# Assignment 3 (ML for TS) - MVA 2021/2022

Eya Ghamgui eya.ghamgui@telecom-paris.fr
Siwar Mhadhbi siwar.mhadhbi@telecom-paris.fr

June 28, 2022

## 1 Introduction

**Objective.** The goal is to present (i) a model selection heuristics to find the number of change-points in a signal and (ii) wavelets for graph signals.

**Warning and advice.**

- Use code from the tutorials as well as from other sources. Do not code yourself well-known procedures (e.g. cross validation or k-means), use an existing implementation.
- The associated notebook contains some hints and several helper functions.
- Be concise. Answers are not expected to be longer than a few sentences (omitting calculations).

**Instructions.**

- Fill in your names and emails at the top of the document.
- Hand in your report (one per pair of students) by Friday 18$^{\text{th}}$ March 11:59 PM.
- Rename your report and notebook as follows:
  `FirstnameLastname1_FirstnameLastname1.pdf` and
  `FirstnameLastname2_FirstnameLastname2.ipynb`.
  For instance, `LaurentOudre_CharlesTruong.pdf`.
- Upload your report (PDF file) and notebook (IPYNB file) using this link: dropbox.com/request/5DKPDBVAJ25hon0ZZsnn.

## 2 Model selection for change-point detection

**Notations.** In the following, $\|x\|$ is the Euclidean norm of $x$ if $x$ is a vector and the Frobenius norm if $x$ is a matrix. A set of change-points is denoted by a bold $\boldsymbol{\tau} = \{t_1, t_2, \dots\}$ and $|\boldsymbol{\tau}|$ (the cardinal of $\boldsymbol{\tau}$) is the number of change-points. By convention $t_0 = 0$ and $t_{|\boldsymbol{\tau}|+1} = T$. For a set of change-points $\boldsymbol{\tau}$, $\Pi_{\boldsymbol{\tau}}$ is the orthogonal projection onto the linear subspace of piecewise constant signals with change-points in $\boldsymbol{\tau}$: for a signal $x = \{x_t\}_{t=0}^{T-1}$,

$$(\Pi_{\boldsymbol{\tau}} x)_t = \bar{x}_{t_k..t_{k+1}} \quad \text{if } t_k \le t < t_{k+1} \tag{1}$$

where $\bar{x}_{t_k..t_{k+1}}$ is the empirical mean of the subsignal $x_{t_k..t_{k+1}} = \{x_t\}_{t_k}^{t_{k+1}-1}$.

**Model selection.** Assume we observe a $\mathbb{R}^d$-valued signal $y = \{y_t\}_{t=0}^{T-1}$ with $T$ samples that follows the model

$$y_t = f_t + \varepsilon_t \tag{2}$$

where $f$ is a deterministic signal which we want to estimate with piecewise constant signals and $\varepsilon_t$ is i.i.d. with mean 0 and covariance $\sigma^2 I_d$.

The ideal choice of $\boldsymbol{\tau}$ minimizes the distance from the true (noiseless) signal $f$:

$$\boldsymbol{\tau}^\star = \arg\min_{\boldsymbol{\tau}} \frac{1}{T} \|f - \Pi_{\boldsymbol{\tau}} y\|^2 . \tag{3}$$

The estimator $\boldsymbol{\tau}^\star$ is an *oracle* estimator because it relies on the unknown signal $f$. Several model selection procedures rely on the "unbiased risk estimation heuristics": if $\hat{\boldsymbol{\tau}}$ minimizes a criterion $\text{crit}(\boldsymbol{\tau})$ such that

$$\mathbb{E}\left[\text{crit}(\boldsymbol{\tau})\right] \approx \mathbb{E}\left[\frac{1}{T}\|f - \Pi_{\boldsymbol{\tau}} y\|^2\right] \tag{4}$$

then

$$\frac{1}{T}\|f - \Pi_{\hat{\boldsymbol{\tau}}} y\|^2 \approx \min_{\boldsymbol{\tau}} \frac{1}{T}\|f - \Pi_{\boldsymbol{\tau}} y\|^2 \tag{5}$$

under some conditions. In other words, the estimator $\hat{\boldsymbol{\tau}}$ approximately minimizes the oracle quadratic risk.

Here, we will consider penalized criteria:

$$\text{crit}(\boldsymbol{\tau}) = \frac{1}{T}\|y - \Pi_{\boldsymbol{\tau}} y\|^2 + \text{pen}(\boldsymbol{\tau}) \tag{6}$$

where pen is a penalty function. In addition, let

$$\hat{\boldsymbol{\tau}}_{\text{pen}} := \arg\min_{\boldsymbol{\tau}} \left[\frac{1}{T}\|y - \Pi_{\boldsymbol{\tau}} y\|^2 + \text{pen}(\boldsymbol{\tau})\right] . \tag{7}$$

**Question 1** *Ideal penalty*

- Calculate $\mathbb{E}[\|\varepsilon\|^2/T]$, $\mathbb{E}[\|f-\Pi_{\boldsymbol{\tau}} y\|^2/T]$ and $\mathbb{E}[\|y-\Pi_{\boldsymbol{\tau}} y\|^2/T]$.
- What would be an ideal penalty $\mathrm{pen}_{\mathrm{id}}$ such that Equation (4) is verified?

## Answer 1

**First Part:**

We consider in this part the norm Frobenius defined as :

$$A \in \mathbb{R}^{n\times m}, \quad \|A\|_F^2 = tr(A^TA) = \sum_{i=1}^{n}\sum_{j=1}^{m}|a_{i,j}|^2$$

$$\textbf{(1)} \quad \mathbf{E}\left[\frac{\|\epsilon\|_F^2}{T}\right] = \frac{1}{T}\mathbf{E}\left[Tr(\epsilon^T\epsilon)\right]$$
$$= \frac{1}{T}Tr\left(\mathbf{E}[\epsilon^T\epsilon]\right) \quad \text{(as trace is a linear operator)}$$

Or we have: $\mathbf{E}\left[(\epsilon-\mu)^T(\epsilon-\mu)\right] = \mathbf{E}\left[\epsilon^T\epsilon\right] = Cov(\epsilon) = \sigma^2 I_d$

Thus,

$$\boxed{\mathbf{E}\left[\frac{\|\epsilon\|_F^2}{T}\right] = \frac{1}{T}\,d\,\sigma^2}$$

$$\textbf{(2)} \quad \mathbf{E}\left[\frac{\|f-\Pi_\tau y\|_F^2}{T}\right] = \frac{1}{T}\mathbf{E}[\textstyle\sum_{i=0}^{T-1}\sum_{j=1}^{d}|f-\Pi_\tau f-\Pi_\tau\epsilon|_{i,j}^2]$$

$$= \frac{1}{T}\mathbf{E}[\textstyle\sum_{i=0}^{T-1}\sum_{j=1}^{d}(f-\Pi_\tau f)_{i,j}^2 + (\Pi_\tau\epsilon)_{i,j}^2 - 2((\Pi_\tau\epsilon)_{i,j}(f-\Pi_\tau f)_{i,j})]$$

$$= \frac{1}{T}(\textstyle\sum_{i=0}^{T-1}\sum_{j=1}^{d}\mathbf{E}[(f-\Pi_\tau f)_{i,j}^2] + \mathbf{E}[(\Pi_\tau\epsilon)_{i,j}^2] - 2\mathbf{E}[((\Pi_\tau\epsilon)_{i,j}\underbrace{(f-\Pi_\tau f)_{i,j}}_{\text{deterministic}})])$$

$$= \frac{1}{T}(\textstyle\sum_{i=0}^{T-1}\sum_{j=1}^{d}\mathbf{E}[(f-\Pi_\tau f)_{i,j}^2] + \mathbf{E}[(\Pi_\tau\epsilon)_{i,j}^2] - 2(\underbrace{\mathbf{E}[(\Pi_\tau\epsilon)_{i,j}]}_{0}(f-\Pi_\tau f)_{i,j}))$$

$$= \frac{1}{T}(\mathbf{E}[\|f-\Pi_\tau f\|_F^2] + \mathbf{E}[\|\Pi_\tau\epsilon\|_F^2])$$

*On the other hand,*

$$\Pi_\tau \epsilon = \begin{pmatrix} (\Pi_\tau\epsilon)_{1,0} & \cdots & (\Pi_\tau\epsilon)_{1,T-1} \\ \vdots & & \vdots \\ (\Pi_\tau\epsilon)_{d,0} & \cdots & (\Pi_\tau\epsilon)_{d,T-1} \end{pmatrix} \in \mathbb{R}^{d\times T-1}$$

$$\begin{aligned}
\mathbf{E}[\|\Pi_\tau\epsilon\|_F^2] \quad &= \mathbf{E}[Tr((\Pi_\tau\epsilon)^T(\Pi_\tau\epsilon))] \\
&= \mathbf{E}[\textstyle\sum_{i=1}^{d}\sum_{t=0}^{T-1}(\Pi_\tau\epsilon)_{i,t}^2] \\
&= \mathbf{E}[\textstyle\sum_{i=1}^{d}\sum_{k=0}^{|\tau|}\sum_{t=t_k}^{t_{k+1}}(\Pi_\tau\epsilon)_{i,t}^2] \\
&= \textstyle\sum_{i=1}^{d}\sum_{k=0}^{|\tau|}\sum_{t=t_k}^{t_{k+1}}\mathbf{E}[(\Pi_\tau\epsilon)_{i,t}^2]
\end{aligned}$$

*And we have,*

$$(\Pi_\tau\epsilon)_t \overset{t_k\leq t<t_{k+1}}{=} \bar{\epsilon}_{t_k,t_{k+1}} = \frac{\sum_{j=t_k}^{t_{k+1}}\epsilon_j}{t_{k+1}-t_k} \quad \Longrightarrow \quad (\Pi_\tau\epsilon)_{i,t} = \frac{\sum_{j=t_k}^{t_{k+1}}\epsilon_{i,j}}{t_{k+1}-t_k}$$

*Thus,*

$$\begin{aligned}
\mathbf{E}[(\Pi_\tau\epsilon)_{i,t}^2] \quad &= \mathbf{E}[(\frac{\sum_{j=t_k}^{t=t_{k+1}}\epsilon_{i,j}}{t_{k+1}-t_k})^2] \\
&= \mathbf{E}[\frac{1}{(t_{k+1}-t_k)^2}(\textstyle\sum_{j=t_k}^{t_{k+1}}\sum_{l=t_k}^{t_{k+1}}\epsilon_{i,j}\epsilon_{i,l})] \\
&= \frac{1}{(t_{k+1}-t_k)^2}\textstyle\sum_{j=t_k}^{t_{k+1}}\sum_{l=t_k}^{t_{k+1}} \quad \underbrace{\mathbf{E}[\epsilon_{i,j}\epsilon_{i,l}]}_{\sigma^2 \ iff \ j=l, \ 0 \ otherwise} \\
&= \frac{1}{t_{k+1}-t_k}\ \sigma^2
\end{aligned}$$

*Therefore,*

$$\mathrm{E}\left[\frac{\|\Pi_\tau\epsilon\|_F^2}{T}\right] = \frac{1}{T}(\textstyle\sum_{i=1}^{d}\sum_{k=0}^{|\tau|}\sum_{t=t_k}^{t_{k+1}}\frac{1}{t_{k+1}-t_k}\ \ \sigma^2) = \frac{(|\tau|+1)\ d\ \sigma^2}{T}$$

$$\Longrightarrow \quad \boxed{\mathbf{E}\left[\frac{\|f-\Pi_\tau y\|_F^2}{T}\right] = \mathbf{E}\left[\frac{\|f-\Pi_\tau f\|_F^2}{T}\right] + \frac{(|\tau|+1)\ d\ \sigma^2}{T}}$$

$$\textbf{(3)}\quad \mathbf{E}[\tfrac{\|y-\Pi_\tau y\|_F^2}{T}] = \tfrac{1}{T}\mathbf{E}[\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}|y-\Pi_\tau y|_{i,j}^2]$$

$$= \tfrac{1}{T}\mathbf{E}[\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}|f+\epsilon-\Pi_\tau y|_{i,j}^2]$$

$$= \tfrac{1}{T}\mathbf{E}[\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}(f-\Pi_\tau y)_{i,j}^2+\epsilon_{i,j}^2+2(\epsilon_{i,j}(f-\Pi_\tau y)_{i,j})]$$

$$= \tfrac{1}{T}(\mathbf{E}[\|f-\Pi_\tau y\|_F^2]+\mathbf{E}[\|\epsilon\|_F^2]+2\underbrace{\mathbf{E}[\sum_{i=1}^{d}\sum_{j=0}^{T-1}(\epsilon_{i,j}(f-\Pi_\tau y)_{i,j})]}_{(\star)})$$

$$(\star)\quad = \textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\mathbf{E}[\epsilon_{i,j}f_{i,j}-\epsilon_{i,j}(\Pi_\tau y)_{i,j}]$$

$$= \textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\mathbf{E}[\epsilon_{i,j}f_{i,j}]-\mathbf{E}[\epsilon_{i,j}(\pi_\tau f)_{i,j}]-\mathbf{E}[\epsilon_{i,j}(\pi_\tau \epsilon)_{i,j}]$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\mathbf{E}[\epsilon_{i,j}(\pi_\tau \epsilon)_{i,j}]\quad (\textit{f deterministic and } \epsilon \textit{ has mean } 0)$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\mathbf{E}[\frac{\sum_{t=t_k}^{t_{k+1}}\epsilon_{i,j}\ \epsilon_{i,t}}{t_{k+1}-t_k}]$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\frac{\sum_{t=t_k}^{t_{k+1}}\mathbf{E}[\epsilon_{i,j}\ \epsilon_{i,t}]}{t_{k+1}-t_k}$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{j=0}^{T-1}\frac{\mathbf{E}[\epsilon_{i,j}^2]}{t_{k+1}-t_k}\quad (\mathbf{E}\left[\epsilon_{i,t}\epsilon_{i,j}\right]=0 \text{ if } t\neq j)$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{k=0}^{|\tau|}\sum_{t=t_k}^{t_{k+1}}\frac{\mathbf{E}[\epsilon_{i,t}^2]}{t_{k+1}-t_k}$$

$$= -\textstyle\sum_{i=1}^{d}\sum_{k=0}^{|\tau|}\sigma^2$$

$$= -d\ (|\tau|+1)\ \sigma^2$$

$$\Longrightarrow \boxed{\mathbf{E}\left[\frac{\|y-\Pi_\tau y\|_F^2}{T}\right] = \frac{1}{T}\mathbf{E}[\|f-\Pi_\tau y\|_F^2]+\frac{1}{T}\mathbf{E}[\|\epsilon\|_F^2]-\frac{2d\ (|\tau|+1)\ \sigma^2}{T}}$$

**Second Part:**

The penalty is given by :

$\mathrm{pen}(\tau) = \frac{1}{T}\mathbf{E}[\|f-\pi_\tau y\|_F^2]-\frac{1}{T}\mathbf{E}[\|y-\pi_\tau y\|_F^2]$

$= \frac{1}{T}\mathbf{E}[\|f-\Pi_\tau y\|_F^2]-\frac{1}{T}\mathbf{E}[\|f-\Pi_\tau y\|_F^2]-\frac{1}{T}\mathbf{E}[\|\epsilon\|_F^2]+\frac{2d\ (|\tau|+1)\ \sigma^2}{T}$ (*using first part of question*)

$= -\frac{1}{T}\ d\ \sigma^2+\frac{2d\ (|\tau|+1)\ \sigma^2}{T}$

$= \frac{(2(|\tau|+1)-1)\ d\ \sigma^2}{T}$

The ideal penalty is

$$\boxed{\mathrm{pen}_{\mathrm{id}}(\boldsymbol{\tau}) = \frac{2|\tau|\ d\ \sigma^2}{T}} \tag{8}$$

**Question 2** *Mallows' $C_p$*

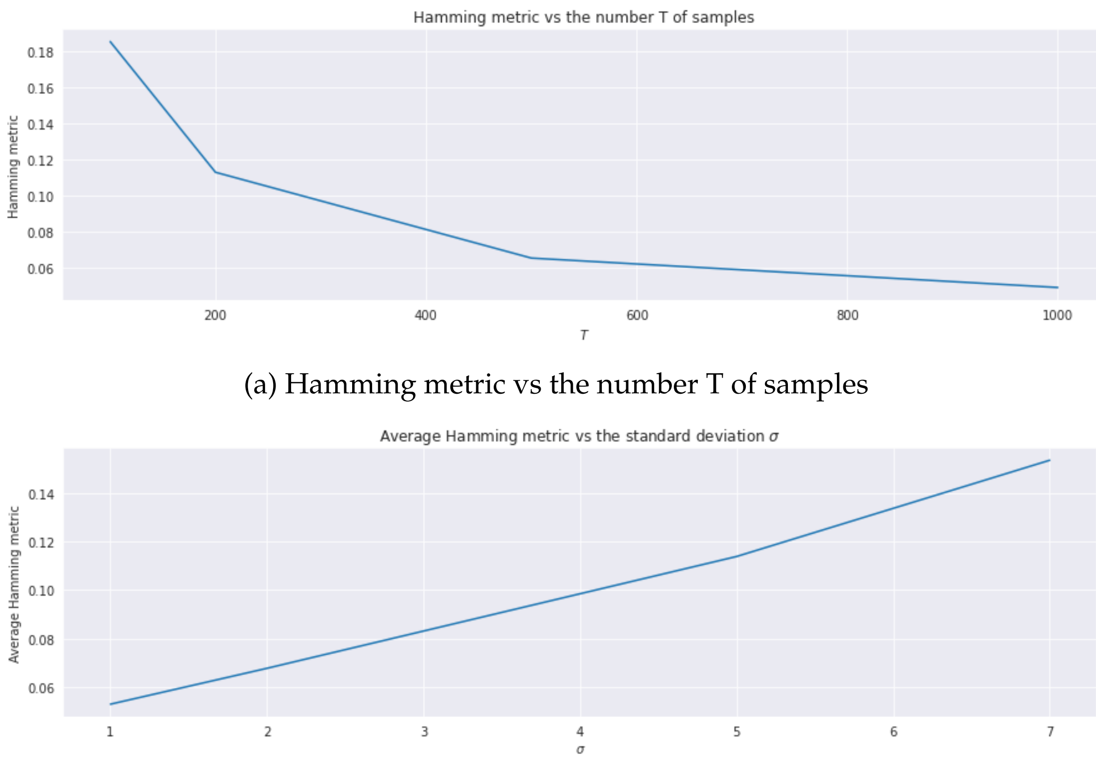
The ideal penalty depends on the unknown value of $\sigma$. Pluging an estimator $\hat{\sigma}$ into $\text{pen}_{\text{id}}$ yields the well-known Mallows' $C_p$. Use the empirical variance on the first 10% of the signal as an estimator of $\sigma^2$.

Simulate two noisy piecewise constant signals with the function `ruptures.pw_constant` (set the dimension to $d = 2$) for each combination of parameters: `n_bkps`$\in \{2, 4, 6, 8, 10\}$, $T \in \{100, 200, 500, 1000\}$ and $\sigma \in \{1, 2, 5, 7\}$.

Using Mallows' $C_p$,

- for $\sigma = 2$ and $T \in \{100, 200, 500, 1000\}$, compute the Hamming metric between the true segmentation and the estimated segmentation and report the average on Figure 1-a;
- for $T = 500$ and $\sigma \in \{1, 2, 5, 7\}$, compute the Hamming metric between the true segmentation and the estimated segmentation and report the average on Figure 1-b.

**Answer 2**




(a) Hamming metric vs the number T of samples

(b) Hamming metric vs the standard deviation $\sigma$

Figure 1: Performance of Mallows' $C_p$

$\Longrightarrow$ We can conclude that for a fixed value of $\sigma$, increasing the number of samples helps the algorithm obtain better segmentation estimate. In other words, we found that the Hamming distance decreases when $T$ increases, thus, we obtained smaller errors. In contrast, for a fixed number of samples, increasing $\sigma$ increases the Hamming distance, which means that for the same number of samples, the higher the noise, the less robust the segmentation estimate and thus the higher the errors.

**Question 3** *Slope heuristics*

The ideal penalty is of shape $\text{pen}(\boldsymbol{\tau}) = Cd|\boldsymbol{\tau}|/T$ where $C > 0$. The slope heuristics is a procedure to infer the best $C$ without knowing $\sigma$.

**Slope heuristics algorithm.**

- Estimate the slope of $\hat{s}$ of $\min_{\boldsymbol{\tau},|\boldsymbol{\tau}|=K} \|\Pi_{\boldsymbol{\tau}} - y\|^2$ as a function of $K$ for $K$ "large enough". Define $\hat{C}_{\text{slope}} := -T\hat{s}$.
- Estimate $\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}} \|y - \Pi_{\boldsymbol{\tau}} y\|^2 / T + \hat{C}_{\text{slope}} d|\boldsymbol{\tau}|/T$.

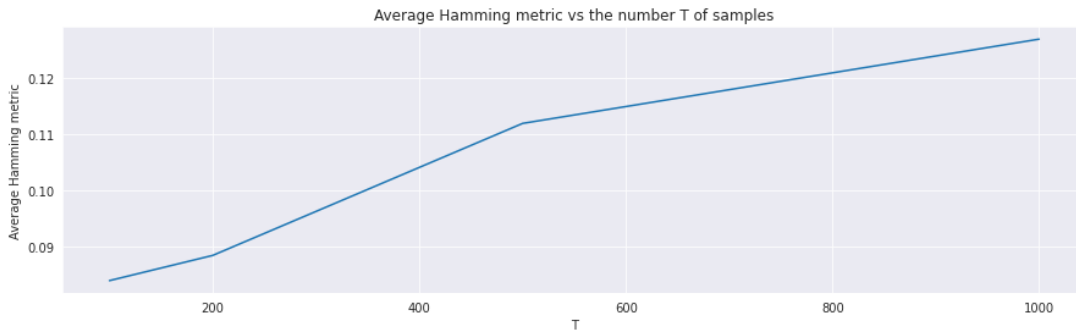
In simulations, "large enough" means for $K$ between 15 and $0.4T$.

Simulate two noisy piecewise constant signals with the function `ruptures.pw_constant` (set the dimension to $d = 2$) for each combination of parameters: `n_bkps`$\in \{2, 4, 6, 8, 10\}$, $T \in \{100, 200, 500, 1000\}$ and $\sigma \in \{1, 2, 5, 7\}$.

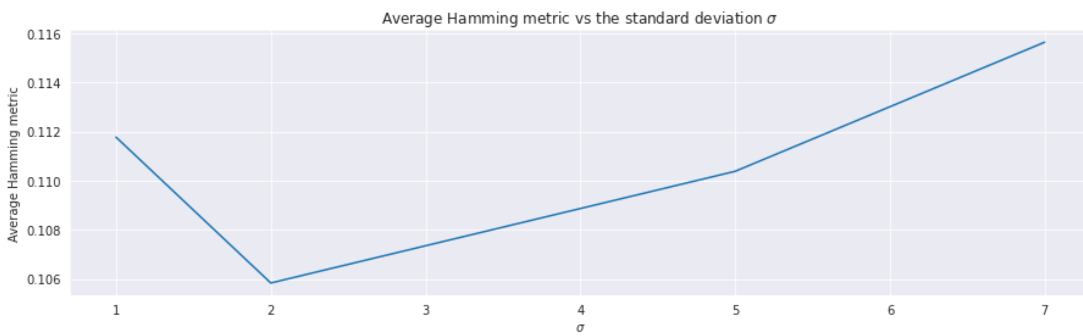
Using the slope heuristics,

- for $\sigma = 2$, $T \in \{100, 200, 500, 1000\}$, compute the average Hamming metric between the true segmentations and the estimated segmentations and report the average on Figure 2-a;
- for $T = 500$ and $\sigma \in \{1, 2, 5, 7\}$, compute the average Hamming metric between the true segmentations and the estimated segmentations and report the average on Figure 2-b.

## Answer 3




(a) Hamming metric vs the number T of samples




(b) Hamming metric vs the standard deviation $\sigma$

Figure 2: Performance of the slope heuristics

$\Longrightarrow$ We can conclude that, increasing the number of samples does not help the algorithm estimate the segmentation. In addition, the value $\sigma = 2$ is the best value that gives the lowest error for a number of samples $T = 500$. Beyond this value, the Hamming distance increases as $\sigma$ increases.

## 3 Wavelet transform for graph signals

Let $G$ be a graph defined a set of $n$ nodes $V$ and a set of edges $E$. A specific node is denoted by $v$ and a specific edge, by $e$. The eigenvalues and eigenvectors of the graph Laplacian $L$ are $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $u_1, u_2, \ldots, u_n$ respectively.

For a signal $f \in \mathbb{R}^n$, the Graph Wavelet Transform (GWT) of $f$ is $W_f : \{1, \ldots, M\} \times V \longrightarrow \mathbb{R}$:

$$W_f(m, v) := \sum_{l=1}^{n} \hat{g}_m(\lambda_l) \hat{f}_l u_l(v) \tag{9}$$

where $\hat{f} = [\hat{f}_1, \ldots, \hat{f}_n]$ is the Fourier transform of $f$ and $\hat{g}_m$ are $M$ kernel functions. The number $M$ of scales is a user-defined parameter and is set to $M := 9$ in the following. Several designs are available for the $\hat{g}_m$; here, we use the Spectrum Adapted Graph Wavelets (SAGW). Formally, each kernel $\hat{g}_m$ is such that

$$\hat{g}_m(\lambda) := \hat{g}^U(\lambda - am) \quad (0 \leq \lambda \leq \lambda_n) \tag{10}$$

where $a := \lambda_n / (M + 1 - R)$,

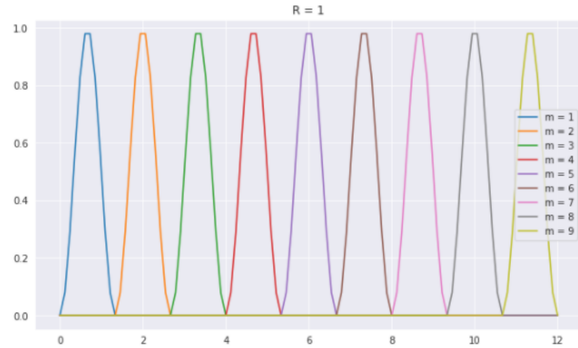
$$\hat{g}^U(\lambda) := \frac{1}{2}\left[1 + \cos\left(2\pi\left(\frac{\lambda}{aR} + \frac{1}{2}\right)\right)\right] \mathbb{1}(-Ra \leq \lambda < 0) \tag{11}$$

and $R > 0$ is defined by the user.

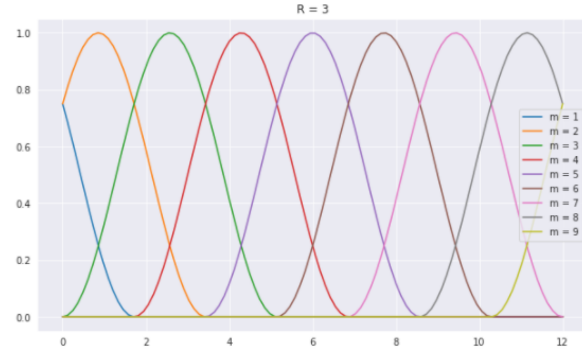
### Question 4

Plot the kernel functions $\hat{g}_m$ for $R = 1$, $R = 3$ and $R = 5$ (take $\lambda_n = 12$) on Figure 3. What is the influence of $R$?

### Answer 4




(a) $R = 1$




(b) $R = 3$


(c) $R = 5$

Figure 3: The SAGW kernels functions

**Influence of $R$:**

The parameter $R$ is the one that controls the overlap of the shifted kernels. When it increases, the overlap of the filters increases. From the previous plots, we can notice that for $R = 1$, there is no overlap between the shifted kernels. However, for higher values, we can notice that the kernels are more and more wrapped.

We will study the Molene data set (the one we used in the last tutorial). The signal is the temperature.

### Question 5

Construct the graph using the distance matrix and exponential smoothing (use the median heuristics for the bandwidth parameter).

- Remove all stations with missing values in the temperature.
- Choose the minimum threshold so that the network is connected and the average degree is at least 3.
- What is the time where the signal is the least smooth?
- What is the time where the signal is the smoothest?

### Answer 5

The stations with missing values are 18 stations which are:

BATZ, BEG-MEIL, CAMARET, PLOUGONVELIN, RIEC SUR BELON, ST NAZAIRE-MONTOIR PLOUAY-SA, VANNES-MEUCON, LANNAERO, PLOUDALMEZEAU, LANDIVISIAU, SIZUN QUIMPER, OUESSANT-STIFF, LANVEOC, ARZAL, BREST-GUIPAVAS, BRIGNOGAN

The threshold is equal to $0.83$.

The signal is the least smooth at $2014-01-21\ \ 06:00:00$

The signal is the smoothest at $2014-01-24\ \ 19:00:00$

## Question 6

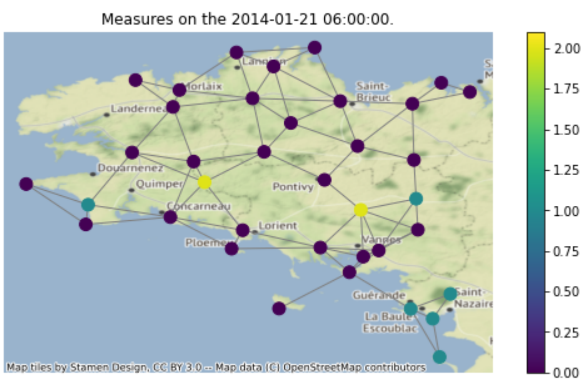
(For the remainder, set $R = 3$ for all wavelet transforms.)

For each node $v$, the vector $[W_f(1, v), W_f(2, v), \dots, W_f(M, v)]$ can be used as a vector of features. We can for instance classify nodes into low/medium/high frequency:

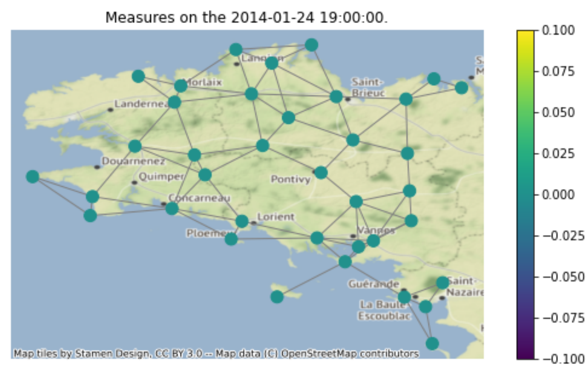
- a node is considered low frequency if the scales $m \in \{1, 2, 3\}$ contain most of the energy,
- a node is considered medium frequency if the scales $m \in \{4, 5, 6\}$ contain most of the energy,
- a node is considered high frequency if the scales $m \in \{6, 7, 9\}$ contain most of the energy.

For both signals from the previous question (smoothest and least smooth) as well as the first available timestamp, apply this procedure and display on the map the result (one colour per class).

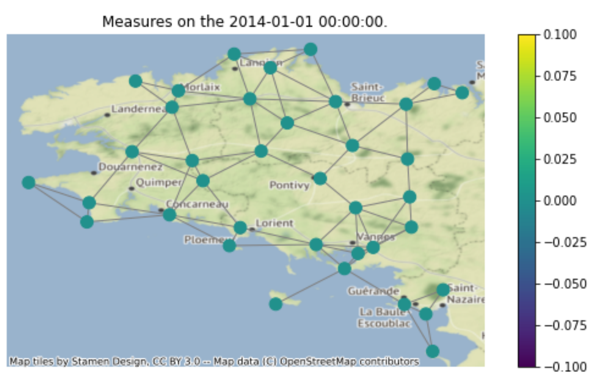
## Answer 6




(a) Least smooth signal




(b) Smoothest signal




(c) First available timestamp

Figure 4: Classification of nodes into low/medium/high frequency

**Comment:**

From the figures above, we can notice that the smoothest signal and the signal at the first time stamp have all nodes with low frequencies. In contrast, the least smooth signal has nodes with low, medium and high frequencies, which is expected as it is the signal with the highest value of smoothness.

### Question 7

Display the average temperature and for each timestamp, adapt the marker colour to the majority class present in the graph (see notebook for more details).

### Answer 7

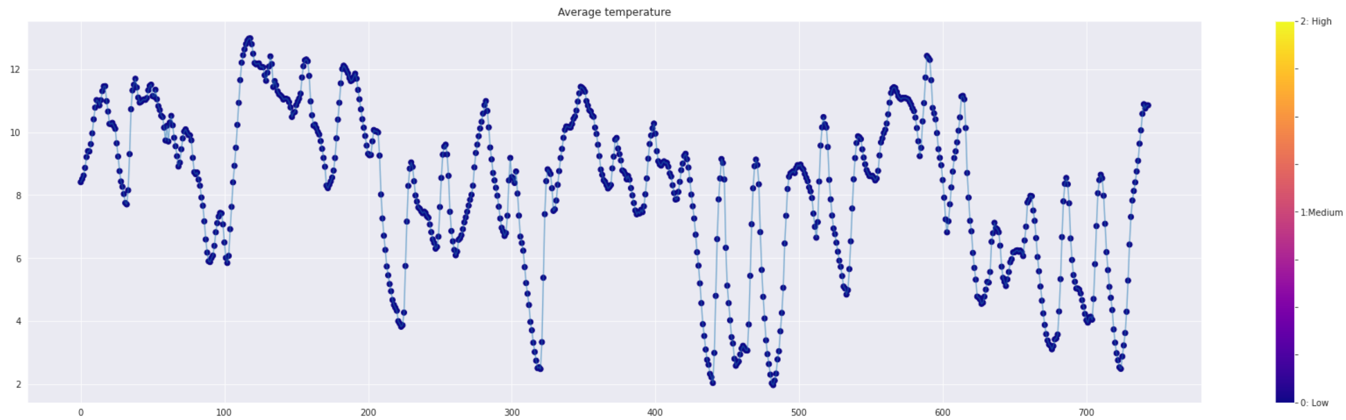



Figure 5: Average temperature. Markers' colours depend on the majority class.

**Comment:**

From the previous plot, we can notice that for all timestamps, the low frequency category is the most dominant.

### Question 8

The previous graph $G$ only uses spatial information. To take into account the temporal dynamic, we construct a larger graph $H$ as follows: a node is now *a station at a particular time* and is connected to neighbouring stations (with respect to $G$) and to itself at the previous timestamp and the following timestamp. Notice that the new spatio-temporal graph $H$ is the Cartesian product of the spatial graph $G$ and the temporal graph $G'$ (which is simply a line graph, without loop).

- Express the Laplacian of $H$ using the Laplacian of $G$ and $G'$ (use Kronecker products).
- Express the eigenvalues and eigenvectors of the Laplacian of $H$ using the eigenvalues and eigenvectors of the Laplacian of $G$ and $G'$.
- Compute the wavelet transform of the temperature signal.
- Classify nodes into low/medium/high frequency and display the same figure as in the previous question.

### Answer 8

- **Laplacian of the Cartesian product of graphs $G$ and $G'$ :**

  The new spatio-temporal graph $H$ is the Cartesian product of the spatial graph $G$ and the temporal graph $G'$. The Laplacian of $H$ is therefore the kronecker sum of the laplacian of $G$ and that of $G'$, which can be written using Kronecker products as follows:

$$\boxed{L(H) = L(G) \oplus L(G') = L(G) \otimes I_n + I_m \otimes L(G')}$$

  where, $\begin{cases} \text{n : number of nodes in graph G} \\ \text{m : number of nodes in graph G}' \end{cases}$

- **Eigenvalues and eigenvectors of the Laplacian of $H$:**

  The eigenvalues of the Laplacian of $H$ are the pair-wise sum of the eigenvalues of the laplacian of $G$ and that of $G'$ : $\{\lambda_1 + \mu_1, \lambda_1 + \mu_2, ..., \lambda_n + \mu_m\}$

  The eigenvectors of the Laplacian of $H$ are the pair-wise kronecker product of the eigenvectors of the laplacian of $G$ and that of $G'$ : $\{u_1 \otimes v_1, u_1 \otimes v_2, ..., u_n \otimes v_m\}$

  *Proof*:
  Let $u_i$ be the $i^{th}$ eigenvector of the graph $G$ associated to the eigenvalue $\lambda_i$, and let $v_j$ be the $j^{th}$ eigenvector of the graph $G'$ associated to the eigenvalue $\mu_j$.

$$\begin{aligned} L(H)(u_i \otimes v_j) &= (L(G) \otimes I_n + I_m \otimes L(G'))(u_i \otimes v_j) \\ &= L(G)u_i \otimes I_n v_j + I_m u_i \otimes L(G')v_j \\ &= \lambda_i u_i \otimes v_j + \mu_j u_i \otimes v_j \\ &= (\lambda_i + \mu_j)\ u_i\ \otimes v_j \end{aligned}$$

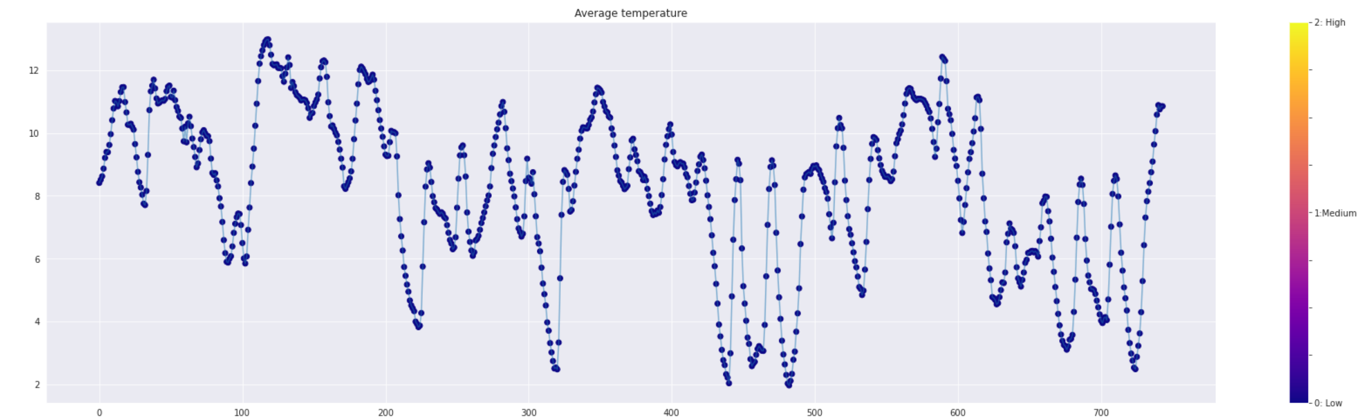



Figure 6: Average temperature. Markers' colours depend on the majority class.

**Comment:**

From the previous plot, if we consider the time dimension in the construction of the graph, we can notice that for all time stamps, the low frequency category is also in the majority.