Assignment 3 (ML for TS) - MVA 2021/2022
refrence : http://www.laurentoudre.fr/ast.html

March 21, 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.

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 $x = \{x_t\}_{t=0}^{T-1}$,

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

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 ε_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:

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

where pen is a penalty function. In addition, let

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

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

- Since ε_t has 0 as mean and $\sigma^2 I_d$ covariance. we get:

$$\mathbb{E}[\|\varepsilon\|^2 / T] = d\sigma^2$$

- Since:

$$\begin{aligned}
< a, \Pi_{\boldsymbol{\tau}} b > &= \sum_{t=0}^{T} < a_t, (\Pi_{\boldsymbol{\tau}} b)_t > \\
&= \sum_{k=0}^{|\tau|} \sum_{t_k \le t < t_{k+1}} < a_t, \bar{b}_{t_k \cdots t_{k+1}} > \\
&= \sum_{k=0}^{|\tau|} \sum_{t_k \le t < t_{k+1}} < a_t, \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t' < t_{k+1}} b_{t'} > \\
&= \sum_{k=0}^{|\tau|} \sum_{t_k \le t < t_{k+1}} \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t' < t_{k+1}} < a_t, b_{t'} > \\
&= \sum_{k=0}^{|\tau|} \sum_{t_k \le t' < t_{k+1}} < \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t < t_{k+1}} a_t, b_{t'} > \\
&= \sum_{k=0}^{|\tau|} \sum_{t_k \le t' < t_{k+1}} < \bar{a}_{t_k \cdots t_{k+1}}, b_{t'} > \\
&= < \Pi_{\boldsymbol{\tau}} a, b >
\end{aligned}$$

Hence $\Pi_{\tau}^* = \Pi_{\tau}$. Besides, it is easy to see that Π_{τ} is a projection: $\Pi_{\tau}^2 = \Pi_{\tau}$, therefore:

$$
\begin{aligned}
\mathbb{E}[\|f - \Pi_{\tau} y\|^2 / T] &= \mathbb{E}[\|f - \Pi_{\tau} f - \Pi_{\tau} \varepsilon\|^2 / T] \\
&= \frac{1}{T}\mathbb{E}[\|f - \Pi_{\tau} f\|^2 + \|\Pi_{\tau}\varepsilon\|^2 - 2 < f - \Pi_{\tau} f, \Pi_{\tau}\varepsilon >] \\
&= \frac{1}{T}(\|f - \Pi_{\tau} f\|^2 + \mathbb{E}[< \Pi_{\tau}\varepsilon, \Pi_{\tau}\varepsilon > -2 < \Pi_{\tau}(f - \Pi_{\tau} f), \varepsilon >]) \\
&= \frac{1}{T}(\|f - \Pi_{\tau} f\|^2 + \mathbb{E}[< \Pi_{\tau}^2\varepsilon, \varepsilon > -2 < \Pi_{\tau} f - \Pi_{\tau}^2 f, \varepsilon >]) \\
&= \frac{1}{T}(\|f - \Pi_{\tau} f\|^2 + \mathbb{E}[< \Pi_{\tau}\varepsilon, \varepsilon > -2 < \Pi_{\tau} f - \Pi_{\tau} f, \varepsilon >]) \\
&= \frac{1}{T}(\|f - \Pi_{\tau} f\|^2 + \mathbb{E}[< \Pi_{\tau}\varepsilon, \varepsilon >])
\end{aligned}
$$

Yet:

$$
\begin{aligned}
\mathbb{E}[< \Pi_{\tau}\varepsilon, \varepsilon >] &= \sum_{t=0}^{T} \mathbb{E}[< (\Pi_{\tau}\varepsilon)_t, \varepsilon_t >] \\
&= \sum_{k=0}^{|\tau|} \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t, t' < t_{k+1}} \mathbb{E}[< \varepsilon_{t'}, \varepsilon_t >] \\
&= \sum_{k=0}^{|\tau|} \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t < t_{k+1}} \mathbb{E}[\|\varepsilon_t\|^2] \ \ (\{\varepsilon_t\}_t \text{ are i.i.d. with 0 mean}) \\
&= \sum_{k=0}^{|\tau|} \frac{1}{t_{k+1} - t_k} \sum_{t_k \le t < t_{k+1}} d\sigma^2 \\
&= (|\tau| + 1) d\sigma^2
\end{aligned}
$$

Therefore:

$$
\mathbb{E}[\|f - \Pi_{\tau} y\|^2 / T] = \frac{1}{T}(\|f - \Pi_{\tau} f\|^2 + (|\tau| + 1) d\sigma^2)
$$

- Finally, it is easy to see that $(Id - \Pi_{\tau})^* = (Id - \Pi_{\tau})$ and $(Id - \Pi_{\tau})^2 = (Id - \Pi_{\tau})$, hence:

$$
\begin{aligned}
\mathbb{E}[\|y-\Pi_{\boldsymbol{\tau}} y\|^2/T] &= \frac{1}{T}\mathbb{E}[\|(Id-\Pi_{\boldsymbol{\tau}})f+(Id-\Pi_{\boldsymbol{\tau}})\varepsilon\|^2] \\
&= \frac{1}{T}\mathbb{E}[\|(Id-\Pi_{\boldsymbol{\tau}})f\|^2+\|(Id-\Pi_{\boldsymbol{\tau}})\varepsilon\|^2-2<(Id-\Pi_{\boldsymbol{\tau}})f,(Id-\Pi_{\boldsymbol{\tau}})\varepsilon>] \\
&= \frac{1}{T}\mathbb{E}[\|(Id-\Pi_{\boldsymbol{\tau}})f\|^2+\|(Id-\Pi_{\boldsymbol{\tau}})\varepsilon\|^2]-2<(Id-\Pi_{\boldsymbol{\tau}})f,(Id-\Pi_{\boldsymbol{\tau}})\mathbb{E}[\varepsilon]> \\
&= \frac{1}{T}(\|(Id-\Pi_{\boldsymbol{\tau}})f\|^2+\mathbb{E}[<(Id-\Pi_{\boldsymbol{\tau}})\varepsilon,(Id-\Pi_{\boldsymbol{\tau}})\varepsilon>]) \\
&= \frac{1}{T}(\|(Id-\Pi_{\boldsymbol{\tau}})f\|^2+\mathbb{E}[<(Id-\Pi_{\boldsymbol{\tau}})^2\varepsilon,\varepsilon>]) \\
&= \frac{1}{T}(\|(Id-\Pi_{\boldsymbol{\tau}})f\|^2+\mathbb{E}[<(Id-\Pi_{\boldsymbol{\tau}})\varepsilon,\varepsilon>]) \\
&= \frac{1}{T}(\|f-\Pi_{\boldsymbol{\tau}} f\|^2+\mathbb{E}[\|\varepsilon\|^2]-\mathbb{E}[<\Pi_{\boldsymbol{\tau}}\varepsilon,\varepsilon>]) \\
&= \frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} f\|^2+(1-\frac{|\tau|+1}{T})d\sigma^2
\end{aligned}
$$

Since:

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

Then:

$$\mathbb{E}[\mathrm{crit}(\boldsymbol{\tau})] = \frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} f\|^2+(1-\frac{|\tau|+1}{T})d\sigma^2+\mathbb{E}[\mathrm{pen}(\boldsymbol{\tau})]$$

Hence, in order to have $\mathbb{E}\left[\mathrm{crit}(\boldsymbol{\tau})\right] \approx \mathbb{E}\left[\frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} y\|^2\right]$, the penalty should then verify:

$$\frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} f\|^2+(1-\frac{|\tau|+1}{T})d\sigma^2+\mathbb{E}[\mathrm{pen}(\boldsymbol{\tau})] \approx \mathbb{E}\left[\frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} y\|^2\right] = \frac{1}{T}\|f-\Pi_{\boldsymbol{\tau}} f\|^2+\frac{|\tau|+1}{T}d\sigma^2$$

Meaning that the ideal penalty could be chosen such that:

$$\mathrm{pen}_{\mathrm{id}}(\boldsymbol{\tau}) = (2\frac{|\tau|+1}{T}-1)d\sigma^2$$

Or by getting rid of the constant terms:

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

Question 2 *Mallows' C_p*

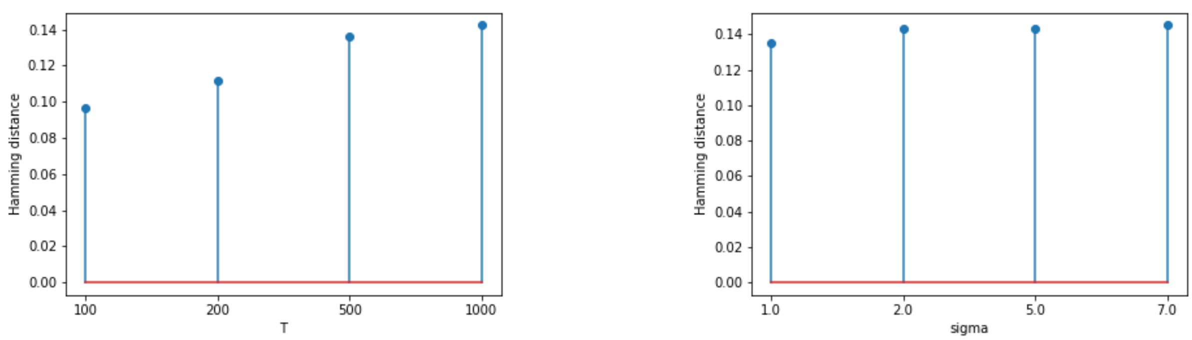
The ideal penalty depends on the unknown value of σ. Pluging an estimator $\hat{\sigma}$ into $\mathrm{pen}_{\mathrm{id}}$ yields the well-known Mallows' C_p. Use the empirical variance on the first 10% of the signal as an estimator of σ^2.

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 σ

Figure 1: Performance of Mallows' C_p

Question 3 *Slope heuristics*

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

Slope heuristics algorithm.

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

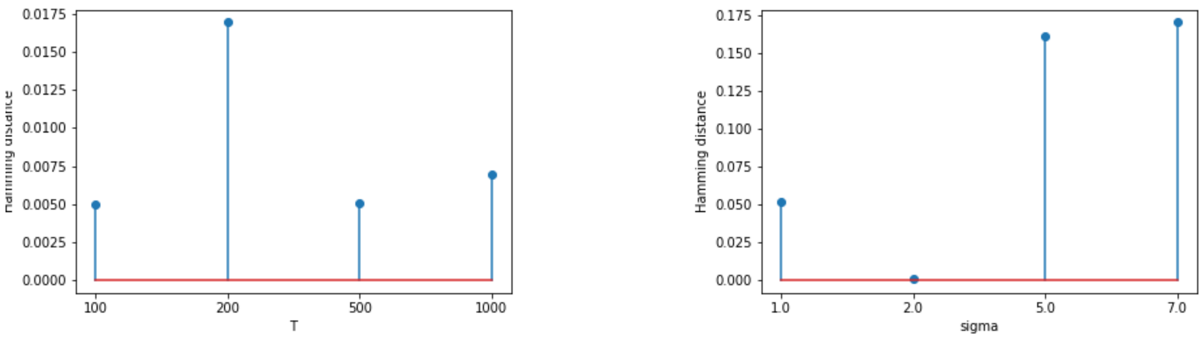
In simulations, "large enough" means for K between 15 and $0.4T$.

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 σ

Figure 2: Performance of the slope heuristics

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

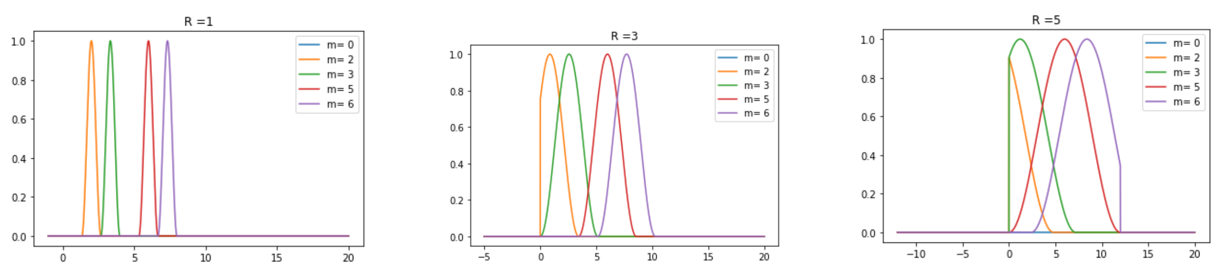
$$\hat{g}^U(\lambda) := \frac{1}{2}\left[1 + \cos\left(2\pi\left(\frac{\lambda}{aR} + \frac{1}{2}\right)\right)\right] \mathbb{1}(-Ra \leq \lambda < 0) \tag{11}$$

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

A bigger R stretches the gap between the zeros of the kernels, i.e. it reduces their support measure.

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:

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

The threshold is equal to 0.83.

The signal is the least smooth at timestamp = "2014-01-21 03:00:00", with a smoothness value equal to 583.85.

The signal is the smoothest at timestamp = "2014-01-24 23:00:00", with a smoothness value equal to 20.67.

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

If Laplacian L is decomposed using spectral theorem: $L = U\Lambda U^T$, then:

$$
\begin{aligned}
W_f(m, v) &= \sum_{l=1}^{n} \hat{g}_m(\lambda_l)(f^T u_l)(u_l^T v) \\
&= f^T (\sum_{l=1}^{n} \hat{g}_m(\lambda) u_l u_l^T) v \\
&= f^T (U \hat{g}_m(\Lambda) U^T) v
\end{aligned}
$$

Hence:

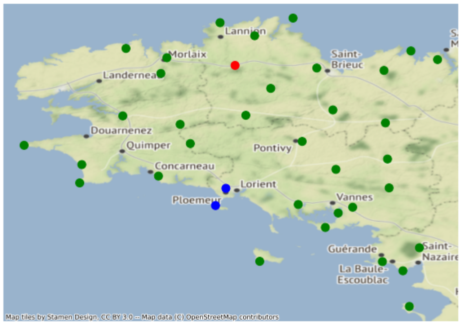
$$
W_f(m, \cdot) = f^T (U \hat{g}_m(\Lambda) U^T)
$$

By coding this formula, one can get the bellow set of figures, where the following color-code has been used:

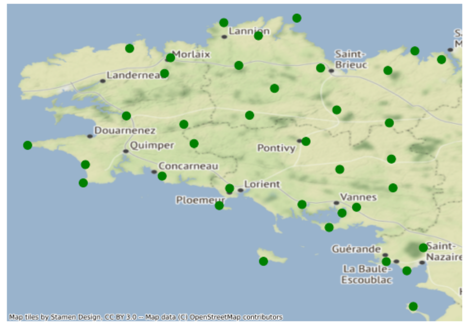
$$
Green \Rightarrow \text{Low frequency node}
$$

$$
Blue \Rightarrow \text{Medium frequency node}
$$

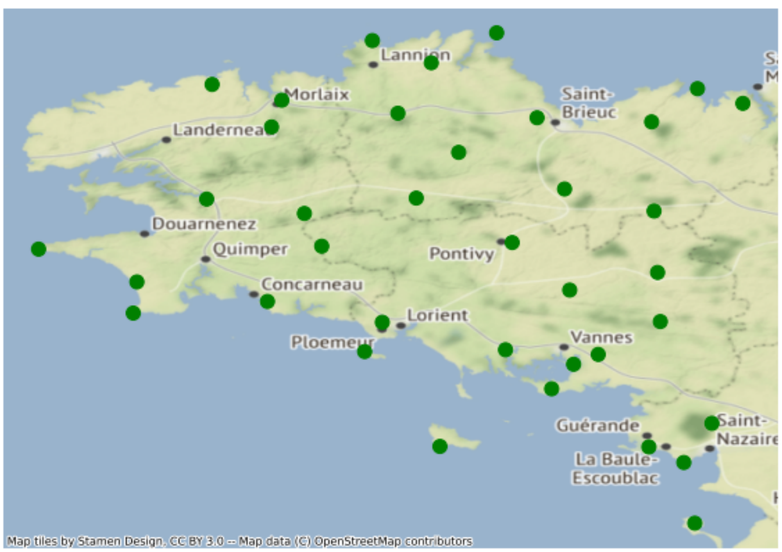
$$
Red \Rightarrow \text{High frequency node}
$$



(a) Least smooth signal



(b) Smoothest signal



(c) First available timestamp

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

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

While using the same previous color-code, we get the following average temperature curve, with low frequency as majority class at all timestamps:

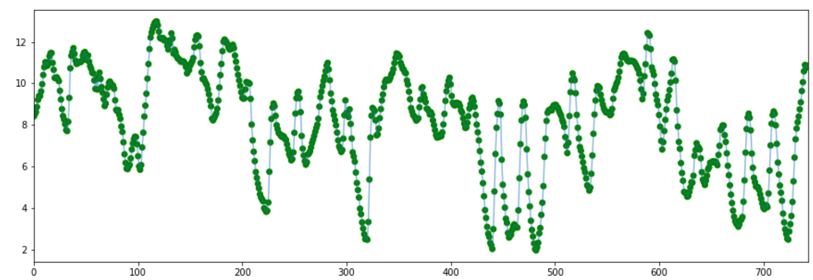


Figure 5: Average temperature. Markers' colours depend on the majority class.

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

- In order to express the Laplacian of H using the Laplacian of G and G', one should first express the adjacency matrix H using the adjacency matrix of G and G'. It is possible then to notice that, $A(H), A(G), A(G')$ are adjacency matrices of H, G, G' respectively then:

$$A(H) = I_T \otimes A(G) + A(G') \otimes I_n$$

In fact, a node in H could be represented as (t, v), where $t \in G'$ and $v \in G$, hence:

$$
\begin{aligned}
(I_T \otimes A(G) + A(G') \otimes I_n)[(t_1, v_1), (t_2, v_2)] &= (I_T \otimes A(G))[(t_1, v_1), (t_2, v_2)] \\
&+ (A(G') \otimes I_n)[(t_1, v_1), (t_2, v_2)] \\
&= \mathbb{I}_{(t_1=t_2)} A(G)[v_1, v_2] + \mathbb{I}_{(v_1=v_2)} A(G')[t_1, t_2] \\
&:= A(H)[(t_1, v_1), (t_2, v_2)]
\end{aligned}
$$

If we denote $D(H), D(G), D(G')$ the degree matrices of H, G, G' respectively then, it is easy to establish that:

$$D(G) = diag(L(G)\mathbb{1}_n)$$

And:

$$D(G') = diag(L(G')\mathbb{1}_T)$$

And finally, it is easy to see that $\mathbb{1}_{nT} = \mathbb{1}_T \otimes \mathbb{1}_n$, therefore:

$$
\begin{aligned}
D(H) &:= diag(A(H)\mathbb{1}_{nT}) \\
&= diag(I_T \otimes A(G)\mathbb{1}_{nT}) + diag(A(G') \otimes I_n\mathbb{1}_{nT}) \\
&= diag(I_T \otimes A(G)\mathbb{1}_T \otimes \mathbb{1}_n) + diag(A(G') \otimes I_n\mathbb{1}_T \otimes \mathbb{1}_n) \\
&= diag(I_T \otimes A(G)\mathbb{1}_T \otimes \mathbb{1}_n) + diag(A(G') \otimes I_n\mathbb{1}_T \otimes \mathbb{1}_n) \\
&= diag((I_T\mathbb{1}_T) \otimes (A(G)\mathbb{1}_n)) + diag((A(G')\mathbb{1}_T) \otimes (I_n\mathbb{1}_n)) \\
&= diag(I_T\mathbb{1}_T) \otimes diag(A(G)\mathbb{1}_n)) + diag(A(G')\mathbb{1}_T) \otimes diag(I_n\mathbb{1}_n)) \\
&= I_T \otimes D(G) + D(G') \otimes I_n
\end{aligned}
$$

The conclusion is that:

$$L(H) := D(H) - A(H) = I_T \otimes L(G) + L(G') \otimes I_n$$

- If $\{u_v\}_{1 \leq v \leq n}$ is the set of eigenvectors of $L(G)$ associated to $\{\lambda_v\}_{1 \leq v \leq n}$, and $\{h_t\}_{1 \leq t \leq T}$ is the set of eigenvectors of $L(G')$ associated to $\{\lambda'_t\}_{1 \leq t \leq T}$, then it is easy to see that for all $1 \leq v \leq n$ and $1 \leq t \leq T$:

$$
\begin{aligned}
L(H)(h_t \otimes u_v) &= I_T \otimes L(G)(h_t \otimes u_v) + L(G') \otimes I_n(h_t \otimes u_v) \\
&= (I_T h_t) \otimes (L(G)u_v) + (L(G')h_t) \otimes (I_n u_v) \\
&= h_t \otimes (\lambda_v u_v) + (\lambda'_t h_t) \otimes u_v \\
&= (\lambda_v + \lambda'_t) h_t \otimes u_v
\end{aligned}
$$

Hence $\{h_t \otimes u_v\}_{t,v}$ is a part of the spectrum of $L(H)$ associated with $\{\lambda_v + \lambda'_t\}_{t,v}$. Yet, since $\{h_t \otimes u_v\}_{t,v}$ is an orthogonal family (since $\{u_v\}_{1\leq v\leq n}$ and $\{h_t\}_{1\leq t\leq T}$ are orthogonal families and $< a \otimes b, c \otimes d >=< a, c >< b, d >$) of size nT, $\{h_t \otimes u_v\}_{t,v}$ is exactly the set of eigenvectors of $L(H)$ and $\{\lambda_v + \lambda'_t\}_{t,v}$ is its spectrum.

- The wavelet transform of the temperature signal was computed in the attached code, leading to the bellow figure, where again we have used the same color code, and again the low frequency is the majority class at all timestamps:

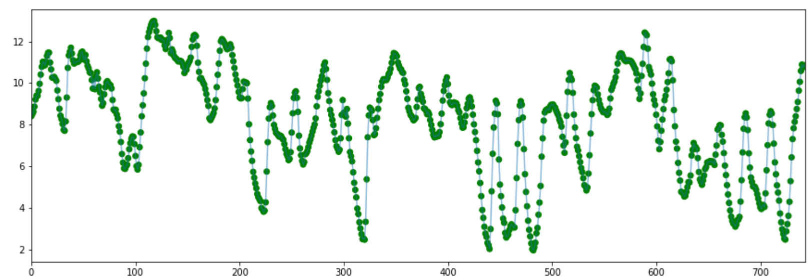


Figure 6: Average temperature. Markers' colours depend on the majority class.