energy changes. Finally, diffusion invariance is a natural parallel to shift invariance since it is the primary use of graphs and graph signals. We introduced the graph WSS definition using WSI through the graph translation:

**DEFINITION 3.2 (GRAPH WSS<sup>7</sup>).** *A stochastic graph signal  $\mathbf{s}$  is graph WSS if and only if it is WSI through the graph translation  $T_G$ :*

$$\mu_{\mathbf{s}} = \mu_{T_G \mathbf{s}} \quad R_{\mathbf{s}} = R_{T_G \mathbf{s}},$$

Note that the requirement on the correlation matrix can equivalently be replaced by the same requirement on the covariance matrix  $\Sigma_{\mathbf{s}}$ . However, due to the fact that the first moment needs to be estimated prior to estimating the covariance matrix, we will favor the correlation matrix  $R_{\mathbf{s}}$  when studying real data. Finally, using the GFT on the signal  $\mathbf{s}$ , we obtain a spectral characterization of WSS graph signals:<sup>7</sup>

**THEOREM 3.3 (GRAPH WSS SPECTRAL CHARACTERIZATION).** *A graph signal  $\mathbf{s}$  is WSS if and only if:*

$$\forall \lambda_l \neq 0, \quad \mu_{\mathbf{s}}(l) = 0 \quad \forall \lambda_l \neq \lambda_k, \quad S_{\mathbf{s}}(l, k) = 0,$$

where  $S_{\mathbf{s}} = R_{\mathbf{s}}$  is the spectral correlation matrix. In other words, graph frequencies are uncorrelated, and the mean is the lowest frequency of the signal (a DC component). This also gives a formal relation between the correlation matrix  $R_{\mathbf{s}}$  and the covariance matrix  $\Sigma_{\mathbf{s}}$ :

$$\Sigma_{\mathbf{s}} = R_{\mathbf{s}} - \mu_{\mathbf{s}}\mu_{\mathbf{s}}^* = R_{\mathbf{s}} - \mu_{\mathbf{s}}(0)u_0u_0^*,$$

using the fact that  $\Sigma_{\mathbf{s}} = S_{\mathbf{s}} - \mu_{\mathbf{s}}(0)E_{1,1}$ , where  $E_{1,1}$  is the matrix with zeros everywhere except for the top left element. From [Thm. 3.3](#), we can define the *graph Power Spectrum Density* as the variance of each spectral component:

$$\hat{\gamma}_{\mathbf{s}}(l) = \mathbb{E} \left[ |\hat{s}(l)|^2 \right] = S_{\mathbf{s}}(l). \quad (6)$$

In turn, we can write the elements of the correlation matrix using the PSD:

$$R_{\mathbf{s}}(i, j) = (T_j \gamma_{\mathbf{s}})(i). \quad (7)$$

Finally, two alternative extensions of stationarity for stochastic graph signals exist in the litterature following our initial work.<sup>7</sup> These are summarize on [Tab. 1](#).

**Perraudin and Vandergheynst approach<sup>17</sup>** This work assumes that the graph power spectrum of a WSS graph signal can be written as a function of the graph frequency. These graph signals are also WSS according to our definition (the converse is not always true).

Table 1. Comparison of graph stationarity approaches.

Paper	$Q$	1 <sup>st</sup> moment	2 <sup>nd</sup> moment	Constraints
[7]	$Q$	$\mu_{\mathbf{s}} = \mu_{T\mathbf{s}}$	$\mathbb{E}[\mathbf{s}\mathbf{s}^*] = \mathbb{E}[(T\mathbf{s})(T\mathbf{s})^*], \quad T = \exp(-i\pi Q)$	$Q$ normal
[17]	$L$	$\mu_{\mathbf{s}} = \alpha \mathbb{1}$	$\Sigma_{\mathbf{s}}(i, j) = (T_j \gamma_{\mathbf{s}})(i), \quad \hat{\gamma}_{\mathbf{s}}(l) = \mathbb{E}[ \hat{\mathbf{s}}(l) ^2]$	no multiple eigenvalues, $\hat{\gamma}_{\mathbf{s}}(l) = f(\lambda_l)$
[18]	$Q$	$\mu_{\mathbf{s}} = 0$	$\mathbb{E}[(Q^a \mathbf{s})((Q^*)^b \mathbf{s})^*] = \mathbb{E}[(Q^{a+c} \mathbf{s})((Q^*)^{b-c} \mathbf{s})^*]$	$Q$ normal

**Marques et al. approach<sup>18</sup>** Here, the invariance is more involved and is with respect to the fundamental matrix  $Q$  of the graph which is not an isometric operator. As a consequence, the first moment is assumed null. The invariance is actually equivalent to  $Q$  and  $R_{\mathbf{s}}$  commuting ( $QR_{\mathbf{s}} = R_{\mathbf{s}}Q$ ) which is itself equivalent to our second moment assumption. These graph signals are also WSS according to our definition (the converse is not true because of the first moment requirement).

### 3.3 A Local Definition for Graphs

Following our work on stationarity for graph signals, and inspired by the work of Perraudin and Vandergheynst,<sup>17</sup> we introduced recently a notion of local stationarity.<sup>13</sup> Given a set of localized atoms  $\{g_{i,m}\}_{i,m}$ , the *local power spectrum* is given by:

$$S_{\mathbf{s}}(i, m) := \mathbb{E} \left[ |W_{\mathbf{s}}(i, m)|^2 \right] = \mathbb{E} \left[ |\langle \mathbf{s}, g_{i,m} \rangle|^2 \right]. \quad (8)$$

Local stationarity is then the property that no matter where we look on the graph, the signal is the same. Using the local PSD above, this translates into:

$$\forall i, j, m, \quad S_{\mathbf{s}}(i, m) = S_{\mathbf{s}}(j, m)$$

Contrary to the global definition in [Sec. 3.2](#), this definition involves the invariance of local descriptors of the signal, hence the name of *local stationarity*.

## 4. GRAPHS FOR EUCLIDEAN SPACE SAMPLINGS

In this section we explore an often overlooked aspect of graph signal processing: the selection of the right graph. In particular, we focus here on sampling Euclidean spaces. These are for example sensor networks, or feature spaces in machine learning. We recall the usual structure used, and then describe our approach using stationary graph signals and graph learning to attain the goal of a better statistical model for the data.

### 4.1 State of the Art

Given a set of measurement point  $\{x_i\}_{i \in \mathcal{V}}$  of a Euclidean space  $\mathcal{S}$ , how do we connect them into a graph structure? Literature on Graph Signal Processing have been using a standard result from differential geometry<sup>9</sup> to justify the use of a Gaussian kernel of the distance as edge weights:

$$w(ij) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right), \quad (9)$$

for some parameter  $\sigma$ . The result in question is stated in [Thm. 4.1](#).

**THEOREM 4.1.** [[9, Theorem 3.1](#)] *Let  $\tilde{s} : \mathcal{S} \rightarrow \mathbb{R}$  be a continuous signal and  $\Delta$  the Laplace-Beltrami operator of  $\mathcal{S}$ . Let  $\mathcal{G}^{(N)}$  be a series of graphs constructed by uniformly sampling  $\mathcal{S}$ , with all edges weighted by a Gaussian kernel of the Euclidean distance. Let  $L^{(N)}$  be the associated graph Laplacians. Then, under some assumptions on  $\sigma$ , and rescaling of the Laplacian, we have the following convergence result for the sampled signals  $s^{(N)}$  of  $\tilde{s}$  on the graphs:*

$$L^{(N)} s^{(N)} \xrightarrow{N \rightarrow \infty} \Delta \tilde{s}.$$

---

<sup>†</sup>For dimensions greater than 2, stationarity is actually defined as the invariance through a set of operators, e.g.  $\{T_x, T_y\}$  in the 2D case.

Application of this result in the context of graph signal processing is challenging since this result only gives guarantees for an arbitrary continuous signal  $\tilde{s}$ . Yet, the ideal convergence result would be that the graph Fourier modes converge to the continuous Fourier modes. Since those are defined as eigenvectors of the Laplacian, rather than sampling of a continuous Fourier mode, this paper<sup>9</sup> does not give any guarantees with that regard.

As a consequence, the graph spectrum of a graph signal does not inform on the continuous spectrum of the underlying sampled continuous signal. The only guarantee is that given a signal  $s$  on the graph, and its spectral representation  $\hat{s}$ , we have that  $s^* L s = \sum_l |\hat{s}(l)|^2 \lambda_l$  is a good representative of the smoothness  $\langle \tilde{s}, \Delta \tilde{s} \rangle$  of the underlying continuous signal  $\tilde{s}$ . In particular, for a continuous Fourier mode  $\tilde{e}$  and  $e$  the sampled graph signal,  $e^* L e = \sum_l |\hat{e}(l)|^2 \lambda_l$  is a good representative of the actual frequency, but  $\hat{e}$  is not necessarily localized in the graph spectral domain.

The second question with this result is the value of  $\sigma$  in Eq. (9). Using the value used in [9, Theorem 3.1] leads to a value so small that the graph is almost disconnected: Most of the edges have weight very close to 0. The question then is how to choose this crucial parameter. The convention is to just use one value that gives interesting result, not too high to avoid all weights being close to 1, and not too small to avoid having almost all weights close to 0. One heuristic states that a good choice of  $\sigma$  is a third of the mean distance between vertices.<sup>19</sup> We adopt this heuristic to compare the Gaussian weights approach to ours as it gives reasonable results.

Furthermore, this complete graph is rarely used without sparsifying it first. In that case, those edges between distant vertices are removed. This step is often desirable for complexity reasons of GSP algorithms: Sparsity of the graph leads to sparsity of matrices and efficient matrix-vector or matrix-matrix multiplications. We investigate two approaches here: Keep the edge between vertices in a given radius, denoted *THR* graph, or keep the  $K$  nearest vertices, denoted *KNN* for  $K$ -nearest neighbors. The complete graph is denoted *CMP*.

The result of Thm. 4.1 does not include the sparsification case. As a consequence, there is no guarantee that those sparse graphs are good representations of the underlying Euclidean space.

## 4.2 Gaussian Random Fields

In this paper, we propose a data-driven approach to define the graph. More precisely, our approach aims at following more closely the result of Thm. 4.1: we define the graph from a given continuous signal. In this section, we describe a continuous signal model that is a good candidate to infer the graph from, and its corresponding graph signal model. We consider here only zero mean signals, but this can easily be extended to arbitrary means.

**Continuous model** From now on, bold characters designate stochastic quantities. Let  $\mathbf{s}(x)$  be a  $D$ -dimensional continuous stochastic signal with  $x \in \mathbb{R}^D$  and:

$$\mathbf{s}(x) = \int \sqrt{\gamma_{\mathbf{s}}(\nu)} \cos\left(i2\pi(x^T \nu + \phi(\nu))\right) d\nu,$$

for some prescribed *Power Spectrum Density* (PSD)  $\gamma_{\mathbf{s}}(\nu)$  of frequency parameter  $\nu \in \mathbb{R}^D$ , and  $\phi(\nu)$  i.i.d. phases uniformly drawn in  $[0, 1]$ .  $\mathbf{s}(x, y)$  is called a *Gaussian Random Field* (GRF). Such a phase randomization is known to stationarize a non-stationary continuous signal, and is used to test stationarity using *surrogates*.<sup>20</sup>  $\mathbf{s}$  is therefore a continuous WSS signal. We can compute its autocorrelation function:

$$R_{\mathbf{s}}(x_1, x_2) = \mathbb{E}[\mathbf{s}(x_1) \mathbf{s}^*(x_2)] = \int \gamma_{\mathbf{s}}(\nu) \cos\left(i2\pi(x_1 - x_2)^T \nu\right) d\nu = R_{\mathbf{s}}(x_1 - x_2), \quad (10)$$

and its spectral autocorrelation function using the Fourier transform of the autocorrelation function:<sup>8, 21</sup>

$$S_{\mathbf{s}}(\nu) = \mathcal{F} \{ \mapsto R_{\mathbf{s}}(\delta x) \} = \gamma_{\mathbf{s}}(\nu). \quad (11)$$



**Graph signal model** The equivalent of this model in a graph setting is called a *Gaussian Markov Random Field* (GMRF). A signal  $\mathbf{s}$  is GMRF if it follows the distribution  $\mathcal{N}(0, \Sigma)$ , a multivariate Gaussian distribution with zero mean, and covariance  $\Sigma$ . When the covariance matrix is jointly diagonalizable with the fundamental matrix of the graph, a condition which is met when the graph Fourier matrix diagonalize  $\Sigma$ , then a GMRF model can be described in the graph spectral domain as:

$$\mathbf{s}(i) = \sum_{l=0}^{N-1} \hat{\mathbf{s}}(l) u_l(i),$$

where the variables  $\hat{\mathbf{s}}(l) \sim \mathcal{N}(0, \gamma_{\mathbf{s}}(l))$  independent variables. Therefore,  $\mathbf{s}$  is graph WSS. Its spectral autocorrelation matrix is a diagonal matrix verifying:

$$S_{\mathbf{s}}(l, l) = \gamma_{\mathbf{s}}(l). \quad (12)$$

### 4.3 A Data-Driven Approach to Choosing a Graph

Observing both Eq. (11) and Eq. (12) reveals an interesting parallel: Both spectral correlations are equal to the power spectrum. Starting from this observation, we approach the problem of choosing a graph by learning it such that the continuous PSD and the graph PSD coincide as functions of the frequencies. To do that, instead of using Gaussian weights with a somewhat arbitrary parameter  $\sigma$ , we learn these weights using a GRF model sampled on the set of vertices. Equation (10) gives then the correlation matrix of the stochastic graph signal obtained using the coordinates of the vertices.

To learn the weights, we use the approach of Egilmez et al.<sup>10</sup> In this work, the Laplacian matrix is optimized to make it coincide with the precision matrix  $\Omega_{\mathbf{s}} = \Sigma_{\mathbf{s}}^{\dagger} = R_{\mathbf{s}}^{\dagger}$  of the data, with  $\cdot^{\dagger}$  the pseudo-inverse:

$$\min_L \text{Tr}(L\Sigma) - \log |L| \quad \text{subj. to } L \text{ is a valid Laplacian,}$$

with  $|L|$  the pseudo-determinant of the matrix. This optimization function is actually a specific instance of a maximum a posteriori estimation of the Laplacian given a GMRF model  $\mathcal{N}(0, L^{\dagger})$  [10, Proposition 5]. This corresponds to the graph PSD  $\gamma_{\mathbf{s}}(0) = 0$  and  $\gamma_{\mathbf{s}}(l) = 1/\lambda_l$ .

Since Thm. 4.1 shows a parallel between the continuous Laplacian and the graph Laplacian, and since the eigenvalues of the continuous Laplacian are squared frequencies, the graph equivalent to continuous the PSD  $\gamma(\nu) = 1/\nu^2$  is the graph PSD  $\gamma(\lambda) = 1/\lambda$ . Therefore, learning the graph from a covariance matrix is here equivalent to considering the data as a continuous GRF signal  $\tilde{\mathbf{s}}$  with PSD  $\gamma_{\tilde{\mathbf{s}}}(\nu) = 1/\nu^2$ .

We propose here to learn the graph structure from a Euclidean GRF model in  $1/\nu^2$ . In Sec. 5, we consider the ideal case for a two dimensional Euclidean plane with GRF model in  $1/\nu^2$ , and in Sec. 6, we look at a real case, where the exact geometry of the underlying space is unknown, and for which we assume the data to be sampled from such a GRF model on this unknown space.

Finally, to solve the question of learning a sparse graph, we use the conventional approach to select vertices. In other words, in this paper, we use  $K$ -nearest neighbors or distance thresholding to limit the number of edges of the graph, and perform the learning algorithm on those weights alone. The goal is to take into account the distance, and counterweight the particular local structure of the graph by learning the graph weights. In Sec. 5 and Sec. 6 we are using the same set of edges to better compare the Gaussian and the learnt approaches.

## 5. EXPERIMENT WITH A GRF MODEL

In this section, we look at the ideal two dimensional Euclidean space with a GRF model  $\tilde{\mathbf{g}}$  whose PSD is  $\gamma_{\tilde{\mathbf{g}}}(\nu_x, 0) = 1/\nu_x^2$  and  $\gamma_{\tilde{\mathbf{g}}}(0, \nu_y) = 1/\nu_y^2$  for  $\nu_x, \nu_y \in \{1, \dots, N/2\}$ , with  $N$  the number of sampled vertices in the space. We compare several properties of the graph Fourier transform and of the power spectrum of the graph signal sampled from this GRF model, using six different graphs. The first group of graphs is weighted by a Gaussian kernel, and the second is weighted by our learning approach. Each group is a set of three graphs that differ only by the sparsity: complete (CMP),  $K$ -nearest neighbor (KNN), and thresholded distance (THR). In this section, the graph has  $N = 200$  vertices,  $K = 6$ , and the radius for THR is set to have the same mean number of neighbors for both KNN and THR.



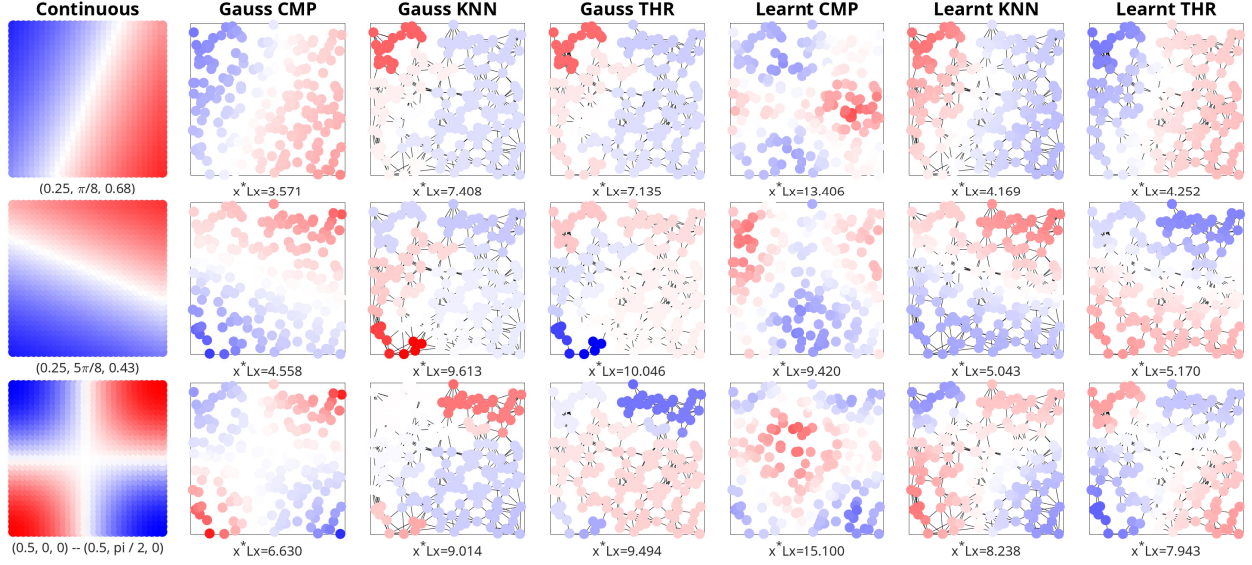


Figure 1. Comparison of the graph Fourier modes #2, #3, and #4 (from top to bottom) for various graphs. Edges are omitted for the complete graphs for clarity. Values range from  $-0.3$  (blue) to  $0.3$  (red) including  $0$  (white). The left columns shows a corresponding 2D Fourier mode of frequency  $\nu$ , rotation  $\theta$  in the 2D plane and phase shift  $\phi \in [0, 1]$ . The parameters  $(\nu, \theta, \phi)$  are given below the figures, with the bottom left one corresponding to a modulation of two Fourier modes. Laplacian variations are computed with respect to Gauss CMP graph Laplacian.

## 5.1 Graph Fourier Transform

Studying the spectral properties of the graph signal is relevant only if the graph Fourier modes do correspond to meaningful modes, that is, if we were to extrapolate them on the underlying Euclidean space, then the graph frequency should be a good indication of the continuous frequency of the extrapolated signal.

Qualitatively, we show on Fig. 1 several graph Fourier modes along with their closest continuous Fourier mode (leftmost column). Two remarks are in order: (i) the sparsified graphs *Gauss KNN* and *Gauss THR* show strong clustering effect (almost piecewise constant), and (ii) the graph *Learnt CMP* has Fourier modes that do not look like continuous Fourier modes.

The first point is important here as the number of edges removed can have a high impact on the graph Fourier modes. Here the mean number of neighbors (around 7) can often be found in the literature, but we see that drawing any conclusion on the graph Fourier spectrum of the sampled signal is difficult. The same remark applies for small values of  $\sigma$  (see Sec. 5.4). The second point can be interpreted as overfitting of the learning approach since the number of parameters (edge weights) is close to the size of the data (covariance matrix).

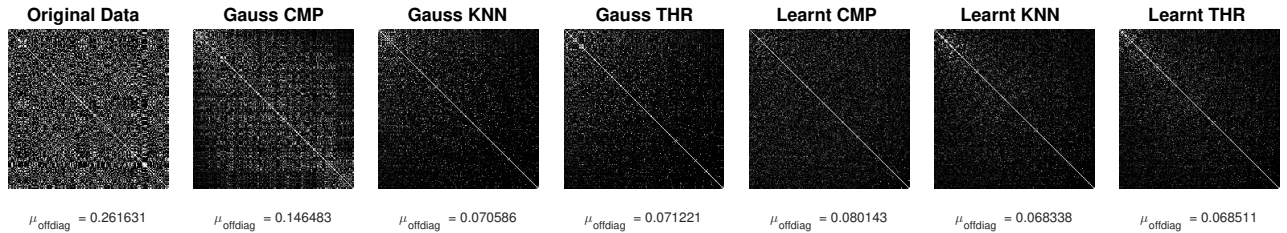


Figure 2. Theoretical Pearson correlation matrix of the sampled continuous GRF signal model  $\mathbf{g}$ . Values range from  $0$  (black) to  $0.5$  (white). Contrast is increased for better visibility and some entries are actually greater than  $0.5$ . For reference, the spatial Pearson correlation matrix is given (left), and the mean Pearson correlation (off-diagonal elements) is given below the matrices.

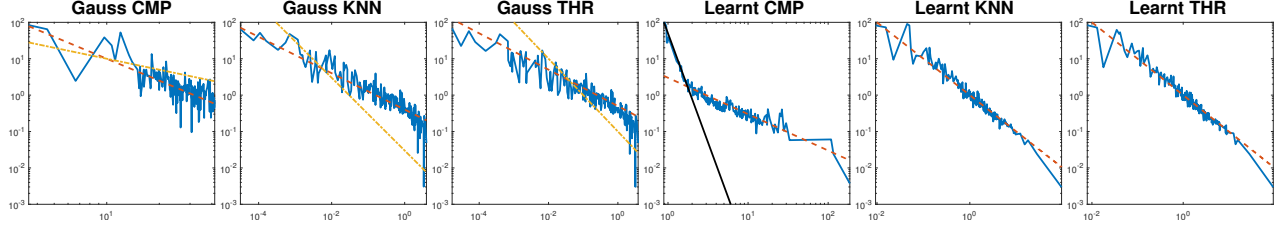


Figure 3. Graph PSD for the sampled continuous GRF signal  $\mathbf{g}$  using different graphs. The red dashed lines corresponds to a function  $\gamma(\lambda) = a\lambda^{-b}$ .  $(a, b)$  parameters are from left to right:  $(e^{1000}, 2.0)$ ,  $(e^{0.4}, 0.5)$ ,  $(e^{0.5}, 0.5)$ ,  $(e^3, 1.0)$ ,  $(e, 1.0)$ ,  $(e, 1.0)$ . The yellow dash-dotted line corresponds to the case  $b = 1.0$ . The plain black line for Learnt CMP corresponds to  $b = 6.0$ . Note that the lowest frequency 0 is not shown here because of the logarithmic scale on the horizontal axis.

We can further study these issues quantitatively using the graph Laplacian variation on *Gauss CMP*. We know<sup>9</sup> that given a continuous signal  $\tilde{s}$ , the output  $Ls$  of the graph Laplacian  $L$  applied to the sampling  $s$  of  $\tilde{s}$  converges exponentially as the number of vertices increases to the output  $\Delta\tilde{s}$  of the continuous Laplacian  $\Delta$  applied to  $\tilde{s}$ . Consequently,  $s^*Ls$  is a good proxy for the quantity  $\langle \tilde{s}, \Delta\tilde{s} \rangle$ . Figure 1 shows the results for the various graph Fourier modes depicted. These numbers confirm the qualitative remarks we made above. In other words, if working on Gauss CMP can be afforded in terms of complexity and a good value of  $\sigma$  is found, then it is a good choice, otherwise, the graphs Learnt KNN and Learnt THR are also good choices.

## 5.2 Spectral Correlations

The goal being that the sampled continuous GRF signal is graph WSS, we look at the correlation matrix  $R_{\mathbf{g}}$  of the sampled graph signal  $\mathbf{g}$  of the GRF signal  $\tilde{\mathbf{g}}$ , and the more informative matrix of Pearson correlations:

$$C_{\mathbf{g}}(k, l) = \frac{S_{\mathbf{g}}(k, l)}{\sqrt{S_{\mathbf{g}}(k, k)}\sqrt{S_{\mathbf{g}}(l, l)}}.$$

This matrix removes the influence of the variance on each mode and shows how much correlation there is between any pair of graph frequencies. Figure 2 shows this matrix for the sampled continuous GRF graph signal  $\mathbf{g}$  using different graphs. From this figure, we observe that (i) for all graphs a large amount of correlation is given by the power spectrum (there less is off-diagonal energy than for the direct domain correlation matrix), and (ii) Gauss CMP has more correlations between frequencies than the other graphs.

With respect to spectral correlations, the graph Gauss CMP should be avoided, and the learnt graphs give good results, especially the sparse ones.

## 5.3 Graph Power Spectrum

In this section, we look at how the graph power spectrum varies with the graph frequency, compared to the underlying continuous PSD of the GRF model  $\tilde{\mathbf{g}}$ . Recall that this PSD is a  $1/\nu^2$  model, such that we are looking for a  $1/\lambda$  graph PSD.

Figure 3 plots the graph PSD against the graph frequency in a loglog graph. Several observations can be made out of this figure. First, all graphs show a colored noise model for the PSD:  $\gamma(\lambda) = a\lambda^{-b}$ . However, the sparsified Gaussian graphs Gauss KNN and Gauss THR are closer to  $b = 0.5$  rather than the target  $b = 1$  of a  $1/\lambda$  model. The graph Gauss CMP is also closer to  $b = 2$ , however, we observe that most of the energy is contained within the low frequencies, and the target  $b = 1$  is not far off factoring in the noise.

On the other hand, the Learnt graphs follow more closely the target of  $b = 1$ , with less noise on the PSD. However, the graph Learnt CMP has a different behavior in the lower spectrum, with  $b$  closer to 6. This shows that the overfitting behavior shown by the GFT is at play here, and lower graph frequency modes are actually polluted by higher continuous frequencies. Consequently energy of the higher continuous frequencies is shifted to the lower spectrum, and a higher value of  $b$  is obtained.

In terms of graph selection, the learnt sparsified graphs work best for the graph power spectrum of the graph signal  $\mathbf{g}$  to match that of the underlying continuous GRF model.

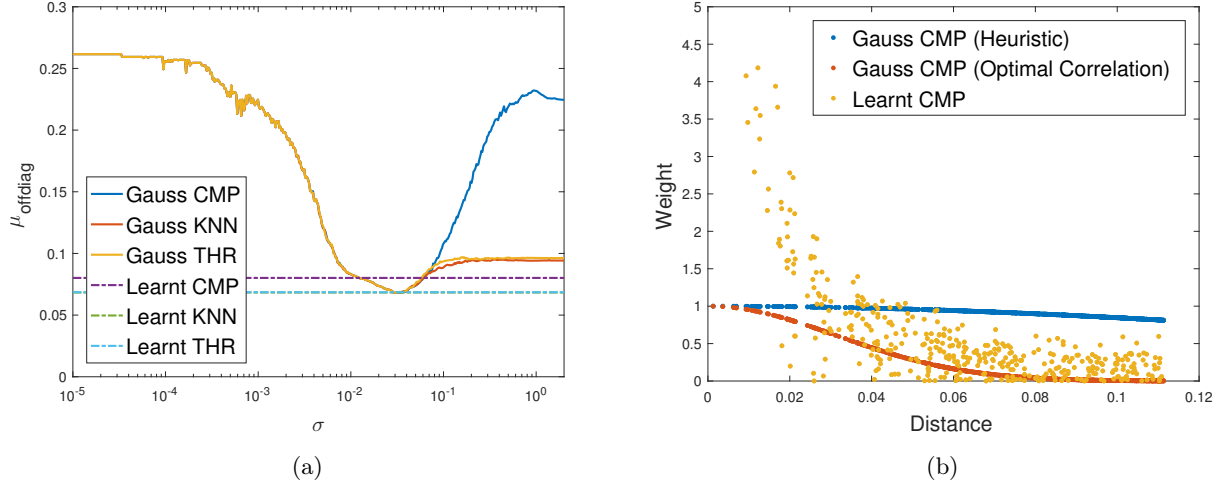


Figure 4. (a) Influence on the mean spectral Pearson correlation of the parameter  $\sigma$  in the Gaussian kernel (the curves for Learnt KNN and Learnt THR almost coincide). (b) Weight against the Euclidean distance for Gauss THR with  $\sigma$  set by the heuristic (1/3 of the mean distance) or selected to minimize spectral correlation on (a), and for the graph Learnt THR. The vertical axis is truncated and the points corresponding to very close vertices have very high weight.

#### 5.4 Influence of $\sigma$

Before studying the local power spectrum of the graph signal, we focus on one crucial parameter of the Gaussian weights approach: the parameter  $\sigma$ . On Fig. 4(a), we study its influence on spectral correlation (*i.e.*, the energy of the off-diagonal elements in the matrices of Fig. 2). The figure shows one important point: There is an optimal value to minimize spectral correlations, around  $\sigma = 0.032$  (versus  $\sigma_{\text{CMP}} = 0.173$ ,  $\sigma_{\text{KNN}} = 0.0255$  and  $\sigma_{\text{THR}} = 0.0242$  when using the heuristic of one third of the mean distance used on Fig. 2).

There are however arguments against using such a method to set the value of  $\sigma$ : The GFT for this small value of  $\sigma$  is very close to that of Gauss KNN and Gauss THR on Fig. 1. More precisely, decreasing the value of  $\sigma$  tends to disconnect clusters as many more edges will have small weights, and as such the low frequency Fourier modes are forced to be almost constant on those clusters. This is hinted by Fig. 4(b) where the weights sharply decay with the distance when using the optimal  $\sigma$ , such that some edges in Gauss THR are almost removed, those between clusters. As a consequence, the graph Fourier modes are not easily interpretable in the underlying Euclidean space, from which the signal  $\tilde{\mathbf{g}}$  is sampled.

In other words, Fig. 4 shows that it is possible to minimize the mean spectral correlation (and on this example it is equal to mean spectral correlation of Learnt KNN and Learnt THR), but doing so is at the cost of a meaningful GFT (not shown here, but included in the matlab code associated to this paper). The learnt sparse graphs on the other hand show the best of both worlds: meaningful graph Fourier modes, and small

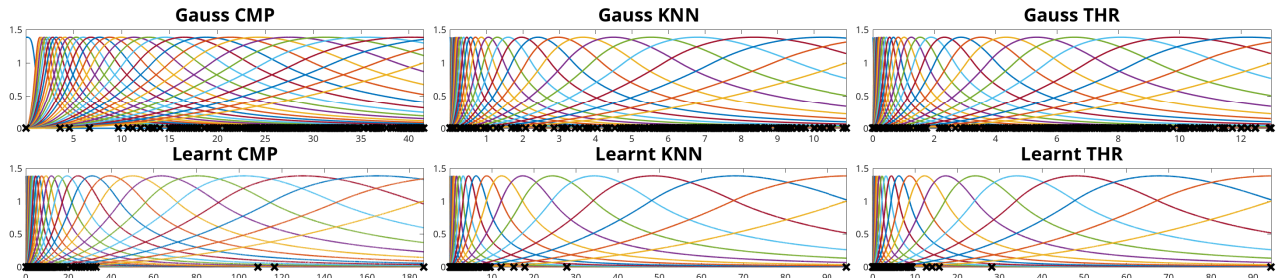
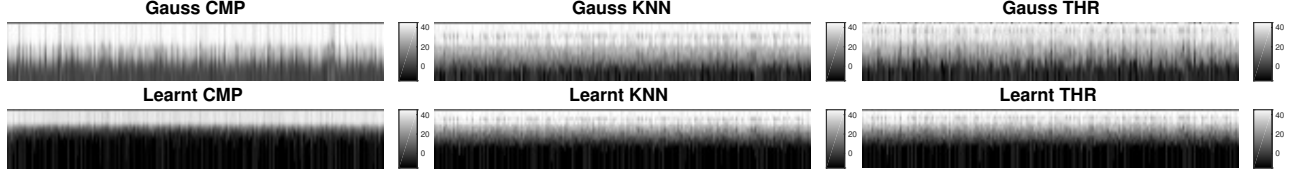
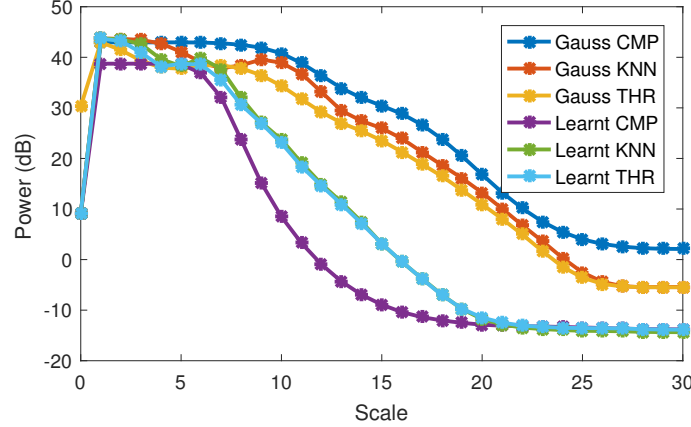


Figure 5. Frequency responses of the SGWT kernels  $g_m$  of various graphs. Bold crosses represent the eigenvalues of the graph.



(a)



(b)

Figure 6. (a) Local power spectrum of the signal  $\mathbf{g}$  in dB with vertices on the horizontal axis, and scales vertically (scaling function and coarse scales on top), and (b) its mean over vertices against the scale.

spectral correlations. Learning the graph allows here to have weights that depend on both the distance and the neighborhoods to overcome the impact of the irregular Euclidean space sampling. This is illustrated on Fig. 4(b) with a non-linear relation between weight and distance.

### 5.5 Graph Local Power Spectrum

To compute the local power spectrum, we use the spectral graph wavelet transform described in Sec. 2. The exact parameters vary from one graph to another to account for the difference of eigenvalue repartitions (see Fig. 5), and can be found in the source code.

Notably, the learnt graphs have several very high eigenvalues. These correspond to those vertices that are extremely close to each other, making them highly correlated, hence being given a high weight to the edge connecting them. Although these high weights could be attenuated using regularization in the learning algorithm, they do represent an important feature of the sampling: If a given signal have a high variation between two such vertices, then there is also a high variation in the underlying continuous signal, hence a high frequency. In other words, these eigenvalues do correspond to very high frequencies in the continuous domain.

However, for the SGWT that relies on the largest eigenvalue to define the wavelets filters, that means that the spectral responses of the filters are skewed towards higher graph frequencies where there are actually few of them. To circumvent this we parametrized differently the span of the wavelet functions in the SGWT from one graph to another. We will further investigate this point in a future communication.

In terms of local stationarity, we observe on Fig. 6(a) that not all vertices have the same local power spectrum, for any graph. However, there is a link between the mean (over vertices) of the local power spectrum and the global spectrum as shown in Fig. 6(b) and studied before.<sup>13</sup> We can therefore study the global power spectrum from the local power spectrum, but deciding global stationarity from it is not possible, even with the exact local PSD matrix.

Using the SGWT, as shown previously,<sup>13</sup> we can infer the graph PSD. However, local and global stationarity are not equivalent definitions, and one does not imply the other, as shown here. One important remark though:



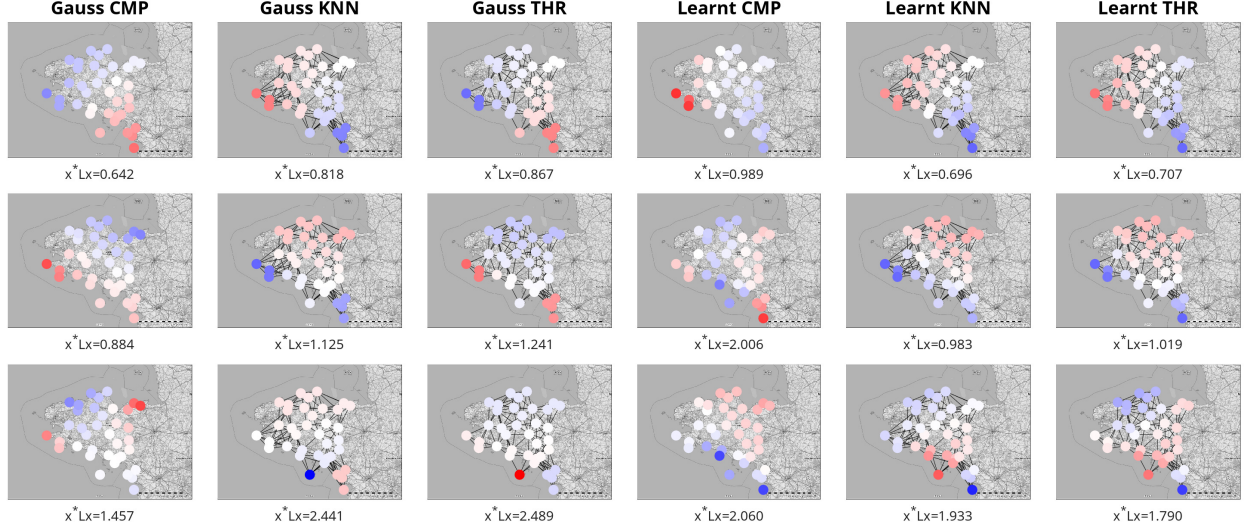


Figure 7. Comparison of the graph Fourier modes #2, #3, and #4 (from top to bottom) for various graphs with the Molène dataset. Edges are omitted for the complete graphs for clarity. Values range from  $-0.7$  (blue) to  $0.7$  (red) including  $0$  (white). One value is above  $0.7$  on the last row for the Learnt KNN graph. Laplacian variations are computed with respect to Gauss CMP graph Laplacian.

The graph used is here important as the SGWT depends on the GFT and the graph frequencies. We will further investigate the relationship between the SGWT and the graph in a future communication to better understand local stationarity.

## 6. GRAPHS FOR REAL SIGNALS

We now look at experimental validation of this data-driven approach of selecting a graph. In a recent communication,<sup>7</sup> we used a sensor network dataset to study stationarity and used the graph Gauss CMP to study stationarity of the sensor readings. We propose to refine the study of this dataset in two ways: (i) we perform a different, less involved, preprocessing, and (ii) we look at graph stationarity with respect to different graph constructions to select one that best describe the statistics of the data.

More precisely, this dataset is called the *Molène dataset*,<sup>22</sup> from the name of an island off the coast of Brittany, France. As the name suggests, the sensor network is spread in the region of Brittany, and we study the temperature readings. Those readings are performed hourly, during the month of January 2014. This makes for 744 readings. Each of the 37 weather stations is associated with a geographical position such that construction of the Gaussian kernel is possible to build the Gaussian graphs.  $K$  for Gauss KNN is chosen equal to 6, and the radius for Gauss THR is chosen to obtain the same mean number of neighbors in both sparsified graphs.

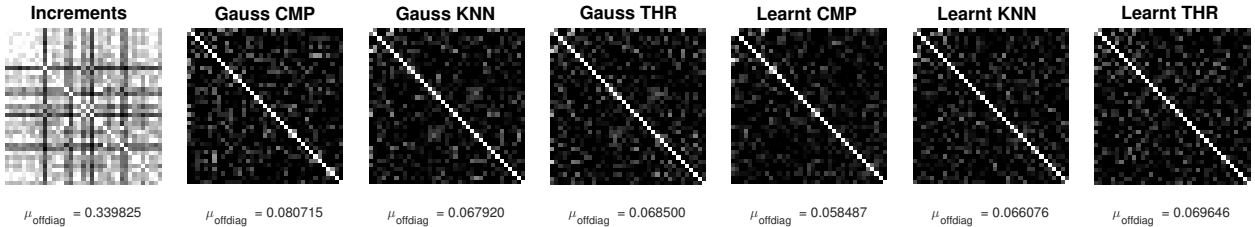


Figure 8. Pearson correlation matrices of the temperature increments. Values range from  $0$  (black) to  $1.0$  (white). For reference, the spatial Pearson correlation matrix is given (left), and the mean Pearson correlation (off-diagonal elements) is given below the matrices.

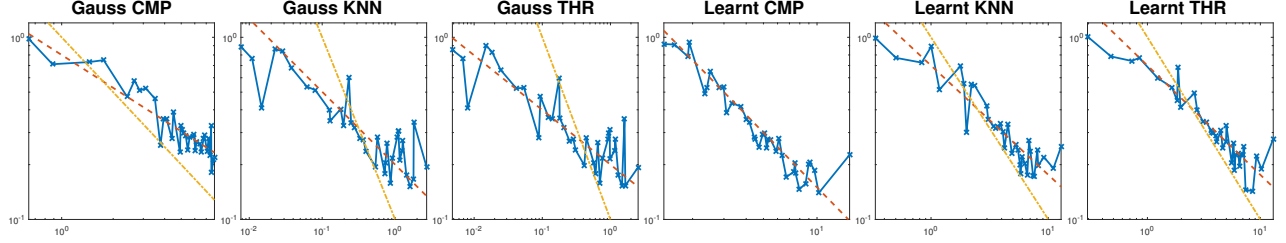


Figure 9. Graph PSD for temperature increments using different graphs. The red dashed lines corresponds to a function  $\gamma(\lambda) = a\lambda^{-b}$ .  $(a, b)$  parameters are from left to right:  $(e^{0.8}, 0.6)$ ,  $(e^{0.2}, 0.4)$ ,  $(e^{0.2}, 0.3)$ ,  $(e^{1.5}, 1)$ ,  $(e^{0.7}, 0.6)$ ,  $(e^{0.7}, 0.6)$ . The yellow dash-dotted line corresponds to the case  $b = 1.0$ . Note that the lowest frequency 0 is not shown here because of the logarithmic scale on the horizontal axis.

Using the sensor readings as many realizations of the same stochastic graph signal, we can compute sample estimators of the first two moments. The goal is to ultimately give a stochastic model for this dataset.

## 6.1 Data Preprocessing

As shown before,<sup>7</sup> preprocessing of the data is necessary because of obvious non-ergodicity shown by the data. Examples of those include cyclo-stationarities (day-night cycles), and trend over the month of the temperatures. We previously used the empirical mode decomposition as preprocessing step, a method empirically known to remove these issues for a more sensible ergodicity assumption. However, this method is particularly involved, and as its name suggests, it is empirical. In,<sup>17</sup> the authors used a simple mean removal from the data, but this method is not ideal since it has to be done twice making for  $37 + 744$  mean estimations: one for each time instant to remove the noise, and one per vertex to cancel out the first moment.

We propose here to model this data using a different assumption: temperatures follow a Brownian motion model. In other words, we assume that temperature increments are independent Gaussian variables. This model has the advantages of not performing any estimation (as opposed to mean removal), of an easy intuition (as opposed to an EMD), and is reasonable in a graph signal processing setting: Correlations between distant vertices are much smaller with temperature increments. One explanation for this phenomenon comes from the study of those stations highly correlated yet very far apart: They are on the coast line such that the sea and its rather constant temperature is attenuating the temperature variation due to the air<sup>‡</sup>.

Further analysis reveals a first moment close to 0 on all vertices (maximum amplitude of the first moment is  $2.5 \times 10^{-3}$ , with temperature increments in  $[-8^\circ\text{C}, 8^\circ\text{C}]$ ). The first moment condition of stationarity is then verified. In the rest of this section we study *spatial correlations of temperature increments*.

## 6.2 Learnt Graph

We use the same data-driven approach to build the learnt graphs. However, instead of having to resort to synthetic graph signals, we here use the sensor readings: the temperature increments. We obtain a covariance matrix fed to the learning algorithm.<sup>10</sup> Figure 7 shows then some low frequency graph Fourier modes of the Gaussian and learnt graphs, along with the Laplacian variations with respect to Gauss CMP of those modes. The same conclusion on Gauss KNN and Gauss THR can be drawn here: There is a noticeable clustering effect. Learnt CMP is again less smooth than one would expect, and the learnt sparse graphs achieve good smoothness, making them the best choice to have good Fourier modes and sparsity.

## 6.3 Graph Stationarity and Graph PSD

Figure 8 shows the spectral Pearson correlation matrices for the different graphs. Results are again very good for the learnt sparse graphs, especially for Learnt THR that shows small spectral correlations in average, but also the smaller extreme correlations. Noteworthy, all graphs are able to substantially decrease the dimension from a dense correlation matrix in the vertex domain, to a highly sparse spectral correlation matrix.

<sup>‡</sup>On this dataset one could very well work with the original temperatures and learn a graph, but the sea would need to be taken into account.

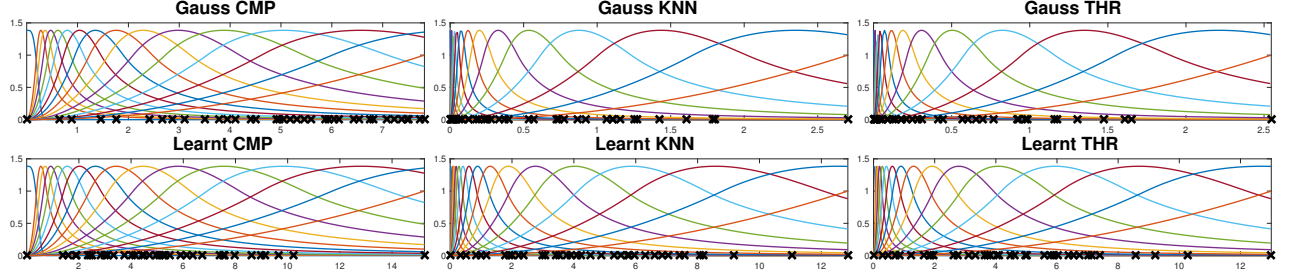


Figure 10. Frequency responses of the SGWT kernels  $g_m$  of various graphs for the Molène dataset. Bold crosses represent the eigenvalues of the graph.

Finally, Fig. 9 shows the associated graph PSD for the different graphs. First of all, except for the graph Learnt CMP, no graph PSD follows the model in  $1/\lambda$ . However, these are data real for which we do not expect the graph PSD to follow this model. Note also that overfitting is at play on Learnt CMP as the graph learning approach will tend to find the graph that best matches this graph signal to a model in  $1/\lambda$ . The second remark is on the fourth graph Fourier mode for the graphs Gauss KNN and Gauss THR (third data point on Fig. 9): This component is very small. Looking at Fig. 7 we see that these modes have high energy on a vertex corresponding to an island station. The temperature of this station should however not be very different than those of the close coastline, such that this mode is absent in the temperature increment graph signal, leading to a small energy. Here the clustering effect goes against the real data, and the graph PSD is not representative of the real data. Finally, the learnt sparse graph achieve good smoothness of the graph PSD, one for which graph filtering using low order polynomial is reasonable.<sup>1</sup>

#### 6.4 Local Power Spectrum

To conclude this experimental section, we look at the local power spectrum on Fig. 11. The associated frequency spectrum of the SGWT is shown on Fig. 10. First of all, for no graph, the signal is local stationary as the local power spectrum varies from one vertex to another. As pointed out in a previous communication,<sup>13</sup> local stationarity is a property that is very hard to verify due to the high number of parameters to the model (number of scales times the number of vertices) compared to the number of samples (vertices). Deciding local stationarity could however be done statistically, and we will further investigate this crucial question in a future communication.

On this other hand, the mean local power spectrum on Fig. 11(b) is smooth for all graphs and can be used to estimate the global power spectrum. Notice that the scaling function coefficient (first data point) is associated to a very high energy. This is due to the fact that the data has some DC component, although it is zero in average. By construction of the SGWT atoms, only the scaling function is affected by this energy.

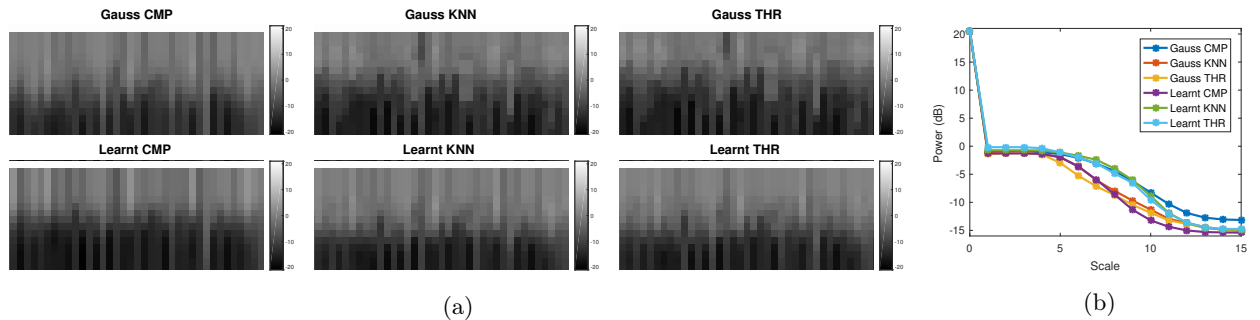


Figure 11. (a) Local power spectrum of the signal  $\mathbf{g}$  in dB (horizontally: vertices, vertically: scales, coarser on top), and (b) its mean over vertices in function of the scale.



## 7. CONCLUSION

For sampled Euclidean spaces, we showed that the usual choice of graph weights deriving from a Gaussian kernel of the Euclidean distance is not ideal, especially with respect to the graph Fourier transform. This is true for both a complete and sparsified graphs. We showed then that learning the weights allows for more flexibility as those weights can depend not just on the distance between vertices but also on the connectivity patterns that are generated by the Euclidean space sampling. We experimentally obtained better results using our learning approach with respect to stationarity, power spectrum, and even the graph Fourier modes smoothness in the lower spectrum.

We illustrated this approach on a real dataset involving temperature measurements. With our approach, we obtained more sensible graph Fourier modes, and a notion of stationarity on a sparse graph that make sense with respect to the data. This is further validated with a smooth graph power spectrum. Preliminary results on the local power spectrum hint at a local decomposition that is interesting to obtain an estimate of the global power spectrum using the SGWT. However, sensitivity of the SGWT parameters to the distribution of eigenvalues, leads to the question of the relation between the two. We will look further into this question, and the broader question of the localized set of atoms in a future communication.

## ACKNOWLEDGMENTS

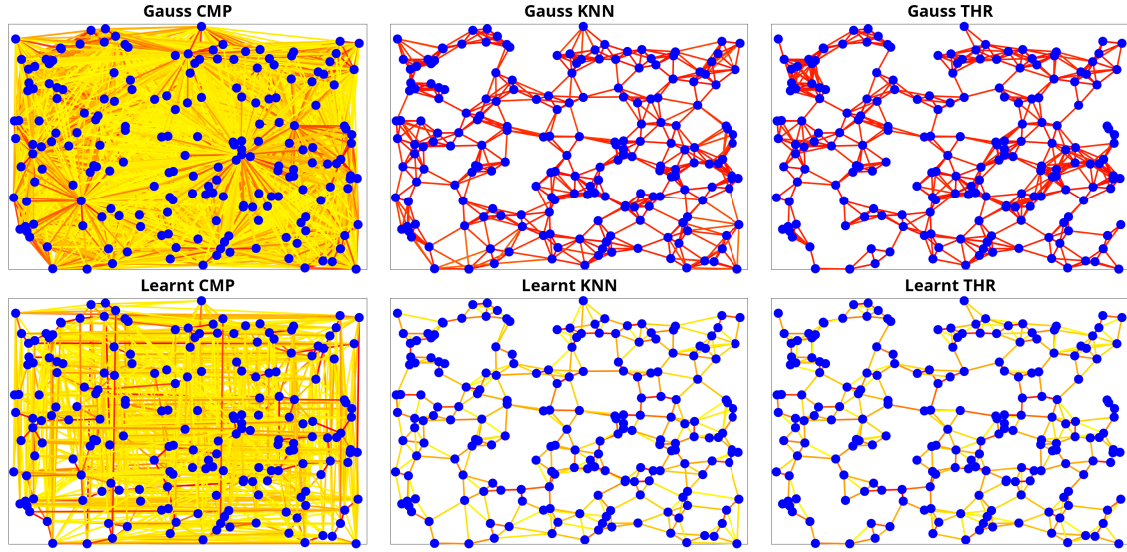
This work was supported in part by NSF under grants CCF-1410009, CCF-1527874, CCF-1029373.

## APPENDIX A. GRAPH WEIGHTS

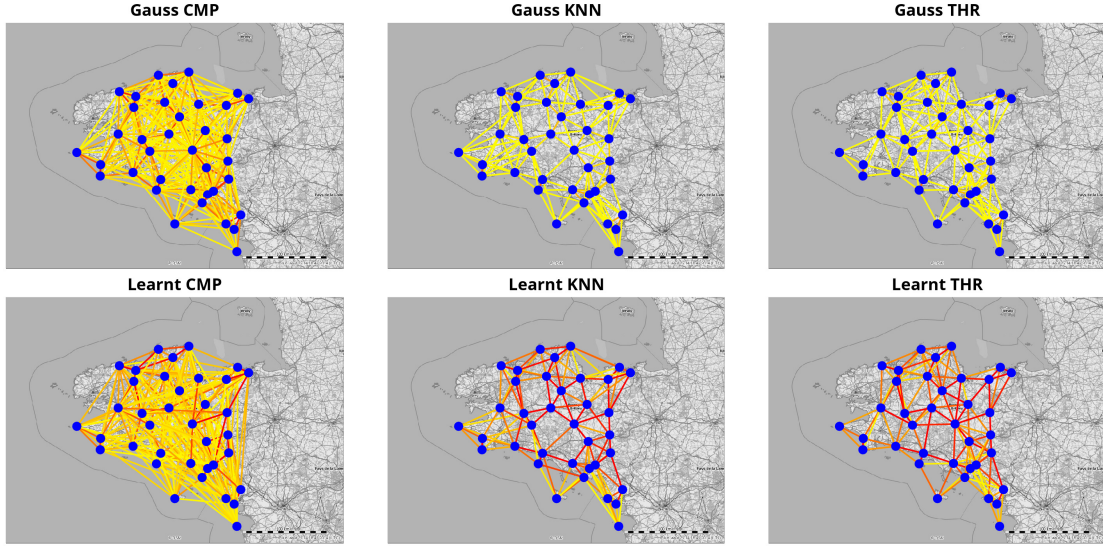
Figure 12 shows using colors the weight of the edges for all the graphs studied in this paper. As discussed in the core of the communication, weights on the sparsified Gaussian graphs are very much similar to one another, making those graphs almost unweighted. The consequence is that the graph Fourier transform depends only on the connectivity, and not on the closeness of those vertices. On the other hand, the learnt graph are able to counterweight the influence of the connectivity to truely fit to the data model at hand.

## REFERENCES

- [1] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P., “The Emerging Field of Signal Processing on Graphs: Extending High-Dimensional Data Analysis to Networks and Other Irregular Domains.,” *IEEE Signal Processing Magazine* **30**(3), 83–98 (2013).
- [2] Sandryhaila, A. and Moura, J. M. F., “Discrete Signal Processing on Graphs,” *IEEE Transactions on Signal Processing* **61**(7), 1644–1656 (2013).
- [3] Hammond, D. K., Vandergheynst, P., and Gribonval, R., “Wavelets on graphs via spectral graph theory,” *Applied and Computational Harmonic Analysis* **30**(2), 129–150 (2011).
- [4] Narang, S. K. and Ortega, A., “Compact Support Biorthogonal Wavelet Filterbanks for Arbitrary Undirected Graphs,” *IEEE Transactions on Signal Processing* **61**(19), 4673–4685 (2013).
- [5] Tremblay, N. and Borgnat, P., “Graph Wavelets for Multiscale Community Mining,” *Signal Processing, IEEE Transactions on* **62**(20), 5227–5239 (2014).
- [6] Agaskar, A. and Lu, Y. M., “A Spectral Graph Uncertainty Principle,” *IEEE Transactions on Information Theory* **59**(7), 4338–4356 (2013).
- [7] Girault, B., “Stationary Graph Signals using an Isometric Graph Translation,” in [*Signal Processing Conference (EUSIPCO), 2015 Proceedings of the 23rd European*], IEEE (2015).
- [8] Girault, B., *Signal Processing on Graphs - Contributions to an Emerging Field*, phd thesis, Ecole normale supérieure de lyon - ENS LYON (Dec. 2015).
- [9] Belkin, M. and Niyogi, P., “Towards a theoretical foundation for Laplacian-based manifold methods,” *Journal of Computer and System Sciences* **74**(8), 1289–1308 (2008).
- [10] Egilmez, H. E., Pavez, E., and Ortega, A., “Graph Learning from Data under Laplacian and Structural Constraints,” (2017).



(a)



(b)

Figure 12. Graph weights represented by colors from yellow (0) to red (1) for the graphs of (a) Sec. 4, and (b) Sec. 6. Edges for the CMP graphs are not shown: when the weight is smaller than  $1/64$  (64 is the number of colors of a colormap in Matlab) corresponding to color associated to 0. For the learnt graphs, several edges have weight greater than 1.

- [11] Segarra, S., Marques, A. G., Mateos, G., and Ribeiro, A., “Network Topology Inference from Spectral Templates,” *IEEE Transactions on Signal and Information Processing over Networks* **PP**(99), 1–1 (2017).
- [12] Padeloup, B., Gripon, V., Mercier, G., Pastor, D., and Rabbat, M. G., “Characterization and Inference of Graph Diffusion Processes from Observations of Stationary Signals,” (June 2017).
- [13] Girault, B., Narayanan, S. S., and Ortega, A., “Towards a definition of local stationarity for graph signals,” in *[2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)]*, 4139–4143, IEEE (2017).
- [14] Girault, B., Gonçalves, P., and Fleury, É., “Translation on Graphs: An Isometric Shift Operator,” *Signal Processing Letters, IEEE* **22**, 2416–2420 (Dec 2015).

- [15] Cressie, N. A. C., [*Statistics for spatial data*], Wiley-Interscience Publication (1991).
- [16] Girault, B., Gonçalves, P., Narayanan, S. S., and Ortega, A., “Localization Bounds for the Graph Translation,” in [*2016 IEEE Global Conference on Signal and Information Processing*], (Dec 2016).
- [17] Perraudin, N. and Vandergheynst, P., “Stationary Signal Processing on Graphs,” *IEEE Transactions on Signal Processing* **65**(13), 3462–3477 (2017).
- [18] Marques, A. G., Segarra, S., Leus, G., and Ribeiro, A., “Stationary Graph Processes and Spectral Estimation,” (Mar. 2016).
- [19] Chapelle, O., Schölkopf, B., and Zien, A., [*Semi-Supervised Learning*], MIT Press (2006).
- [20] Borgnat, P., Flandrin, P., Honeine, P., Richard, C., and Xiao, J., “Testing stationarity with surrogates: a time-frequency approach,” *IEEE Trans. Signal Processing* **58**(7), 3459–3470 (2010).
- [21] Papoulis, A., [*Probability, random variables, and stochastic processes*], McGraw-Hill, 3rd ed. (1991).
- [22] Météo France, “Données horaires des 55 stations terrestres de la zone Large Molène sur un mois,” (Jan. 2014). Source: <https://www.data.gouv.fr/fr/datasets/donnees-horaires-des-55-stations-terrestres-de-la-zone-large-molene-sur-un-mois/>.