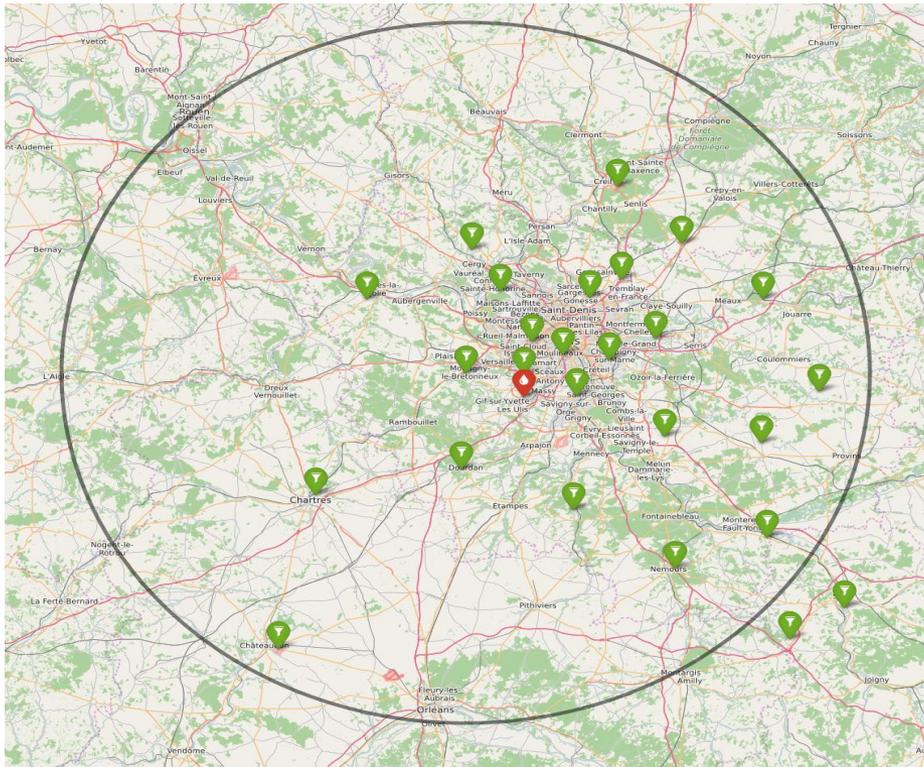


Iterative Multi-Scale Dynamic Time Warping (IMs-DTW): Estimation de la dissimilarité entre des séries temporelles de précipitations et application à la classification des événements pluvieux

M. D. DILMI , L. Barthes , C. MALLET

Problématique :

Caractérisation de la variabilité et de l'intermittence des précipitations à partir des observations de séries temporelles.



- Des données de la station SRITA (2 pluviomètres, disdromètre,...)
- Des données issues de 26 pluviomètres installés sur différentes stations en Île-de-France
- Chaque station fournit une série temporelle de mesures

Fig 1: les 26 stations de météo-France en Île-de-France + station SIRTA (Rouge)

Contexte

les précipitations : sont un phénomène intermittent caractérisé par une forte variabilité spatiale et temporelle. De plus, les précipitations résultent de plusieurs processus atmosphériques complexes. Pour cette raison, il existe différents types de précipitations (ex. stratiformes, convectives).

Dimensions du problème: l'étude de la variabilité des précipitations peut être abordée de plusieurs façons :

- 1. Classification des événements** : chaque type d'événements présentent des spécificités qui lui sont propres.
- 2. Étude de la variabilité Spatiale:** les précipitations manifestent des spécificités différentes selon l'endroit.
- 3. Étude de la variabilité Temporelle:** les précipitations manifestent des spécificités différentes d'une année à l'autre, une saisonnalité

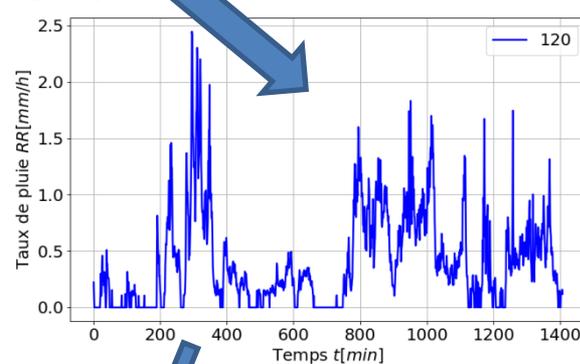
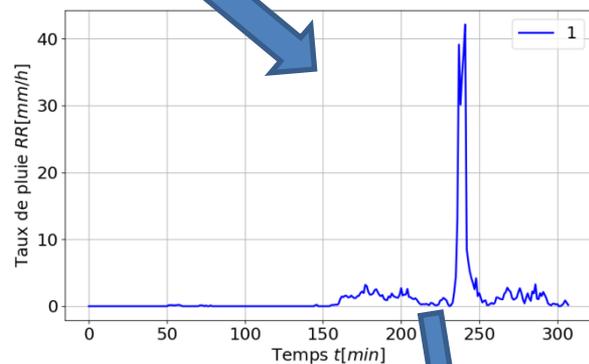
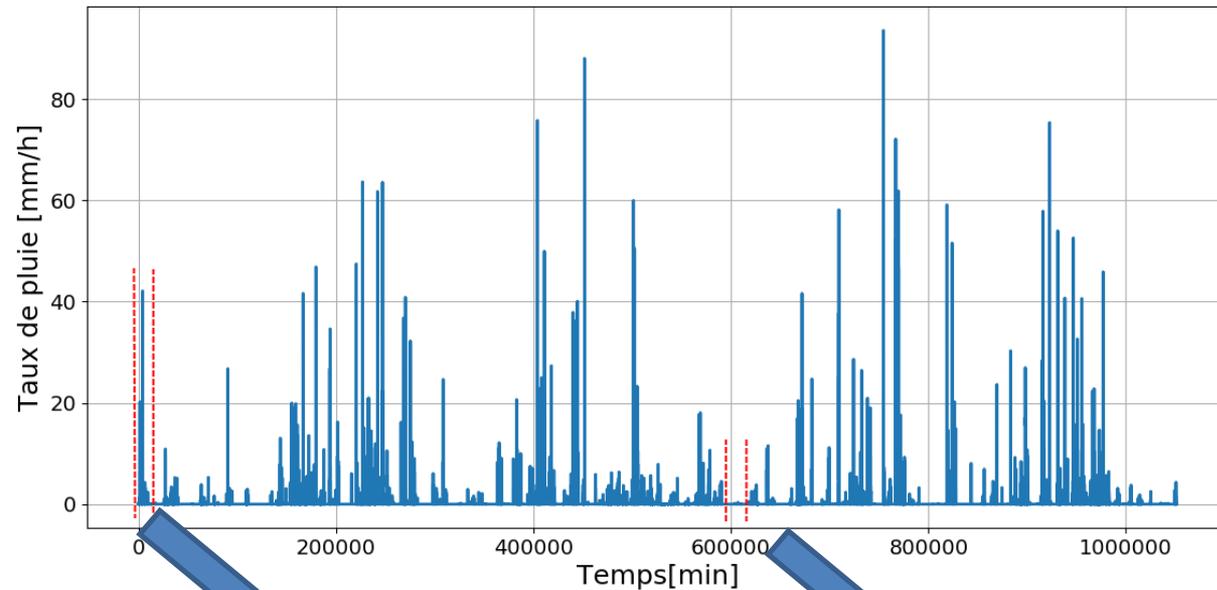
Plan de la présentation :

La présentation contient :

1. Caractérisation et classification des événements de précipitations en utilisant les données d'un disdromètre avec pas d'échantillonnage d'1min
2. L'IMs-DTW une dissimilarité adaptée à la comparaison des séries temporelles (événements) de précipitations + tests et validations
3. Exploitation de l'IMs-DTW et étude la variabilité des précipitations
 1. Présentation de la méthodologie
 2. Étude de la variabilité Spatiale: comparaison et classification (CAH, K-médoides) des 26 stations franciliennes différentes.
 3. Étude de la variabilité Temporelle:
 - classification des événements de précipitations des 26 stations
 - analyse de 142 années de données de paris-montsouris (climatique)

1-Characterisation et classification des événements de précipitations

1- Caractérisation et classification des événements de précipitations (1)



Description

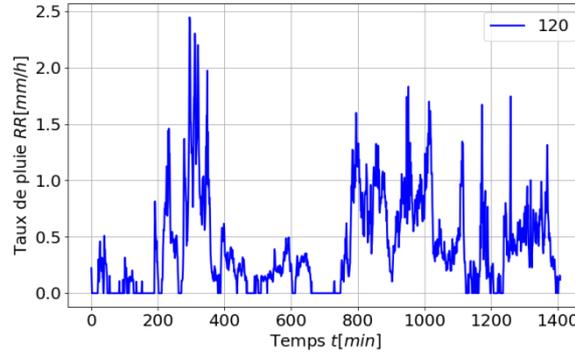
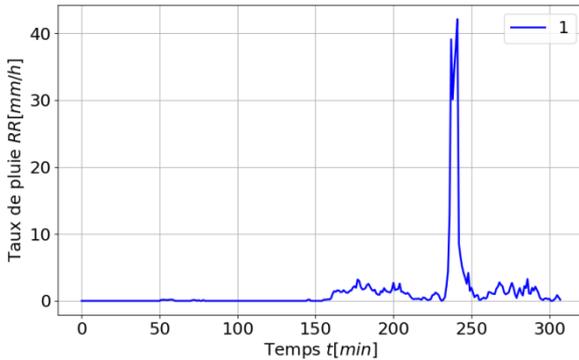
1. Série temporelle de taux précipitant mesurée entre 2013 et 2014 au SIRTA. Temps d'intégration = 1min. Instrument : Disdromètre.

2. Découpage de la série en 234 événements.

Incomparabilité => 3. recours à une liste variables caractérisant chaque événement

⇒ **Comment choisir ces variables ?**

1- Caractérisation et classification des événements de précipitations (2)



$$(D_e \quad \dots \quad R_m \quad \sigma_{RR} \quad P_C \quad \dots \quad R_{max} \quad \dots \quad R_d) \in \mathbb{R}^{23}$$



Selection des variables

(*algo génétique + SOM*)¹



$$(D_e \quad \sigma_{RR} \quad P_C \quad R_{max} \quad R_d) \in \mathbb{R}^5$$

(1) DILMI & al. 2017

Diversité des variables descriptives.

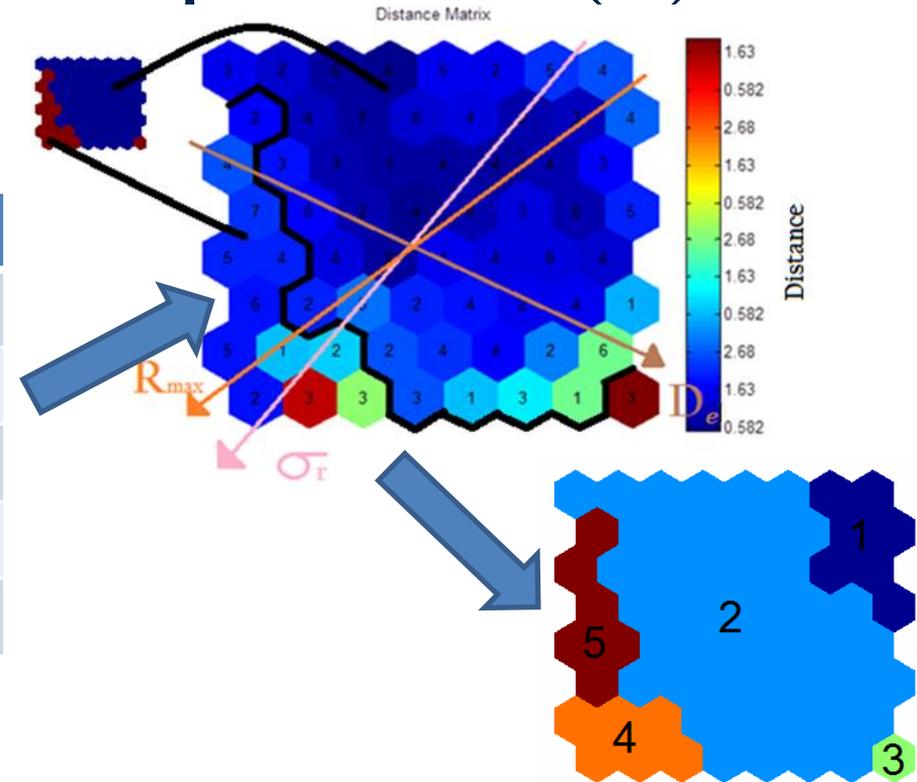
Problèmes liés à la grande dimension :

- Problème lié au poids différent accordé à chaque caractéristique.
- « Malédiction de la dimension » et problème de représentativité.
- sélection d'une description à 5 variables caractérisant chaque événement à 98%

(durée, écart-type, valeur max, contour vertical, cumul d'eau)
 $\in \mathbb{R}^5$

1- Caractérisation et classification des événements de précipitations (3)

	D_e	σ_{RR}	R_{max}	P_C	R_d
$Event_1$					
\vdots					
$Event_i$					
\vdots					
$Event_{234}$					



4. Classification SOM sur l'espace R^5 + CAH sur la carte obtenue:

- Un résultat correspondant à la classification stratiformes/convectives des événements.
- **Une CAH en 5 classes à l'échelle macro** est proposée ensuite (DILMI & al. 2017)

1-Characterisation et classification(5)

Transposition de l'étude sur des données de nature différentes

Description par caractéristiques :

1. sensible à la résolution,
2. perte d'information non quantifiée

=> trouver un autre moyen de comparer les séries de pluie

2-L'IMs-DTW une dissimilarité adaptée à la comparaison des séries temporelles

1-comparaison des séries temporelles de précipitations

2-la dissimilarité IMs-DTW (le concept)

3- application (analyse et interprétation)

4-comparaison des approches (description vs dissimilarité)

2-1-Comparaison des séries temporelles de précipitations

Deux séries temporelles présentent pour différentes raisons :

- Une différence d'amplitudes.
- Insertion/suppression
- Un décalage temporel variable

Distance euclidienne, corrélation, et corrélation croisée représentent des outils intéressants mais ne sont pas adaptés aux décalages temporels variables

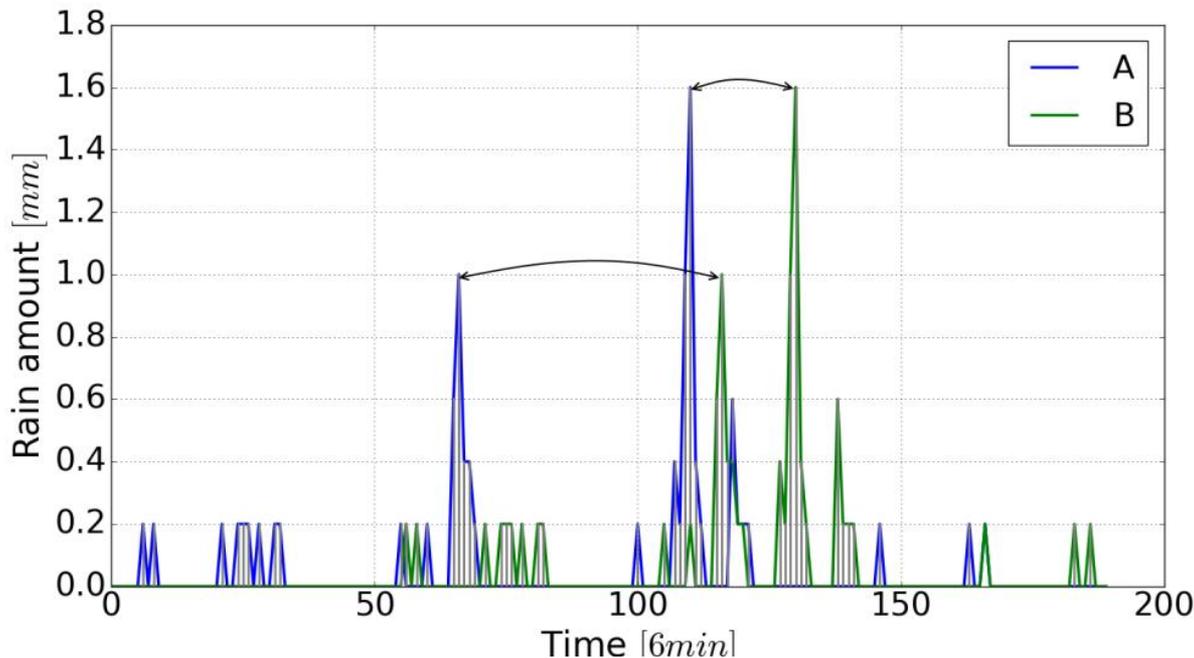


Fig. deux stations voisines enregistrant le même évènement avec des taux de pluie différents et des décalages temporels

⇒ **Besoin d'une méthode adaptée**

2-2-La dissimilarité IMs-DTW(concept)

Au lieu de comparer classiquement point à point, on propose de :
considérer une déformation élastique des deux axes du temps

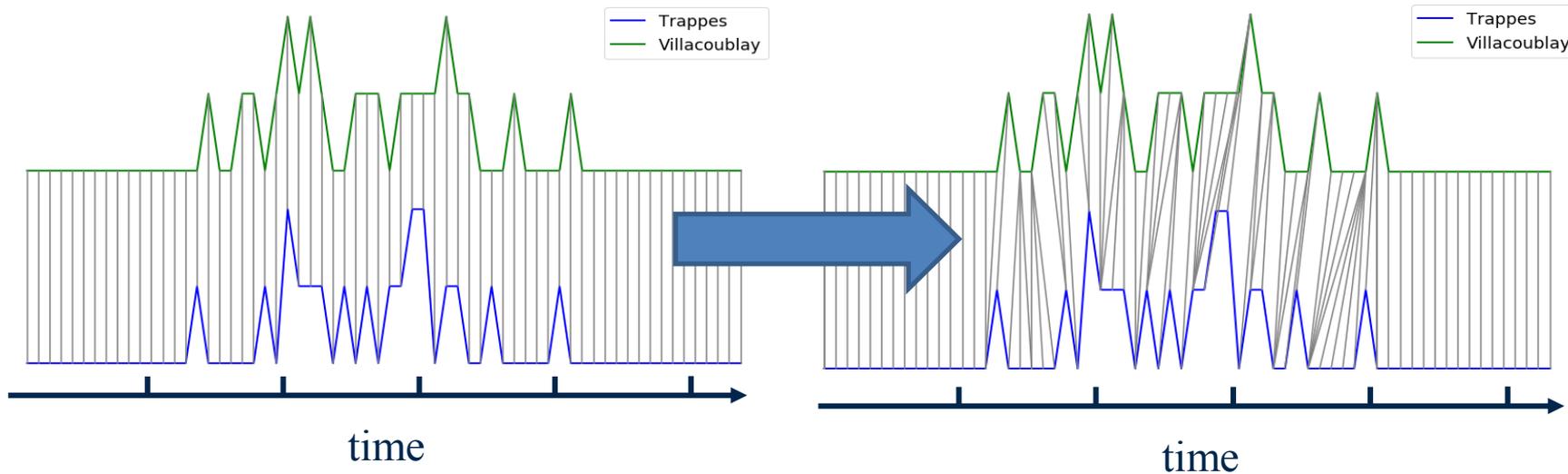


Fig. comparaison point à point des deux séries

Fig. comparaison déformée des deux séries

⇒ afin de comparer des motifs similaires mais décalés dans le temps.

2-2-La dissimilarité IMs-DTW(concept)

Pour clarifier la différence temporelle, une façon de voir est de développer les deux séries tout au long de l'axe i et l'axe j respectivement sur un plan $i - j$ (en Face).

Nous avons proposé un algorithme d'optimisation avec **contraintes d'échelles** capable de trouver le bon décalage (en rouge)

Contrainte : ressemblance sur toutes les echelles

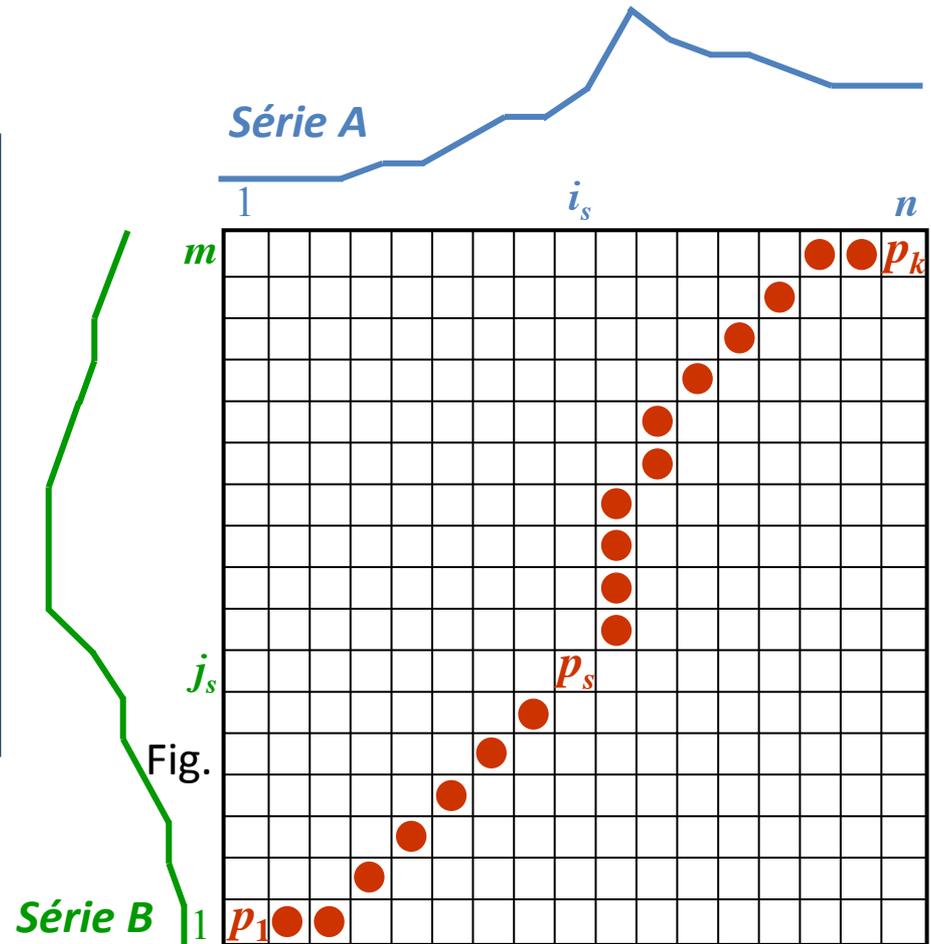
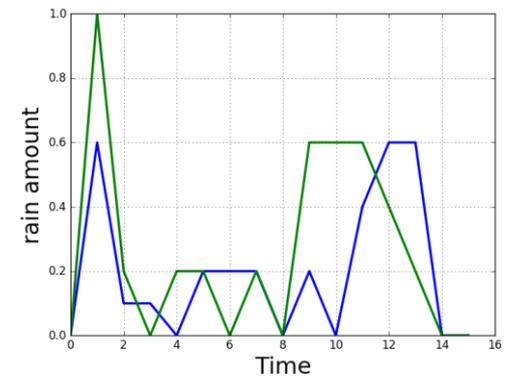
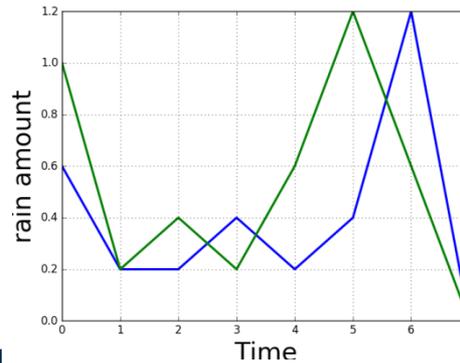
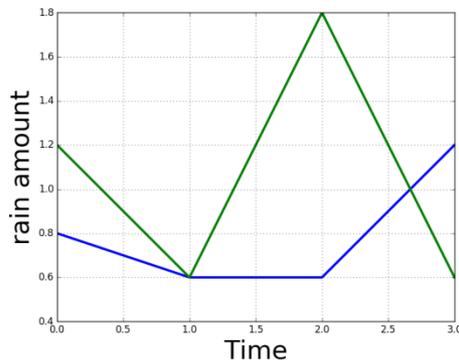
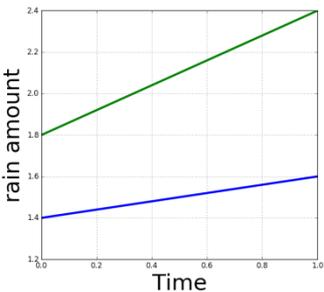
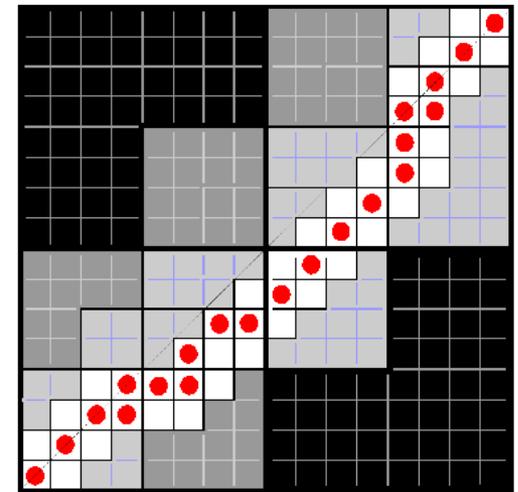
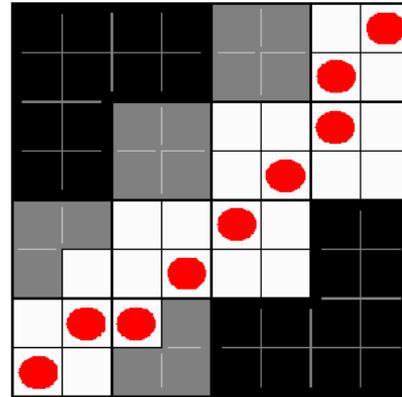
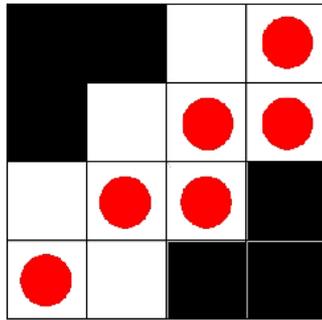
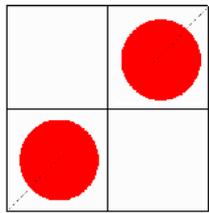


Fig. (en rouge)l'alignement nécessaire reliant les deux séries A et B

2-La dissimilarité IMs-DTW(concept)

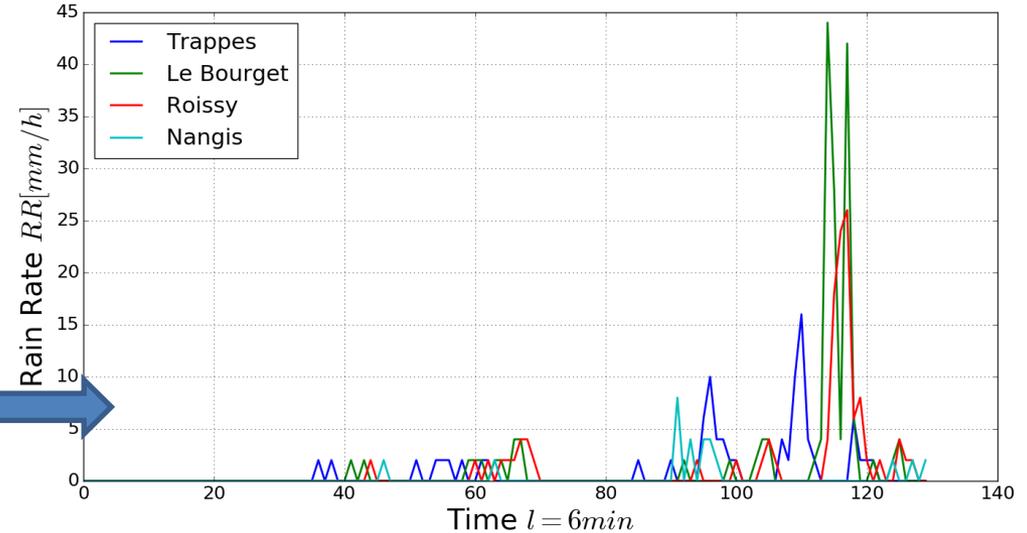
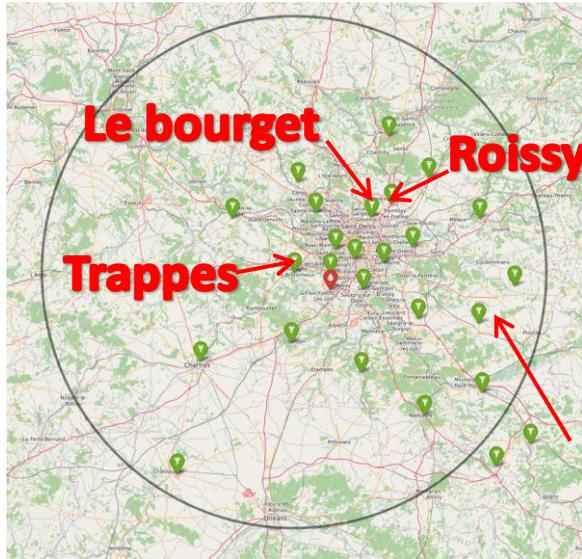
contraintes d'échelles :



Contrainte : ressemblance à toutes les echelles

2-3-Application de l'IMs-DTW (Analyse et interprétation)

L'événement du 11 juin 2009 enregistré à Trappes, Le Bourget, Roissy et Nangis entre 6h00 et 19h00

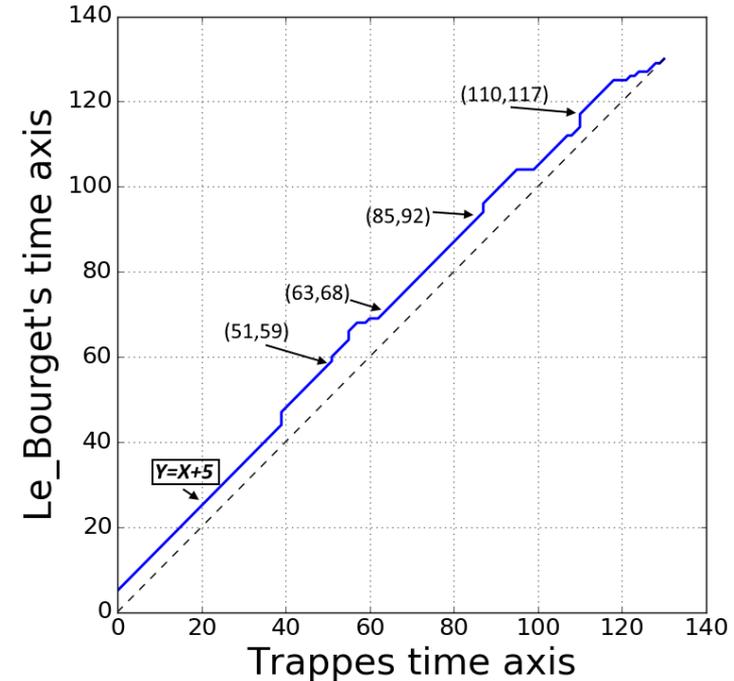


- N.B.:** ♦ Une cellule de pluie est passée par Trappes, ensuite Roissy en passant par Le Bourget
♦ Une cellule différente est passée par Nangis

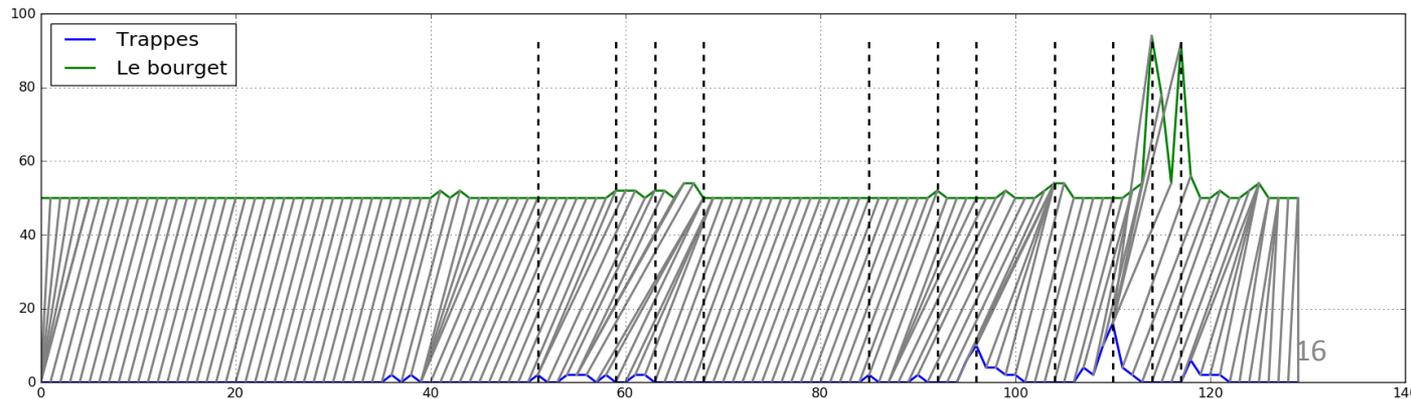
2-3-Application de l'IMs-DTW (Analyse et interprétation)

Le path associe les événements pluvieux enregistrés à trappes à ceux enregistrés à la station Le bourget (les pics de pluie, les moments de début et de fin, les durées inter-événements...)

Les périodes seches des deux séries sont associées avec un décalge de **5 pas = 30 min**. ça représente le temps moyen que prend un événement pour se déplacer entre les deux stations



Matching parfait->
(trappes est en
avance par 30min)



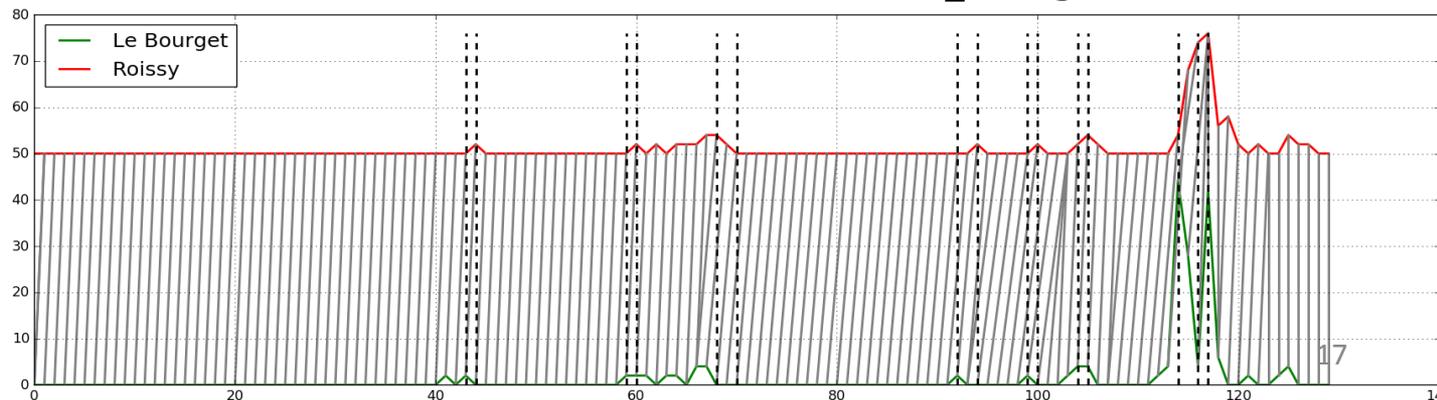
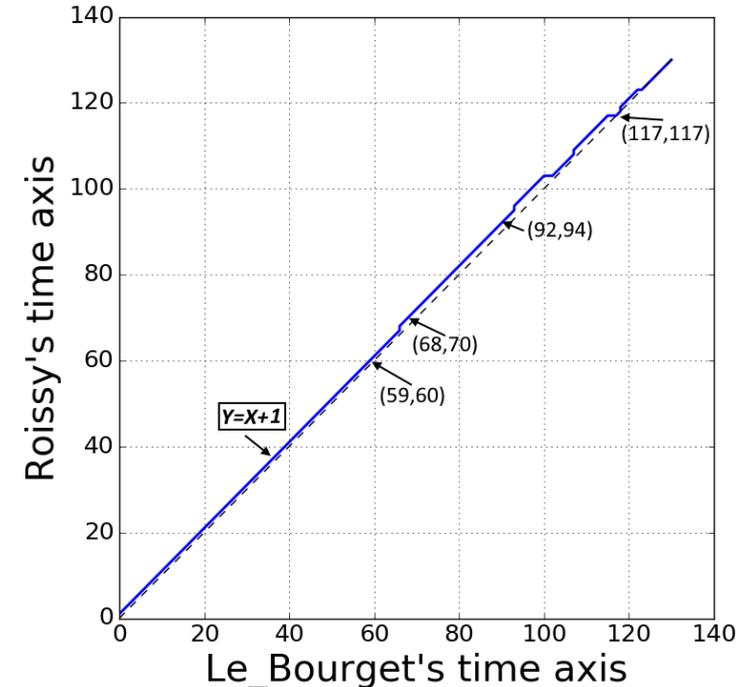
2-3-Application de l'IMs-DTW (Analyse et interprétation)

La station le bourget est en avance par rapport à Roissy et enregistre les événements en premier

un décalage entre 1 (6min) et 2 pas (12 min) en cohérence avec la distance géographique

Résultat attendu : entre Trappes et Roissy 42 min (si déplacement de la cellule et méthode efficace)

Le path admet comme interprétation un retard (déclage)

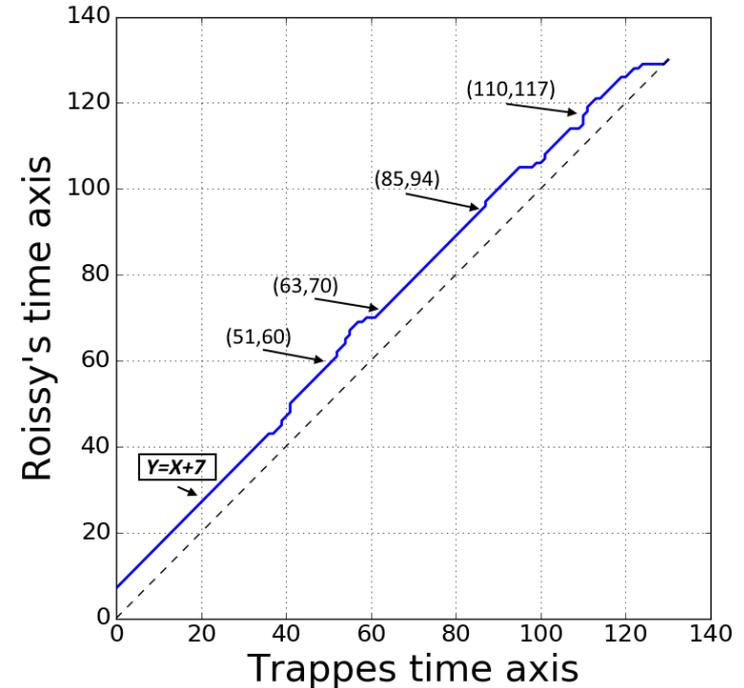


2-3-Application de l'IMs-DTW (Analyse et interprétation)

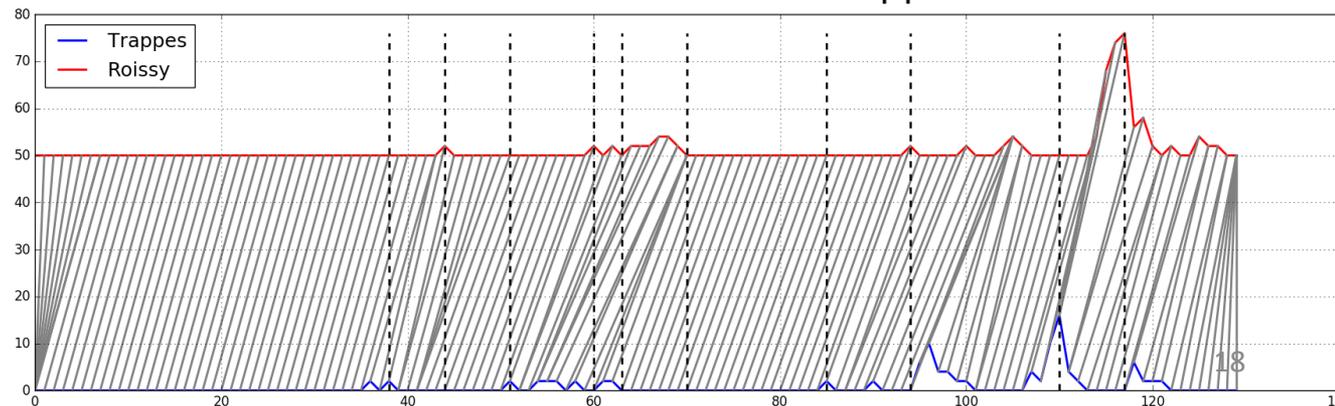
Un matching parfait entre les deux événements des stations Trappes et Roissy
La station de Trappes est en avance par 7 pas (42 min) par rapport à Roissy

Résumé :

Entre Trappes et Le bourget 30 min
Entre Le bourget et Roissy 12 min
Entre Trappes et Roissy 42 min
(reflète parfaitement le déplacement de la cellule)



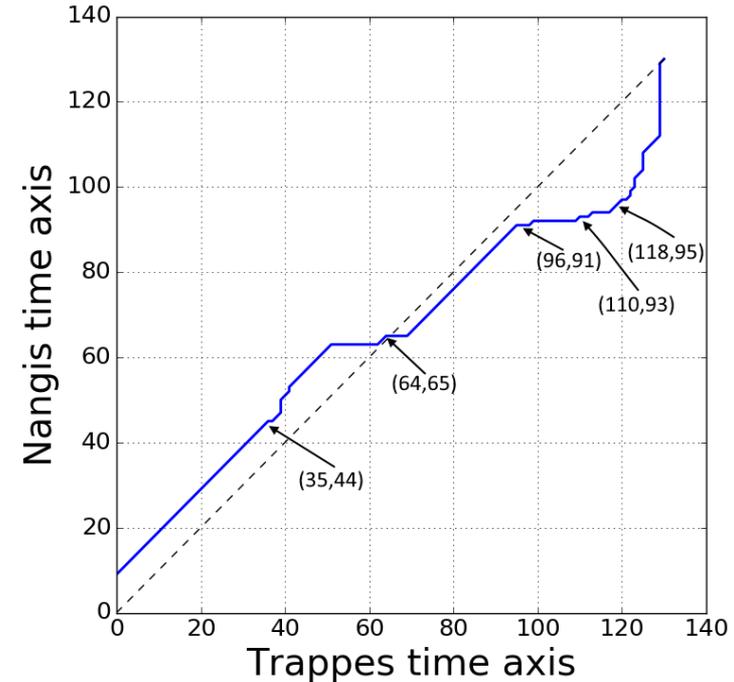
La relation entre les décalages des différentes stations est un moyen de validation du sens des paths



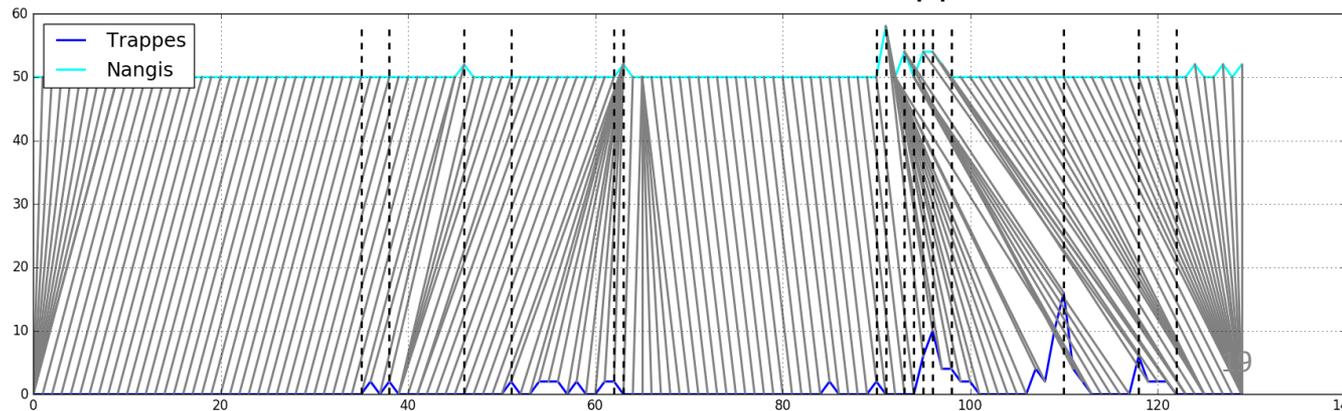
2-3-Application de l'IMs-DTW (Analyse et interprétation)

Les cartes radar montre que trappes et Nangis n'ont pas enregistré les mêmes épisodes pluvieux (cellules nuageuses)

Les associations du path donc essayent de repérer des similarités de motifs entre les deux stations et pas une déformation.



Le path ne représente pas des retards parce que un retard suppose un déplacement



2-3-Application de la méthode (Analyse et interprétation)

les mesures retournées par
l'IMs-DTW opérateur : carré de
la différence

Couple de Stations	Dissimilarité (IMs-DTW)	Distance euclidienne Normalisée
Trappes-Roissy	0.02	0.40
Trappes-Le Bourget	0.13	0.18
Le Bourget -Roissy	0.10	0.19
Trappes-Nangis	0.02	0.05
Le Bourget/Nangis	0.22	0.36
Roissy/Nangis	0.07	0.14

Interprétabilité?

2-Comparaison des deux approches

L'approche par caractéristiques

- Avantages:
 - Interprétable
 - Faible complexité
- Inconvénients:
 - Difficulté de transposition (instrument, échelle)
 - Perte d'information possible

Approche par IMs-DTW

- Avantages:
 - Facilement adaptable à l'échelle et à l'instrument
 - Comparaison globale sans réduction d'information
- Inconvénients:
 - Interprétation difficile
 - Complexité importante

⇒ Les inconvénients de l'une sont les avantages de l'autre

Idée : au lieu de les considérer comme des approches concurrentes, pourquoi ne pas combiner les deux approches et les considérer comme complémentaires?

3-Exploitation des données MF (travail en cours)

1-méthodologie

2-variabilité spatiale

3- variabilité temporelle

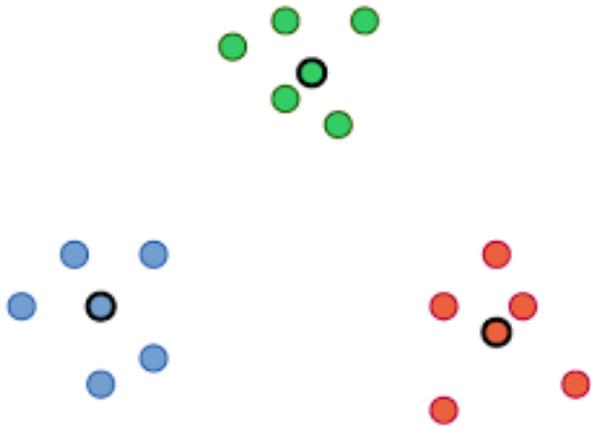
3-1-Méthodologie : combinaison des approches

- **La classification (un moyen d'étude de la variabilité)** : Une façon de réduire la dimensionnalité des données sans perdre leur variabilité est de les représenter sous forme de classes en les regroupant par similitude. Les classes permettent de caractériser des situations types.

- Approche proposée :
 - Étape 1 : effectuer une classification en K classes sur le jeu de données en utilisant l'IMS-DTW
 - Étape 2 : choisir un représentant de chaque classe
 - Étape 3 : analyser les K-représentants et interpréter le résultats en utilisant les variables descriptives

Les deux premières étapes représentent un algorithme de classification connu sous le nom de K-médoïdes

Algorithme des K-médoïdes



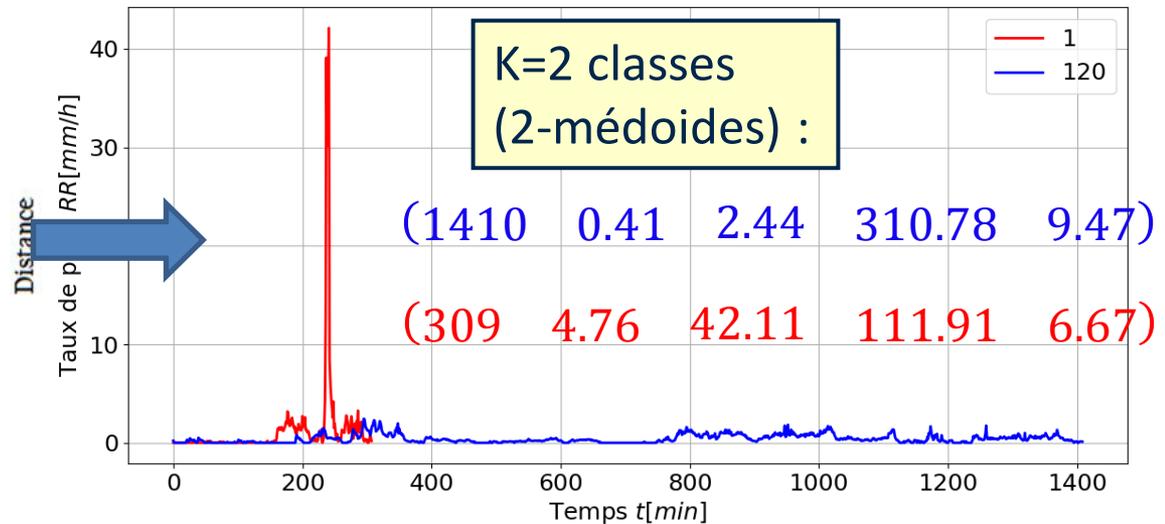
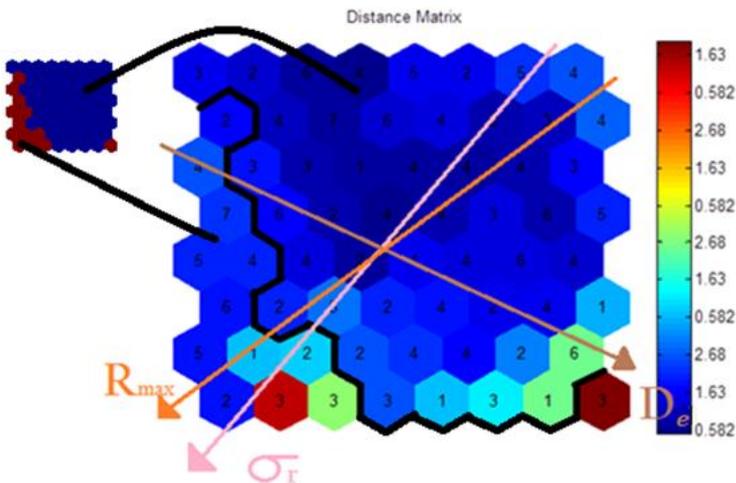
- Comme les k-moyennes, l'algorithme des k-médoïdes minimise l'erreur quadratique moyenne qui est la distance (dans notre cas dissimilarité) entre les points de la classe et le point central (ou médoïde).

Fig 1: exemple de K-médoïdes avec K=3

- **Avantage pratique** : les K-médoïdes proposent des situations réelles représentatives à analyser

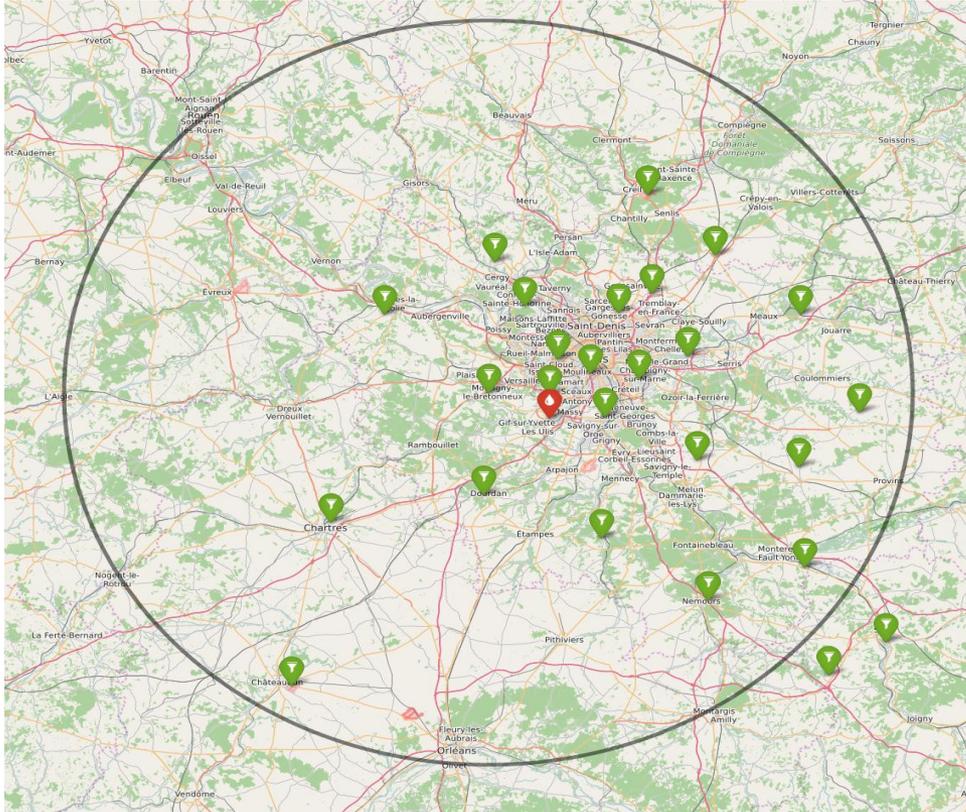
Exemple d'application : classification des 234 événements

$$(D_e \quad \sigma_{RR} \quad P_C \quad R_{max} \quad R_d) \in \mathbb{R}^5$$



Classification validée également pour K=5

3-2-Variabilité spatiale des précipitations en ile de France

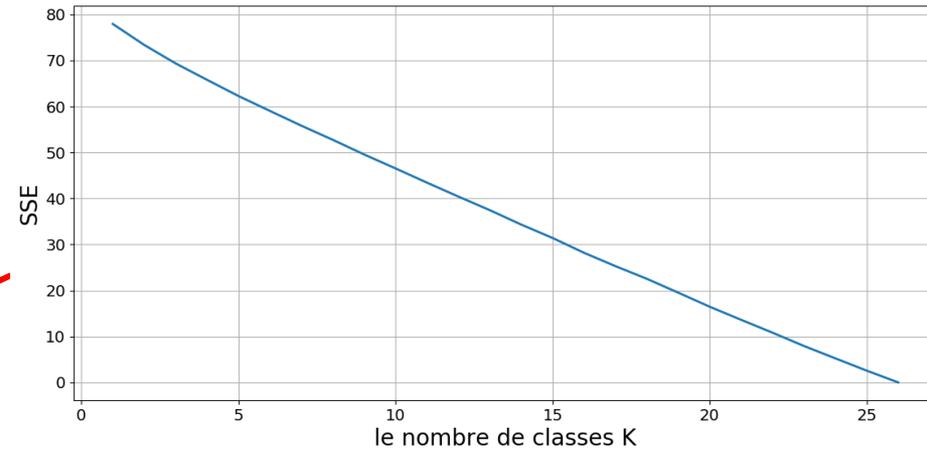
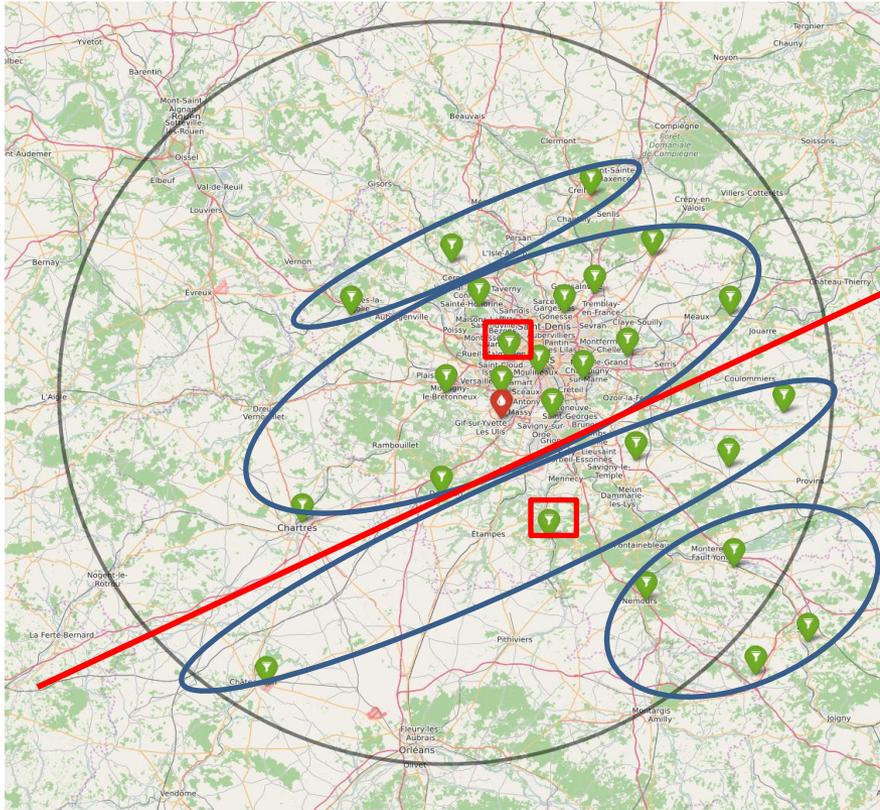


- Des données issues de 26 pluviomètres installés sur différentes stations en Île-de-France entre jan. 2009 et déc.2015
- Chaque station fournit une série temporelle de mesures avec un pas d'échantillonnage de 6 min(soit des séries de longueurs $240 \times 365 \times 7 = 613200$ échantillons chacune

- On applique la méthode précédente sur les 26 stations en variant $K=1:26$

Fig 1: les 26 stations de météo-France en Île-de-France (en vert)

3-2-Variabilité spatiale des précipitations en ile de France



$\forall K = 1:26,$
*la classification conserve
la topologie géographique*

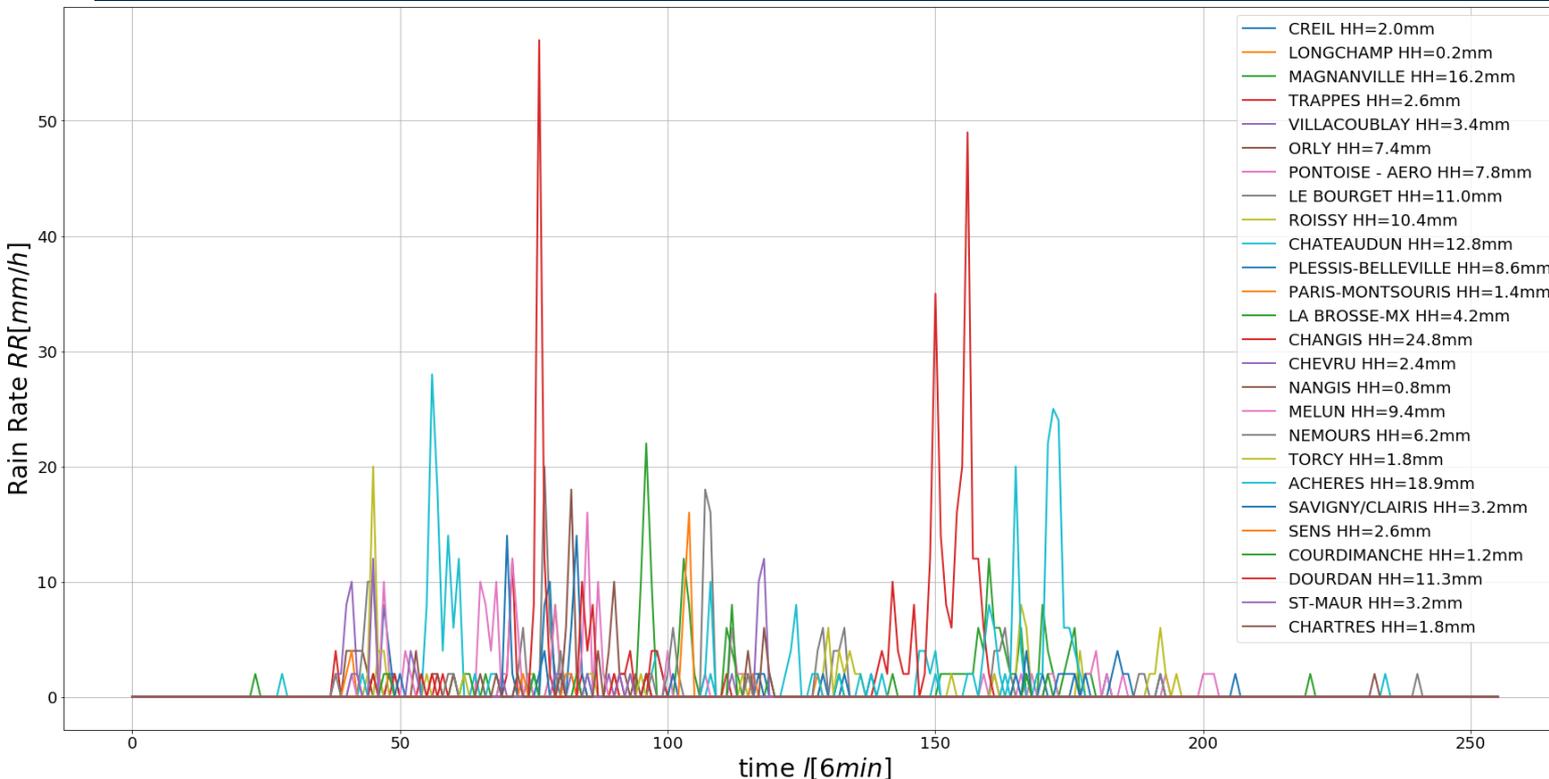
à ce niveau de comparaison pas
d'apport de nouvelles informations

Fig : classification en 2 (resp.4) classes en rouge (resp bleu) des mesures des 26 stations de météo-France en Île-de-France en 2015

⇒ Au lieu de comparer les stations sur tout l'axe, étudions la variabilité spatiale pour un événement précis

3-2-Variabilité spatiale des précipitations en ile de France (09 juin 2009)

Analyse par événement : les 26 stations ont toutes enregistré une trace de l'évènement pluvieux du 09 juin 2009

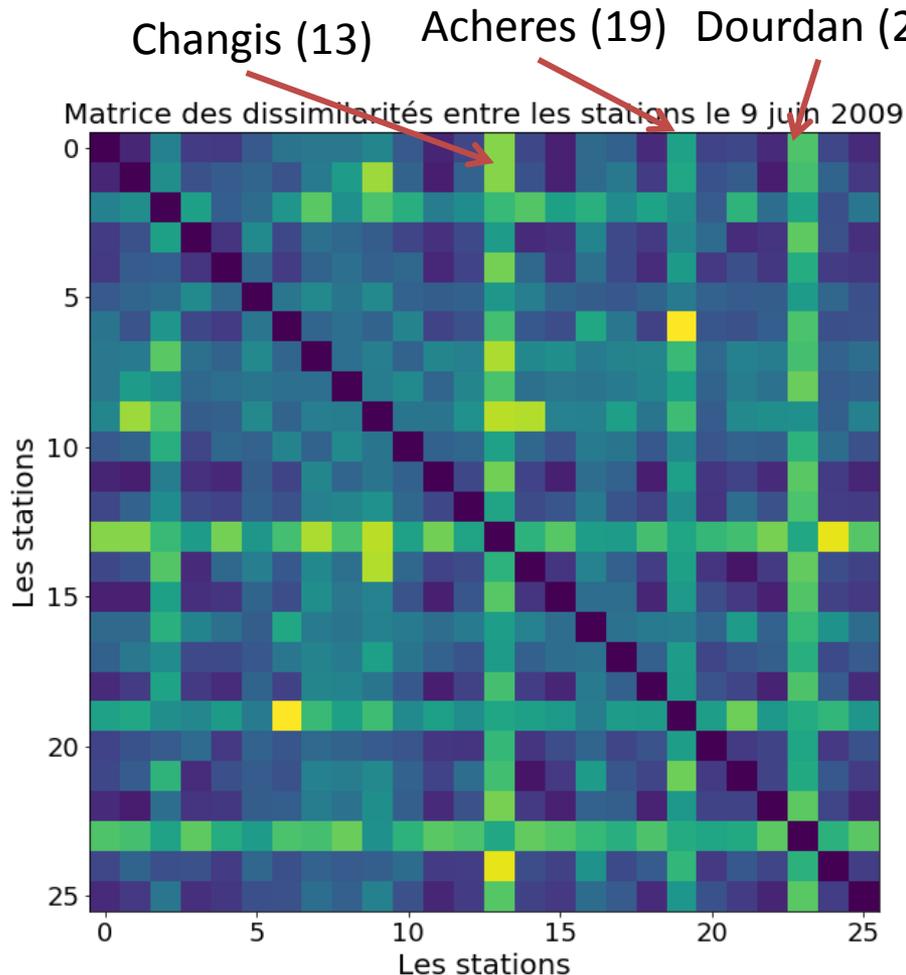


Visuellement, les stations **Changis, Acheres et Dourdan** se démarquent et enregistrent des quantités d'eau importantes.

⇒ Que fournit l'IMs-DTW appliquée sur les données enregistrées ce jours là?

3-2-Variabilité spatiale des précipitations en ile de France (09 juin 2009)

Les mesures de dissimilarités sont stockées dans la matrice des dissimilarités :



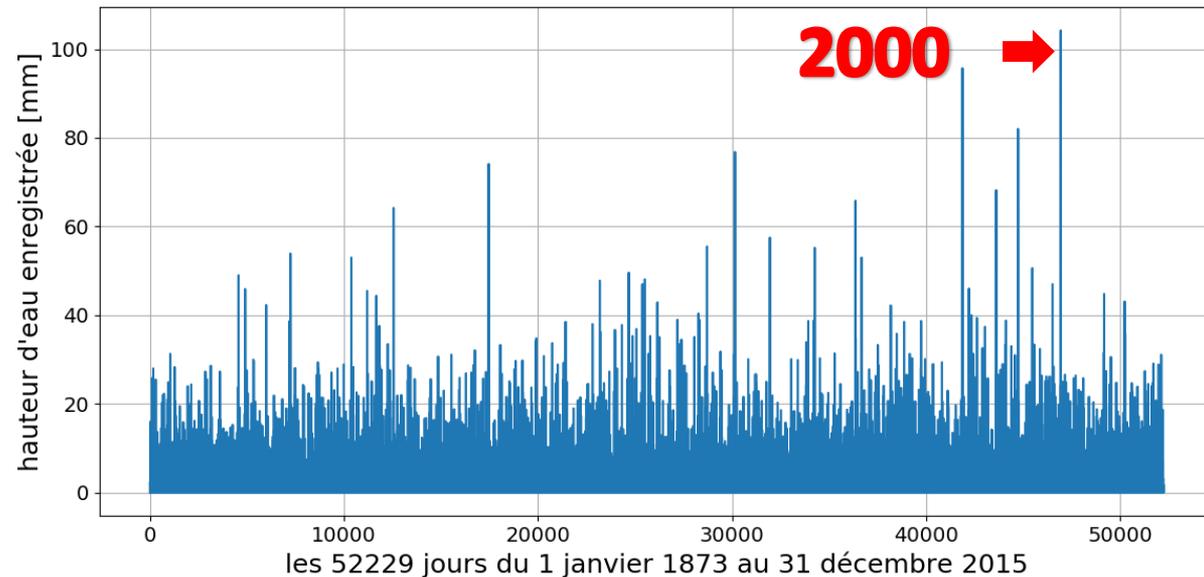
La case (i, j) de la matrice des dissimilarités contient la mesure de dissimilarité entre la série temporelle enregistrée sur la station i et la série temporelle enregistrée sur la station j

⇒ Les stations de Changis(13), Acheres(19) et Dourdan(23) se démarquent aussi sur la matrice des dissimilarités.

3-3- Variabilité temporelle des précipitations (impact climatique)

- La série historique/ la plus longue en France est la série de mesures faites à paris-montsouris en Île-de-France depuis sept. 1873

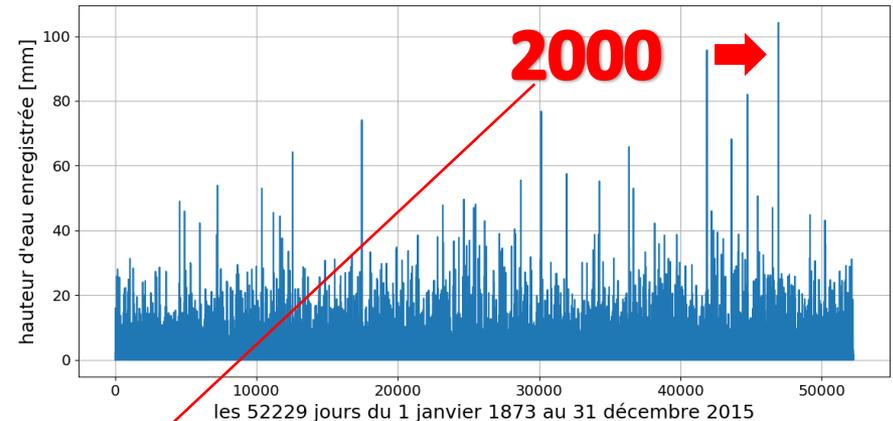
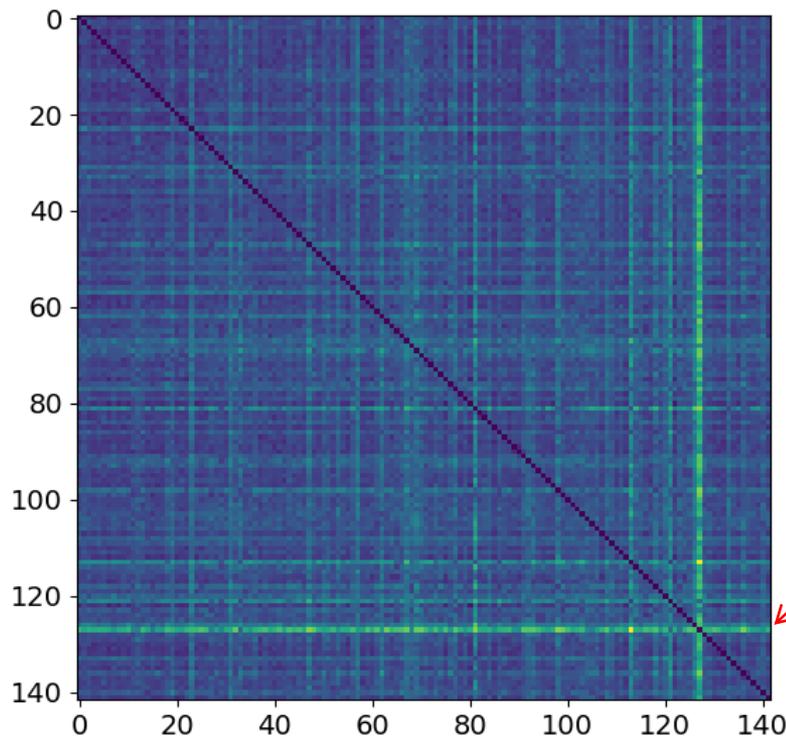
- On dispose donc d'une série temporelle de mesures couvrant 142 ans avec un pas d'échantillonnage d'un jour



- On calcule l'IMs-DTW sur les couples formés de paires des 142 années (ex. $D(1913,1919)$) ce qui donne une matrice de dissimilarité symétrique $D_{142 \times 142}$ et 10082 paths à analyser

3-3- Variabilité temporelle des précipitations (impact climatique)

- On calcule l'IMS-DTW sur les couples formés de pairs des 142 années (ex. $D(1913,1919)$) ce qui donne une matrice de dissimilarité symétrique $D_{142 \times 142}$

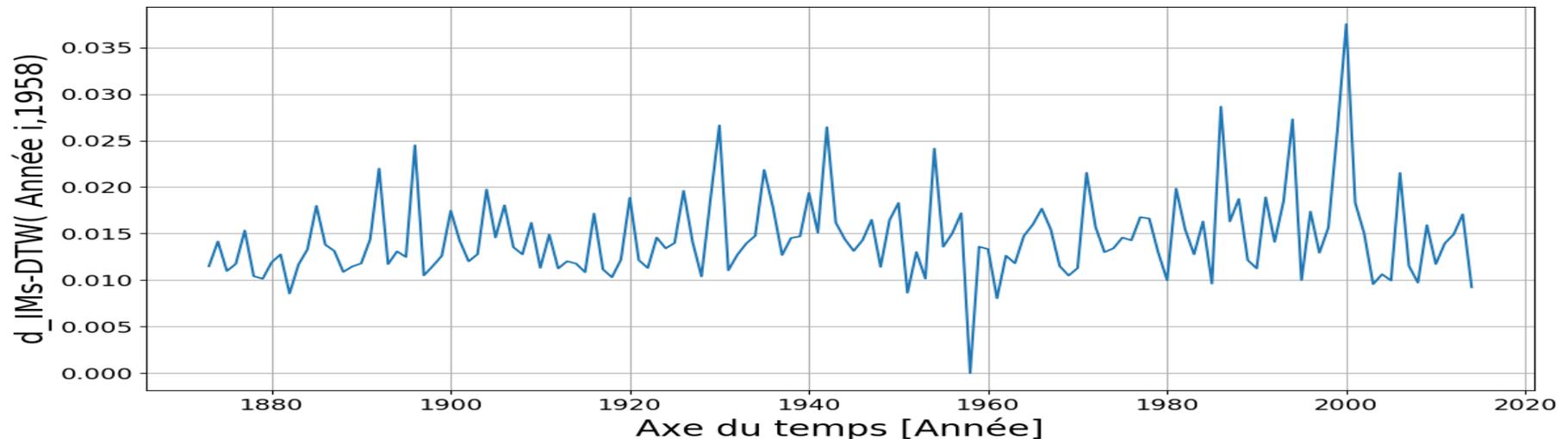
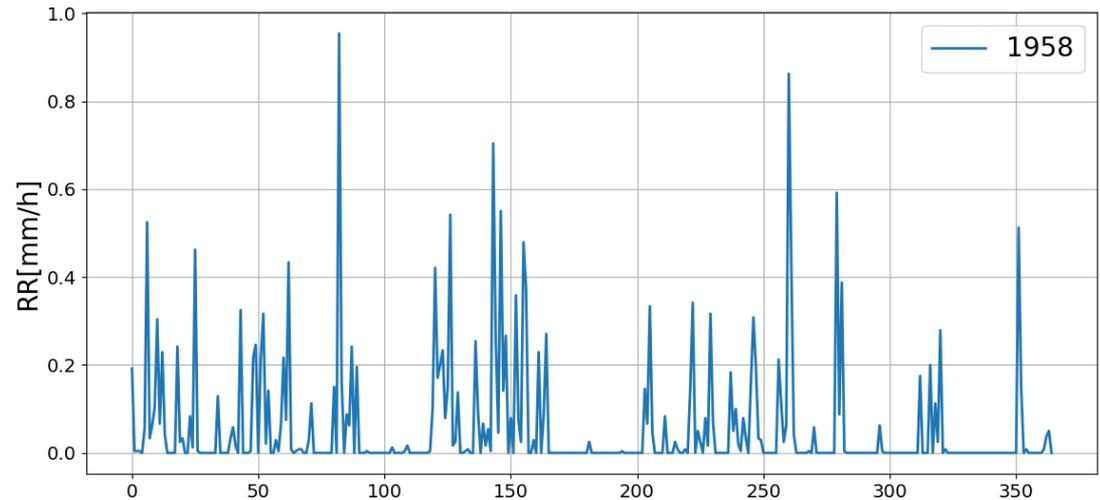


- L'année 2000 (sept.2000 à aout 2001)

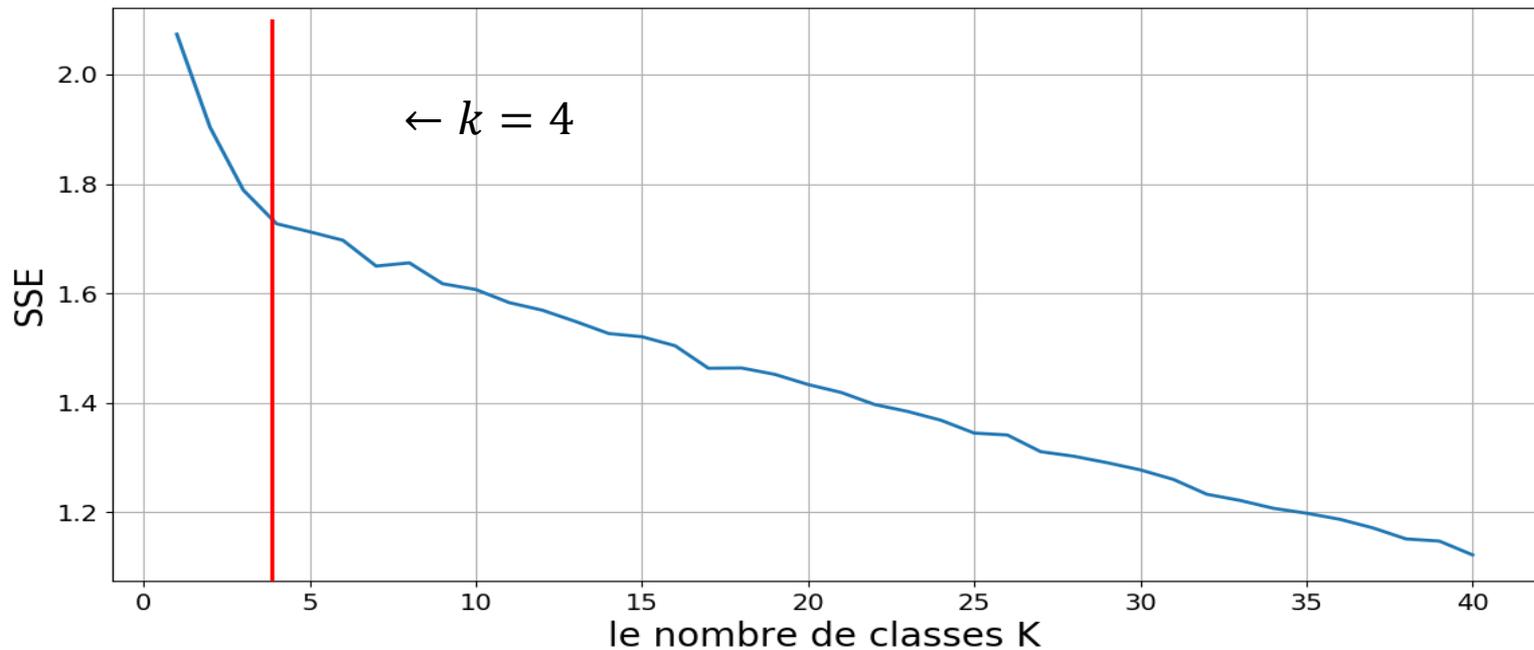
3-3- Variabilité temporelle des précipitations (impact climatique)

Application des K-médoides avec $k=1$ pour trouver l'année représentative

➤ Évolution de la dissimilarité au fil des années

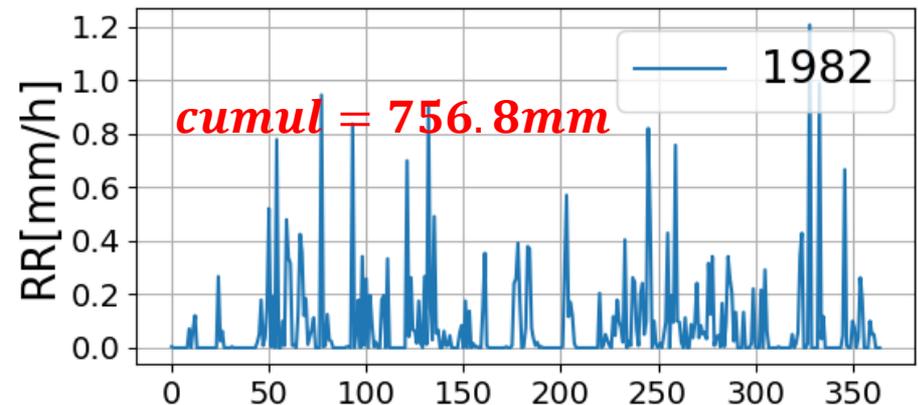
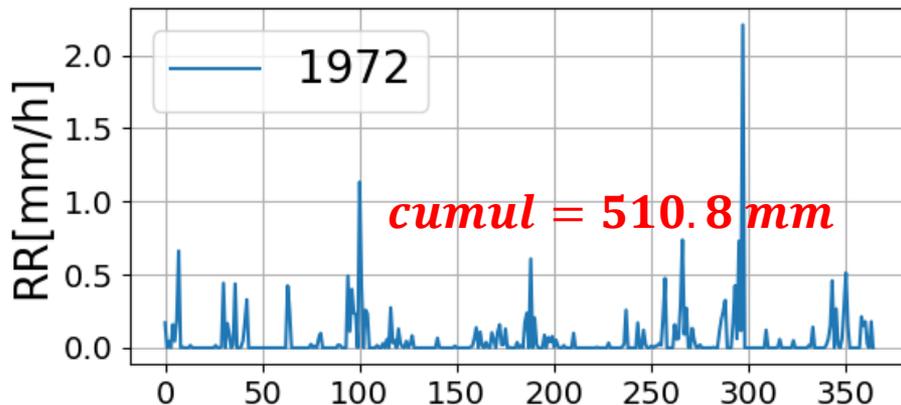
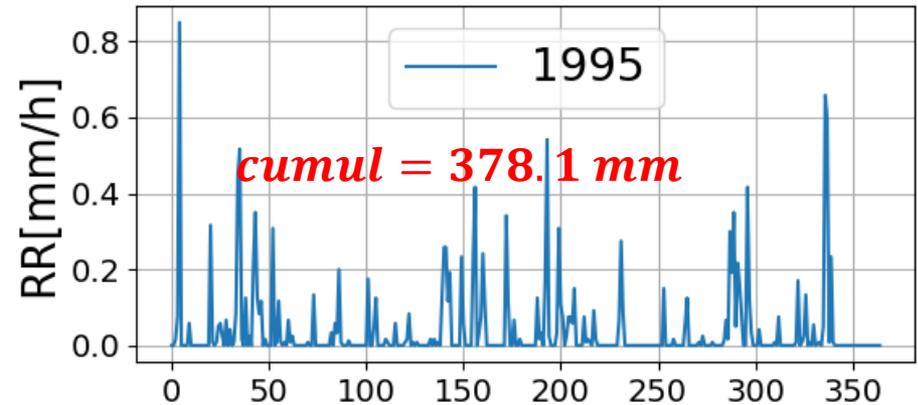
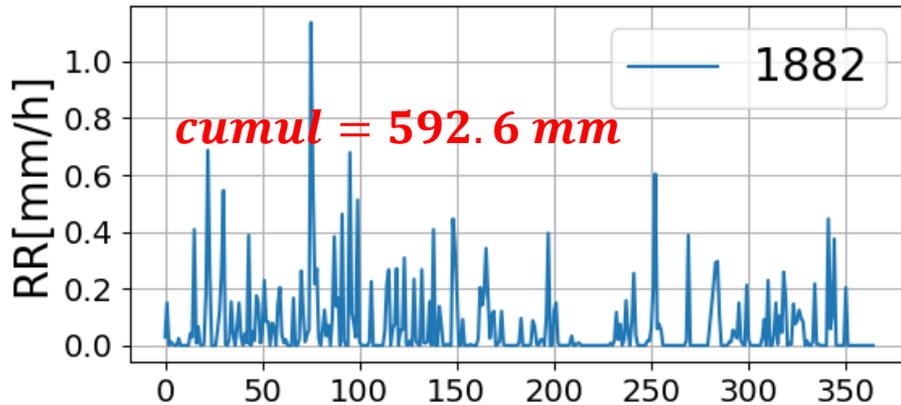


3-3- Variabilité temporelle des précipitations (impact climatique)



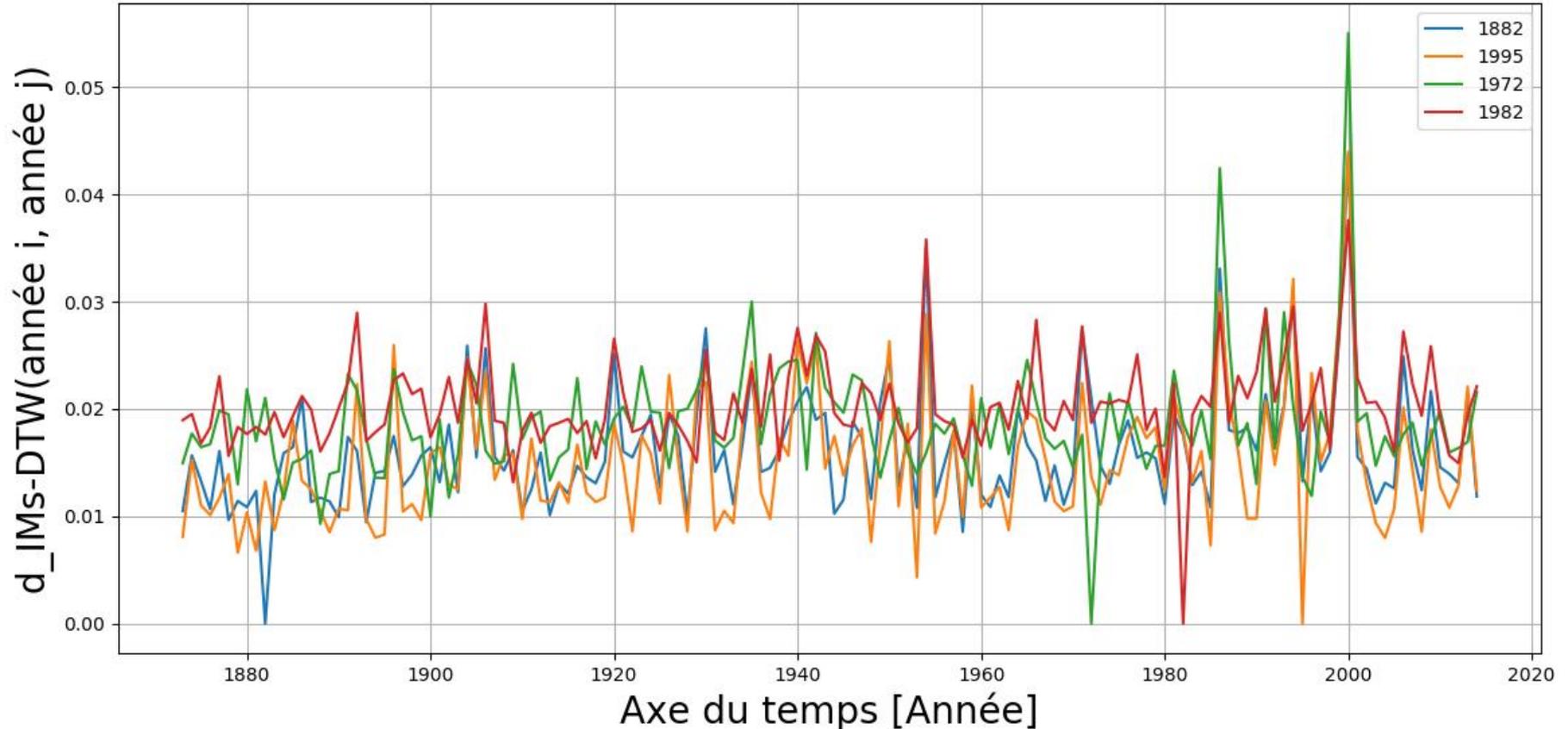
➤ Application des K-médoides avec k=1:40

3-3- Variabilité temporelle des précipitations (impact climatique)



- Les 4 représentants des 4 classes
- L'année 2000-2001 appartient à la 4^{ème} classe avec un cumul de 959 mm et des orages intense l'été

3-3- Variabilité temporelle des précipitations (impact climatique)



➤ Les six années exceptionnelles (1909, 1929, 1938, 1982, 1986, 2000)

Conclusion et perspectives

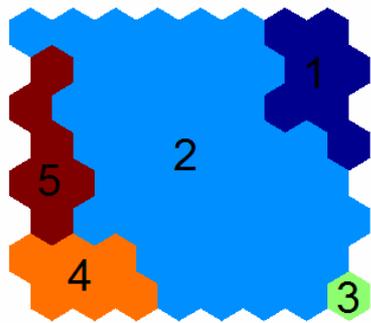
Conclusion :

- L'IMs-DTW est un Outil d'analyse multi-échelle des séries temporelles des précipitations.
- L'approche combinée (IMs-DTW+ K-médoïdes) permet une classification des séries temporelles de pluie sans passer par la caractérisation

Perspectives :

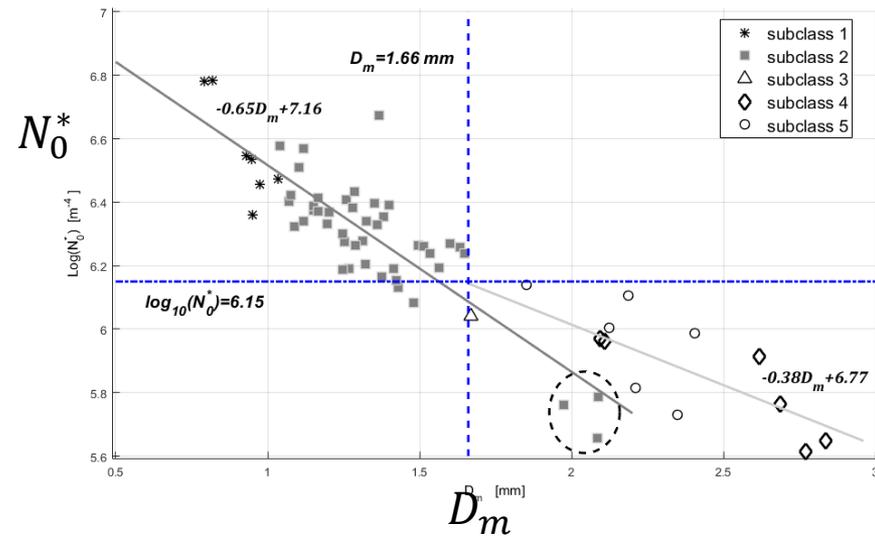
- classification croisée évènements /stations ou années/stations
- interprétation descriptive

1- Caractérisation et classification des événements de précipitations (4)



Classification à l'échelle de la macro-observation

Projection sur la microphysique



Projection des 64 neurones sur l'espace (D_m, N_0^*)

la projection des classes sur l'espace de la microphysique conserve la séparation
⇒ relation d'échelles entre la macro-observation et l'échelle de la microphysique

1-Caractérisation et classification(5)

Transposition de l'étude sur des données de nature différentes

L'étude précédente traitait des données à fine résolution (pas d'échantillonnage d'1 min) issues d'un instrument sensible (DBS),

La configuration des 5 paramètres (durée, écart-type, valeur max, contour vertical, cumul d'eau) $\in \mathbb{R}^5$ permettant une bonne description des événements de pluie est-elle sensible à:

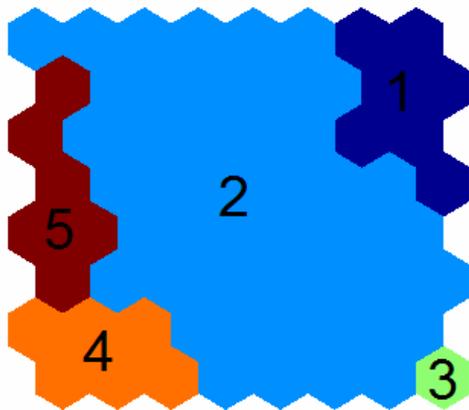
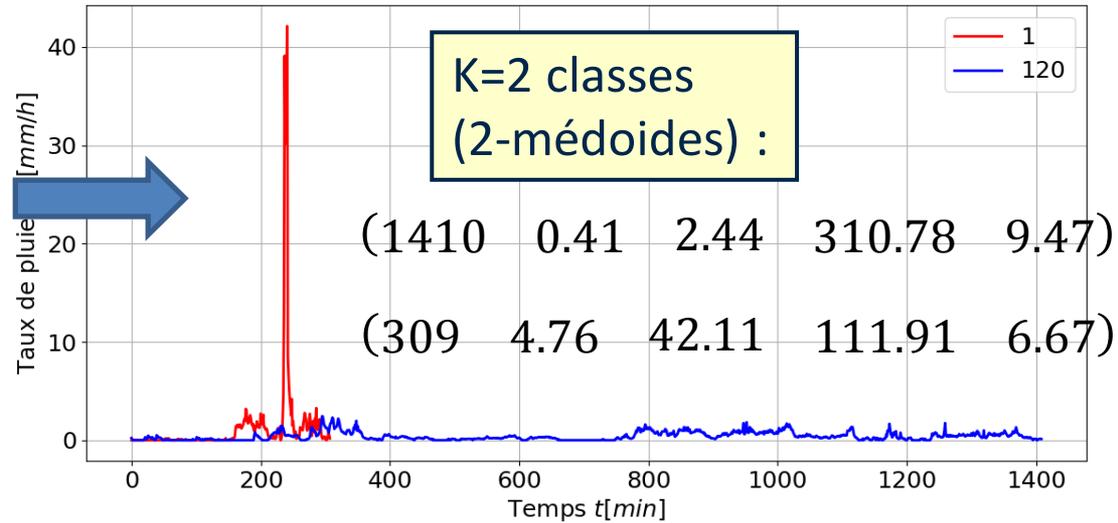
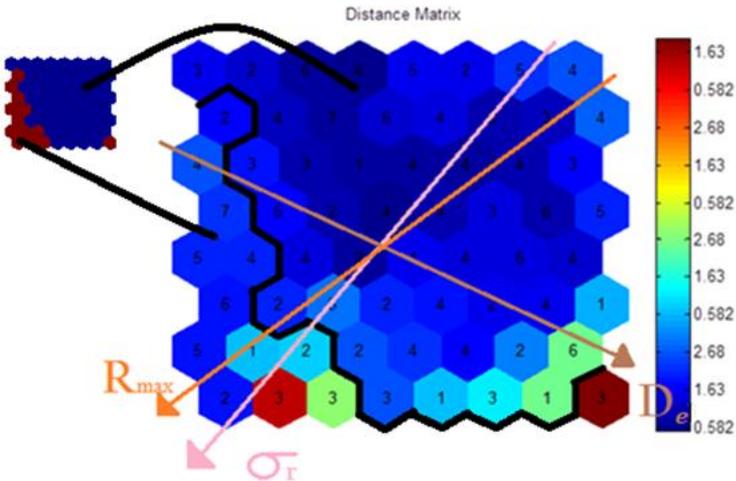
1. **La résolution temporelle** : est-ce que la même description est utilisable à une résolution grossière?
2. **Un changement d'instrument** : quelle sont les performances de cette description sur les données d'un pluviomètre?

cette description était sensible, et d'une façon générale toute description par caractéristiques est sensible à ces deux paramètres

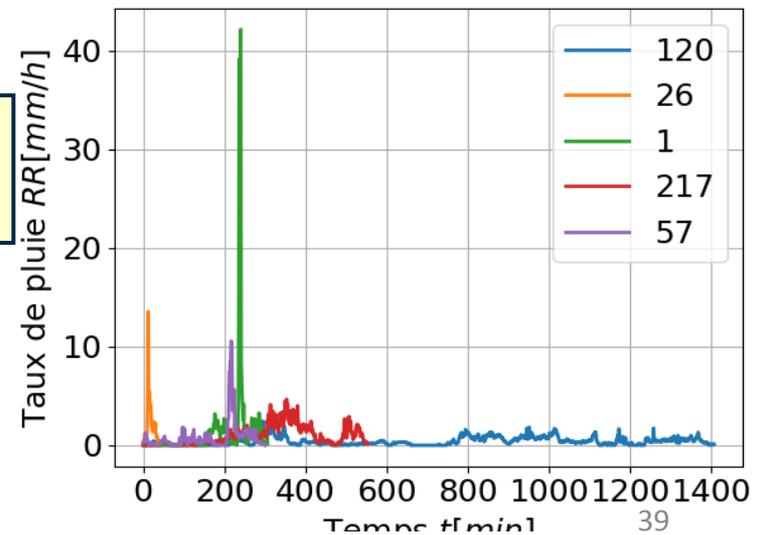
Or, la nécessité d'une description vient du problème d'incomparabilité des séries temporelles de précipitations

=> **trouver un autre moyen de comparer les séries de pluie**

Exemple d'application : classification des 234 événements



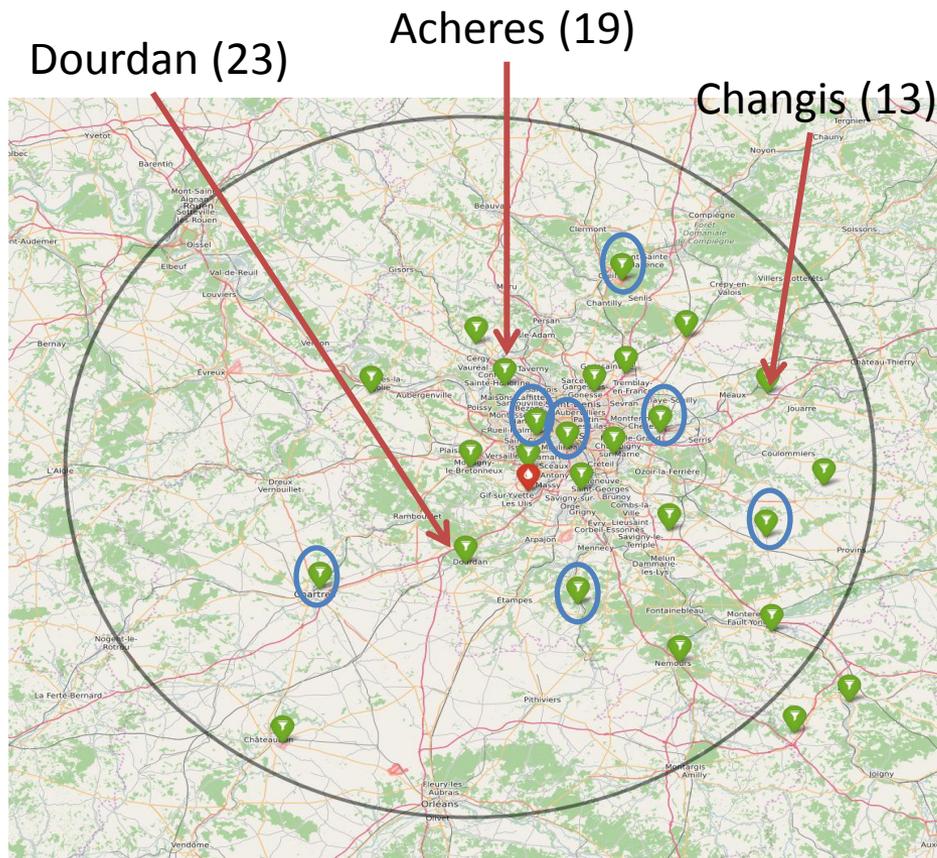
K=5 classes
(5-médoides) :



Le choix du K (nombre de classes) n'est pas discuté car il s'agit d'une comparaison

3-2-Variabilité spatiale des précipitations en ile de France (09 juin 2009)

Classification k-médoides k=1:26

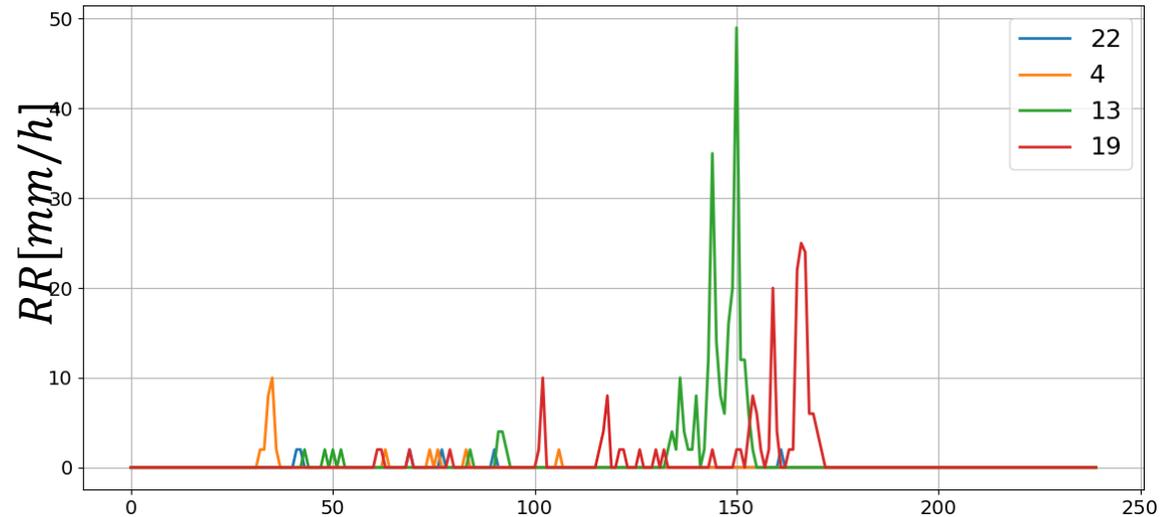
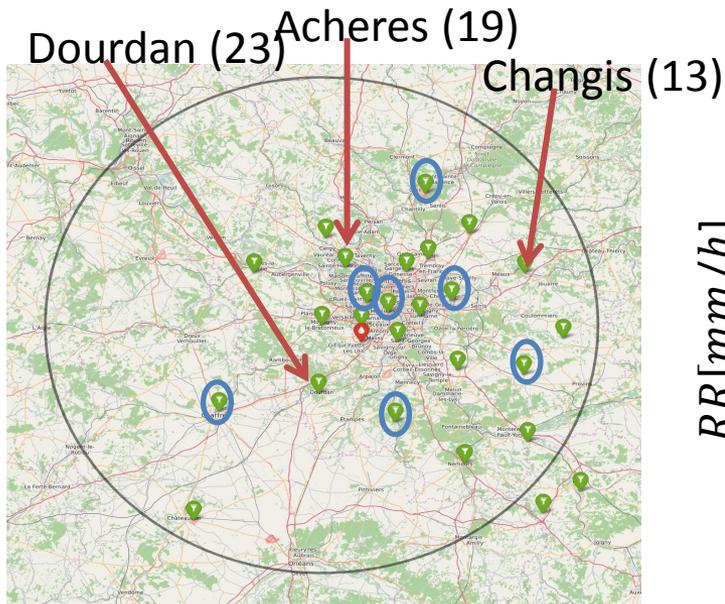


- ⇒ Les stations de Changeris(13), Acheres(19) et Dourdan(23) se démarquent aussi sur la classification et s'isolent rapidement k=3 ([19],[13,23],[le reste])
- ⇒ Les stations en bleu restent groupées $\forall k$, elles enregistrent juste des traces
- ⇒ Pas de conservation de la topologie géographique par la classification
- ⇒ En attente d'une validation sur les cartes radars

Fig 1: les 26 stations de météo-France en Île-de-France + station SIRTa (Rouge)

3-2-Variabilité spatiale des précipitations en ile de France (09 juin 2009)

Classification k-médoides k=4



- Les 4 représentants des 4 classes
- Les stations 13 et 23 sont regroupées