Multiresolution analysis (MRA) classification of plurennial to multi-decadal climate drivers to streamflow in France using Wavelet Transform and Geostatistical Euclidean Distance Clustering

Manuel Fossa[1], Marie Nicolle[1], Nicolas Massei[1], Matthieu Fournier[1], Benoit Laignel[1]

[1]University of Rouen, UMR CNRS 6143 M2C, 76821 Mont-Saint-Aignan, France

Contact : Manuel Fossa (manuel.fossa@etu.univ-rouen.fr)

1 Introduction

It is interesting to try and classify the spatial data to infer the spatial scales associated with certain characteristics within a group of time series. It is indeed very important for operational practices such as flood management and/or climate change adaptation to not only know the severity of a potential impact but also predict the spatial extent of the impact, since similar stations should suffer similar consequences. Classifications on streamflow have already been studied (Hannah et al.(2000; Renard et al.(2006; Wilson et al.(2013) and climate to rainfall or streamflow classification has also been realized by Boé and Terray(2008), Martinez et al.(2008)Martinez and Garavaglia et al.(2010)Garavaglia. Most of those studies establish classification based on annual hydrological regimes. This is in deep contrast with the study that focuses on low frequency time series. There are two axes that the present study aims at improving on. First most of the previous works were carried on raw signals while it is known that different scales of variability (TSV) characterize variabilities of geophysical signals. For instance, atmospheric pressure, precipitation, annual cycle, aquifer contribution or even tidal cycle variability may all be contributing to streamflow variabilities (Labat et al.(2000; Massei and Fournier(2012)). Several studies have highlighted the correlation between some time scales of variability in streamflow or rainfall and the corresponding scales in climatic fields (Massei et al.(2007; Massei et al.(2010; Feliks and Ghil(2010); Fritier et al.(2012; Pinault(2012)). More specifically on streamflow, Boé and Habets(2013) have shown that multi-decadal variability in streamflow in France was impacted by large multi-decadal climate oscillation especially the Atlantic Multi-decadal Variability. Dieppois et al.(2013)Dieppois highlighted the importance of considering the different TSV in assessing the link between climate fields and local temperature and rainfall variability. Relation of hydroclimate time scales to spatial scales has also been recently also investigated by Laepple and Huybers(2014). Those studies highlight the importance of considering each time scale separately. Thus prior to doing a clustering analysis, the input signals must be decomposed into nearly independent components (as Palus(2014) and Jajcay et al.(2016)Jajcay pointed out, pure independence is rarely seen). The second axis is the classification method itself. Classifying fields is always difficult because the extension in two dimensions brings the notion of "general similarity" which is different from "local similarity" i.e two fields may have differences pointwise but an otherwise similar global shape. To deal with it, fields are either represented as index (Boé and Habets(2013); Lopez and Frances(2013)) which necessarily lose a lot of spatial information or a mean of taking into account the global shape has to be found like, for example, the Teweles-Wobus score. However the performance of such score is very variable dependent (it works well on pressure fields but less on other field variables) (Teweles and Wobus(1954)). This study presents a method called Geostatistical Euclidean Distance Clustering (GeoEDC) whose performance, by considering the field as an image, is less variable-dependent.

36 The present study aims at finding out if a classification of climate fields that are correlated with the 6, 10 and 21 year TSV
37 streamflow over France is possible and does so by introducing a clustering method called GeoEDC.

38 **2 Case Study**
39 Data used in this study comes in two types: streamflow data and climate field data. In 2009, IRSTEA (Institut national de
40 Recherche en Sciences et Technologies pour L'Environnement et l'Agriculture) compiled a set of daily streamflow time
41 series spanning 1948-2008, over 200 stations in France. Those stations were based on their small influence by anthropogenic
42 activities (they are said to be "climate sensitive"). For this study, 152 of those stations were chosen (those which presented
43 continuous data from 1968 to 2008) and the sample rate was converted to monthly. The spatial repartition of stations is
44 shown in Figure 1. Areas without stations are known to be heavily influenced by human activities especially the Northern
45 part, were one third of the total France population lives.

46 Climate data comes from NCEP/NCAR reanalysis data set. Four climate variables were selected on the basis of their known
47 role in the water cycle of Europe: geopotential height, air humidity, zonal and meridional wind. The area selected spans -
48 77/40E,-5/78N and from 1968 to 2008, at monthly rate. This area was selected based on the known influence of both
49 Atlantic and African climate fields on precipitation and streamflow over France. When applicable, geopotential at level
50 850mb was chosen which approximately ensures a close to surface measurement while preventing relief thresholding in most
51 of the area (except in the Alps for instance).

52 It is commonly reported in the literature that links between large-scale circulation and local variable such as precipitation or
53 streamflow are clearer in winter in the Euro-Atlantic sector. Thus, only winter months are presented here.

54 **3 Methods**
55 The basic workflow can be seen in Figure 2. Both streamflow and climate field data are first decomposed using
56 multiresolution analysis. The climate field of interest is associated with streamflow via a point-wise correlation analysis
57 giving a so-called correlation climate field (CCF), i.e. map of correlation between each climate variable and a streamflow
58 station. Once the correlation fields are obtained, a Geostatistical Euclidean Distance Clustering is applied. The analysis
59 produces both cluster maps (stations that share the same climate drivers) and sets of CCFs inside each cluster. We first assess
60 the existence of a classification according to a time scale of variability (TSV) and climate driver then look at the shape of the
61 representative climate drivers in each cluster.

62 Wavelet transform are used to decompose the data into different TSVs.

63 The wavelet transform theory has been and is still extensively studied. Detailed explanation of the theory dedicated to
64 hydrologists can be found in Labat et al.(2000)Labat. A signal can be represented in a Hilbert space, with a complete
65 orthonormal system basis generated by an orthonormal wavelet. The basis is formed by a family of wavelet functions that are
66 all identical, except that each is dilated/contracted, translated version of the others.

67 The signal can thus be represented as follows:

68 $x(t) = \sum_{j,k=-\infty}^{\infty} c_{j,k} \psi_{j,k}(t)$ (1)

69 This means that the value at time t of the variable x from the time series, $x(t)$ can be perfectly represented the sum at
70 different scales j and position k of wavelet function $\psi_{j,k}$ and a wavelet coefficient $c_{j,k}$.

71 Equation 1 states that the signal can be both represented on this orthonormal basis (wavelet analysis), and perfectly
72 reconstructed from its Hilbert representation (wavelet synthesis).

73 The wavelet functions families that form the orthonormal basis of the Hilbert space are defined as:

74 $$\psi_{j,k}(t) = 2^{\frac{j}{2}}\psi(2^j t - k) \tag{2}$$

75 As equation 2 states, the wavelet at a given scale and position is a scaled and translated version of the same wavelet from the
76 next/previous scale. One can see that wavelet transforms are by nature multiresolution transforms like Fourier transforms but
77 they differ in that the wavelet is not periodic.

78 This allows the wavelet transform to act as a time-frequency transform, providing not only information on the time scales
79 (that can also be approximated as Fourier frequencies), but also on the temporal evolution of those scales. $c_{j,k}$, the wavelet
80 coefficients, express how the signal varies from one position to another, those position depending on the scale (Precise
81 interpretation of wavelet coefficients depends on the wavelet function family chosen and can sometimes be very tedious).
82 The choice of the wavelet function family is important. This choice must be based on three parameters: the function itself
83 (which conditions the interpretation of wavelet coefficients and the reconstruction process), the size of its support (the size
84 where the wavelet value is non zero), and the number of its vanishing moments. The choice will mainly be a compromise
85 between frequency and time resolution. A wavelet with a short support will have a very good time resolution (being able to
86 detect singularities) but a bad frequency resolution and vice versa. A number of wavelet functions have been developed and
87 more information can be found in the reference previously cited.

88 Using equation 1, wavelet transforms can be used mainly in two ways: data analysis and data reduction. The former involves
89 studying wavelet coefficients to gather information about the variation of the data at different scales and times (equation 1
90 solved for $c_{j,k}$), while the latter mainly involves decomposing the signal into a set of orthonormal components (at chosen
91 scales) and retaining only the most significant ones (in term of variability). Example of the first use is time series analysis
92 while image/video compression is a common application of the second.

93 The wavelet method used in this study is the multiresolution analysis (MRA) with the aim of decomposing both streamflow
94 and climate fields into components of increasing time scale of variability (TSV). Detailed theory explaining the multi
95 resolution nature of wavelets can be found in Mallat(1989) and one of its most famous applications for computers in Percival
96 and Walden(2000). The MRA analysis is based on equation 1, but the scales and positions are chosen so as to decrease the
97 number of needed information as much as possible. The scales chosen are dyadic (i.e: power of 2), and positions are chosen
98 so that the number of positions where the coefficients are calculated is decreased by 2 at each scale. This peculiar MRA
99 technique is a discrete wavelet transform (DWT) process that satisfies the orthogonality of the basis both in translation and
100 scale. It is also possible to use an undecimated discrete wavelet transform (MODWT) where the number of coefficients isn't
101 down sampled at each scale. This kind of MRA loses orthogonality in scale, but allows for shift invariance of the
102 coefficients. For this study the DWT was chosen as the MRA technique. The wavelet chosen for the MRA is a D4
103 (Daubechies 4 wavelet), which offers a good frequency-time resolution compromise. Both streamflow and climate data are
104 decomposed into 9 components according to the following time scales: 3 Months, 7 months, Annual, 1.5 years, 4 years, 6
105 years, 10 years and 21 years. The last component is composed of the rest of the signal (usually contains the mean). For this
106 study, only the 6, 10 and 21 year components are kept.

107 Correlation Analysis between a streamflow time series and a climate field is done pointwise.

108 Once the correlation maps are obtained, a Geostatistical Euclidean Distance Clustering (GEDC) is undertaken to see which
109 stations have similar climate fields. GEDC is a method adapted from the IMage Euclidean Distance (IMED) proposed by
110 Wang et al.(2005)Wang who designed a Euclidean distance algorithm for images taking into account the distance between
111 the pixels. The idea is that pixels that are close do not exhibit strong differences in gray levels and that when it happens it is
112 due to a deformation. Thus the method gives importance to differences of gray levels that are spatially far enough from each
113 other. The aim there is to be able to recognize global patterns that are similar even if locally small distortions exist. Such a
114 method is thus adapted to the clustering of climate fields whose impact on streamflow depends both on actual values
115 (pressure, wind strength etc), but also the shape of the fields. For example it is well known that the pressure field pattern
116 changes in the North Atlantic often identified as "NAO+/-, Atlantic Ridge or Scandinavian Blocking weather regime", will
117 dictate a large part of the climate variability over Western Europe. The IMED method was thus modified and adapted in this
118 work into the GeoEDC method. The difference lies into the calculation of the distance between pixels which now becomes a
119 distance between coordinates on the map and thus takes into account Earth curvature.

120 The basic steps are as following:

121 The Euclidean distance between two correlation maps is calculated as in:

122 $$DT(i,j) = ((\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{G}(\boldsymbol{x} - \boldsymbol{y}) \qquad (3)$$

123 $DT(i,j)$, the distance between correlation map of station i and correlation map of station j can be calculated by the outer
124 product of the two correlation maps $\boldsymbol{x}, \boldsymbol{y}$ and the matrix of weights $\boldsymbol{G}$:

125

$$\boldsymbol{G} = \frac{1}{2\pi\sigma^2} exp(-\frac{\left(d_{ij}^s\right)^2}{2\sigma^2})$$

126 d_{ij}^s is the matrix of distance between each coordinates. $\boldsymbol{G}$ can be seen as a Gaussian ponderation of the variable (e.g.
127 geopotentials correlation) difference between two coordinates based on their spatial distance. Instead of calculating the
128 spatial distance using trigonometric functions the coordinates are translated into n-vectors. This prevents calculation errors
129 when angles approach certain values.

130 A n-vector is a 3 dimensional vector defined as:

$$n_e = \begin{array}{c} \cos(latitude) * \cos(longitude) \\ \cos(latitude) * \sin(longitude) \\ \sin(latitude) \end{array}$$

131

132 When used on a computer, calculation of the spatial distance between two n-vectors is defined as:

$$D_{ab} = R * \text{atan2}(|\boldsymbol{n_a} \times \boldsymbol{n_b}|, \boldsymbol{n_a}.\boldsymbol{n_b})$$

133 The spatial distance between a n-vector a and n-vector b, D_{ab}, is thus the product of radius of Earth R and the angle between
134 the two coordinates represented by the n-vectors, $atan2(|\boldsymbol{n_a} \times \boldsymbol{n_b}|, \boldsymbol{n_a}.\boldsymbol{n_b})$.

135   To sum up, due to matrix G, neighbor's coordinates will see their variable difference decreased and far away coordinates will
136   have it untouched. Extremely far away coordinates are also filtered.

137   The GeoEDC gives a distance matrix between correlation map, and an ascendant hierarchical clustering (AHC) is applied on
138   the distance matrix using the complete metric distance. A silhouette optimization algorithm gives the best number of clusters
139   for each climate driver by minimizing the inverse of the sum of the mean total silhouette width and the intra silhouette
140   standard deviation.

141

142   4 Results

143   4.1 Cluster maps of streamflow based on their climate drivers
144   Figure 3 shows the results of the GeoEDC applied to the correlation climate fields for the four climate drivers and for the
145   three TSVs.

146   Clear spatial patterns are visible for all climate drivers as well as for all TSV even when the number of clusters is only two.
147   Even at the highest TSV clear spatial repartition shows up.

148   Inter climate driver comparison shows that clusters are not necessarily the same depending on the climate driver. This
149   highlights the complexity of climate influence, even for relatively close stations. Nevertheless as the TSV increases the inter
150   climate driver variability decreases, that is more stations share the same climate drivers.

151   Inter TSV comparison indicates that for a given climate driver the clustering changes quite significantly according to time
152   scale. However one can see that as TSV increases the clusters shape and repartition are more and more dipolar (often with
153   north/south shape and repartition).

154   Spatially, the Mediterranean area is often a singularity even at high TSV.

155   4.2 Climate Correlation Fields
156   Figure 4 represents the average correlation fields corresponding to each cluster, for each climate driver, and each TSV. The
157   GeoEDC method first distinguishes the general shape of the climate correlation field (CCF). If two are very similar then
158   discrimination is made on the actual correlation values.

159   The results can be broken down into two components. First the intra TSV comparison, for one given TSV, assesses the
160   difference between clusters. The second one is an inter TSV comparison. The latter answers the question "Are CCFs
161   different depending on the TSV?". Indeed, it must be stressed that a certain cluster at one TSV, say cluster 1, do not contain
162   the same stations at another TSV so this latter way of tackling the results cannot be called "intra-cluster". It will be referred
163   as inter-TSV.

164   Intra TSV results show that climate driver clustering is sometimes based on shapes i.e for a given TSV, clusters CCF differ
165   on their very shape, and sometimes only correlation values are distinct, which is the case as one go to higher TSV.

166   Inter TSV results show significant variability in CCF depending on the TSV. Geopotential and wind show very different
167   structures for each of their TSV.

168 Of course fields themselves are of particular interest. At the 6 year TSV, geopotential is organized with clear pressure dipole
169 in the Euro-Atlanic zone with a negative correlation zone over England and a positive correlation zone, located along an arc
170 starting from Western Africa coast to South Mediterranean basin for stations of Central to Eastern France (Figure 5a).
171 Interestingly, for the Mediterranean stations (cluster 2) streamflow is not correlated with Mediterranean geopotential (Figure
172 5a). Correlation values are not very high with values up to ±0.6. Zonal Wind patterns clearly show Westward reinforcing
173 influence with a positive correlation zone over the North Atlantic. Differences between clusters result from the latitudinal
174 localization of the high positive correlation band (Figure 5b). For example southern stations are mainly correlated with
175 southernmost westerlies. Meridional wind shows evolution towards two vortices identified by their opposing poles, one
176 corresponds to a counterclockwise vortex over England, the other to a clockwise vortex over western coast of Africa which
177 is in line with the geopotential CCFS previously described (Figure 5c). Correlation values for winds are high, with some
178 areas having higher than ±0.7 correlation values. Air humidity show a local pattern with high correlation over France for all
179 but the southern stations where streamflow is positively correlated with air humidity over the Eastern North Atlantic (Figure
180 5d). Correlation values are high with values up to ±0.7.

181 At the 10 year TSV, geopotential shows a clear dipole correlation pattern with a strong negative band from the Eastern US
182 coast to England, and a strong positive correlation band from Eastern Canada to Scandinavia. The difference between
183 clusters is the presence or absence of a positive correlation area over the African continent (Figure 5a). Correlation values
184 are high up to ±0.8. Zonal wind shows a similar arrangement with the decrease in the western wind from North America
185 Eastern Coast to England, and a positive correlation zone at the south of that band. Difference between clusters seems to be
186 the presence or absence of vortex like systems with some clusters more uniform along the band (Figure 5b). Indeed
187 Meridional wind show vortex-like circulations on the eastern coast of North America and over Northern Europe (Figure 5c).
188 Correlation values are high for winds, going up to ±0.9. Air humidity shows positive correlation value at the location of the
189 wind and geopotential vortex-like systems. Cluster differences are based on the strength of a positive correlation band going
190 from England to the Middle-East (Figure 5d). Correlation values are high (±0.9).

191 At the 21 year TSV, Geopotential shows a structure reminiscent of the 10 year band with the band of negative correlation
192 over the Eastern US coast to England, and a positive band over Eastern Canada to Scandinavia. The difference between
193 clusters is the general level of correlation. While most stations are highly (anti)correlated with geopotential, the
194 Mediterranean and South Britany stations are not (Figure 5a). Both winds CCFs are also very close to their 10 year TSV
195 counterpart and they are both having high correlation values. Note that the number of clusters has gone from 3 in the 10 year
196 TSV, to 2 in the 21 year TSV (Figure 5b and 5c). Air humidity on the other hand shows a different shape from its previous
197 TSV with a generalization of positive correlation zones over the Northern Atlantic (Figure 5d). High correlation values are
198 still present.

199

200

5 Discussion

5.1 About GeoEDC clustering

The GeoEDC method is not the first to focus on the shape of fields rather than strict values of the variables, however the
method differs from those of Boé and Terray(2008), Martinez et al.(2008)Martinez or Garavaglia et al.(2010)Garavaglia in
that GeoEDC is essentially a pattern recognition clustering technique. The former are focused on pressure fields, and look at

their gradient because it translates into the notion of wind patterns (gradient of pressure effectively being the main driver with Coriolis force behind wind direction). It follows that such technique work well with pressure fields, but may not be adaptable to humidity fields for example. By treating fields as images, the GeoEDC method seems quite suitable for any type of field where global shape is more important that the mean values. The real strength of GeoEDC lies in its ability to account for local deformation and thus concentrate on global shapes. When those are comparable between stations, the method performs classification on mean values as is it still a Euclidean Distance based method.

One important aspect of the GeoEDC method is its capacity to produce smooth clustering i.e. making smooth transitions between stations belonging to different clusters but close spatially. Here we will verify this property by the following two step approach: The first one only consists in comparing CCF at stations located at the interface between two clusters, to see if they share similar CCFs. The second step is more complex, and requires finding the point at the center of the cluster, based on the rational that closest stations to this point are also the farthest from the interfaces and thus should have the most representative characteristics of the cluster's CCF. Finding a center for a cluster is easy; however this center is only relevant if the cluster is physically bounded by other clusters. If the boundaries aren't made of other clusters (for example, the cluster is limited on one of its side by sea or if the cluster has sides on the borders of the stations coverage) then the cluster is not totally representative of the CCF, and the real physical center may lie elsewhere. This is the case for the cluster maps produced for this study. The clusters all have an artificial boundary and thus are never completely bounded by other clusters. Finding the geometric center of the clusters as they are may be of some use however. Computing it and comparing stations close to that center with stations on interfaces will give information about the real spatial extent of the cluster. If stations close to that center are very different from those at interfaces, this means that the real extent of the cluster is close to the one provided by the study, otherwise this indicates that one should look at stations close to artificial boundaries as they may provide very different CCFs.

Figure 5 illustrates the latter point. The spatial structure of CCFs on the Mediterranean part of the cluster 2 (black points) of geopotential for the 6 year TSV is shown. The cluster has North and West boundaries with another cluster, but the South and East boundaries are artificial (respectively coastline and France/Italy border). Analyzing the stations at the interface between clusters shows that indeed the GeoEDC provides smooth transitions as stations from two different clusters that are the closest do produce similar CCFs (Arrows on Figure 5 point at stations from different clusters that are close together). The geometrical center of the cluster was calculated using a convex set minimization problem with linear constrains taking into account the Earth curvature. This problem can be explained as finding the smallest spherical cap that contains all the stations of the cluster, and then finding the center line of the spherical cap. This is equivalent to finding the smallest vector from the center of Earth to the base plane of the spherical cap. One can see that the stations at the interface seem to follow a counter clockwise differencing scheme. Starting from the Southeast most stations, going counter clockwise, CCFs change gradually towards a Scandinavia-Equatorial/Mid latitudes pressure dipole. The stations that are closest to the geometric center actually take the CCF of the northern stations even those on interfaces. Since those stations are not different from stations at the interfaces, this means the real geometric center of the cluster lies probably much further south (possibly over the Mediterranean Sea).

5.2 Cluster maps

The results show that for one TSV, clusters are sometimes different depending on the climate drivers. The exploration of the physical processes behind this repartition is beyond the scope of this study, and should be treated in future work. However one thing that appears in the results is that with increasing TSV, the clustering is more and more the same between climate drivers especially between geopotential and wind. Indeed, already from the 10 year TSV, the clusters are all along the same lines, only the number of clusters differs. This translates into stations sharing more and more of the same climate drivers. This suggests that the spatial scales of climate drivers correlated with stations increase with increasing TSV. This is a

248 commonly accepted hypothesis that large TSV are correlated with large spatial scales. This is however something that would
249 have to be verified in future work. One interesting detail is that for each TSV, geopotential always displays the smallest
250 number of clusters, and the most homogeneous spatial patterns. This suggests that the causal chain of climate drivers
251 correlated with streamflow in France is first geopotential, then winds and humidity. This is what was highlighted in the
252 results section i.e. winds and air humidity clusters differ in either latitude or longitude, while the geopotential CCF are more
253 similar and have a far more homogeneous structure.

254 Looking at inter-TSV difference in clustering, the same remark can be made. With increasing TSV, the clusters align more
255 and more along a latitudinal repartition with Southern, Central and Northern France apart. The number of clusters decreases
256 for each climate driver as the TSV increases. This again suggests that spatial scales increase with TSV.

5.3 Climate Correlation fields

258 On a general note, it is important to understand one nuance that comes from wavelet analysis. Components of a wavelet
259 transform are subtracted not from the mean but from the original signal. It means that the CCF should be understood not as
260 e.g. "streamflow increases when winds are westward", but as "streamflow increases when the winds tend to be accentuated
261 westward" (an eastward wind becoming less eastward falls in this last category). Thus, the correlation field map should be
262 treated as correlation between streamflow and climate driver variations of variation (Correlation analysis treats with
263 variations, but wavelet analysis do so too hence the "variations of variation"). As an example, the zonal wind CCF at 10 and
264 21 year TSVs doesn't show easterly winds at the North (negative correlation area), and westerly at the south of the North
265 Atlantic. Similarly, they do not show that vortices develop on eastern US coast and over north England. Rather, they show
266 that streamflow reacts to some dynamics that tend towards those structures. The total (the actual field of wind values) field
267 for this TSV may still be different as shown in Figure 6. One can see that the field is in fact mostly made of westerlies on the
268 Northern part of the North Atlantic, while the inverse is true for the Southern part.

269 Establishing the climate fields that drive either precipitation or streamflow over France has already been achieved, for
270 example by Boé and Terray(2008) and Garavaglia et al.(2010)Garavaglia focusing on spatial extent like this study. However
271 other studies have been realized focusing more on the temporal evolution of climate forcing at some TSV. For example,
272 Sutton and Hodson(2005), Sutton and Dong(2012) inspected the role of multi-decadal variability in SST to explain changes
273 in Europe climate. In particular Boé and Habets(2013) studied the multi-decadal streamflow variability in France for a
274 number of rivers, and studied their link with large scale fields such as sea level pressure and Atlantic Multidecadal
275 Variability. Their work concentrated on spring months but they noted that in winter (the months chosen for this study),
276 multidecadal variabilities were not synchronously connected with local climate variation (e.g temperature, precipitation).The
277 authors found that when correlation existed, streamflow were correlated with negative pressure fields over continental
278 Europe, and positive pressure fields over central Atlantic. The results of this study are consistent with those from the authors
279 in that the same field at that scale was found. However, the present study found good correlation between geopotential and
280 streamflow during winter months in contrary to the previous authors. It should be noted that Boé and Habets(2013) used an
281 index, while "original" fields were used in this study. This may explain the differences in correlation results. In addition, it
282 should be noted that stations from the second cluster at the 21 year scale are far less correlated that the one in the first
283 cluster. Those stations almost all lie in the Western Mediterranean area.

284 Finally, the spatial to temporal scale relationship is really interesting. As noted previously, station clustering in France tends
285 to get less complex as TSV increases both intra-climate driver and inter-TSV. This is confirmed by CCF maps which show
286 that the patterns at the higher TSV become more "global", and the difference between clusters more due to differences in
287 correlation values instead of field shape (weather regimes tend to be similar). However a close look highlights areas where
288 changes of shape exist. For instance in zonal wind 21 year CCF (Fig 5b), difference between clusters is indeed due to the

289 main dipole value differences, but some distinct differences in shape are visible on the Eastern coast of the South African
290 continent. Similarly, the Western Siberia seems to be under westerlies reinforcement area in the first cluster but completely
291 out in the second. The same can be said of air humidity 21 year TSV CCF (Figure 5d), where Western Mediterranean basin
292 is either under a highly positive correlation area or not at all. This leads to think that several spatial scales actually coexist
293 within one TSV. Future work is needed to investigate this temporal-spatial scale relationship.

294

6 Conclusion

296 This study aimed at assessing whether a regional classification of streamflow variability in France based on large-scale
297 climate drivers was possible according to interannual to interdecadal time-scales. The study covers on 152 climate sensitive
298 streamflow watersheds over France. Our work focused on three time scales of variability (TSV): 6, 10 and 21 years. In order
299 to test this hypothesis, a clustering by hierarchical ascendant classification was applied to 4 types of correlation maps
300 between climate fields and streamflow time series in France: Geopotential, zonal and meridional wind and air humidity (all
301 at 850mb). To this end, we developed a method called Geo Euclidean Distance (GeoEDC), adapted from Image Euclidean
302 Distance first proposed by Wang et al.(2005)Wang for image processing. This method computes similarity between fields
303 based on their global shapes before looking at mean values. By treating field as images, GeoEDC works on more types of
304 fields that methods previously developed. The results of the clustering show distinct spatial patterns for every climate drivers
305 and for all time scales of variability considered. The general trend is that at the lowest TSV (6 year) stations that share one
306 climate driver do not necessarily share the others. As the TSV increases, more stations share the same climate drivers.
307 Results also show that for a given climate driver, the clustering of stations differs depending on the TSV. As TSV increases
308 the clustering becomes simpler, with less clusters for a given climate driver. Correlation climate fields (CCF) associated with
309 streamflow in France are consistent with those found in the literature but that study brings up more information by
310 considering different TSV as well as considering true fields and not indices. The clustering also suggests hypothesis on the
311 order of causality of climate drivers in the forcing chain. The results suggest that, as expected, geopotential are the source for
312 most of the variation of streamflow in France with wind and air humidity following, even considering the large TSV
313 involved. Future work focusing on the time-space evolution characteristics of the CCFs, exploration of the large-scale and
314 watershed physics explaining the clustering maps obtained would be needed in order to better understand how climate fields
315 influence streamflow in France.

316

References

318 Boé, J. and Habets, F.: Multi-decadal river flow variations in France, Hydrological and Earth Systems Science, 18, 691–708,
319 2013.

320 Boé, J. and Terray, L.: A Weather-Type Approach to Analyzing Winter Precipitation in France: Twentieth-Century Trends
321 and the Role of Anthropogenic Forcing, Journal of Climate, 21, 3118–3133, 2008.

322 Dieppois, B., Durand, A., Fournier, M., and Massei, N.: Links between multidecadal and interdecadal climatic oscillationsin
323 the North Atlantic and regional climate variability ofnorthern France and England since the 17th century, Journal of
324 Geophysical Research: Atmospheres, 118, 4359–4372, 2013.

325 Feliks, Y. and Ghil, M.: The Atmospheric Circulation over the North Atlantic as Induced by the SST Field, Journal of
326 Climate, 24, 522–542, 2010.

327 Fritier, N., Massei, N., Laignel, B., Durand, A., Dieppois, B., and Deloffre, J.: Links between NAO fluctuations and inter-
328 annual variability of winter months precipitation in the Seine River watershed (north-western France, C. R. Geoscience, 344,
329 396–405, 2012.

330 Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garcon, R., Don, R., and Benardara, P.: Introducing a rainfall compound
331 distribution model based on weather patterns sub-sampling, Hydrol. Earth Syst. Sci., 14, 951–964, doi: 10.5194, 2010.

332 Hannah, D., Barnaby, S., Gurnell, A., and Mcgregor, G.: An approach to hydrograph classification, Hydrol. Process., 14,
333 317–338, 2000.

334 Jajcay, N., Hlinka, J., Kravtsov, S., Tsonis, A., and Palus, M.: Time scales of the European surface air temperature
335 variability: The role of the 7–8 year cycle, Geophysical Research Letters, 43, 902–909, 2016.

336 Labat, D., Abadou, R., and Mangin, A.: Rainfall–runoff relations for karstic springs., Journal of Hy, 238, 2000.

337 Laepple, T. and Huybers, P.: Global and regional variability in marine surface temperatures, Geophysical Research Letters,
338 2014.

339 Lopez, J. and Frances, F.: Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir
340 indices as external covariates, HESSD, 10, 3103–3142, 2013.

341 Mallat, S.: A Theory for Multiresolution Signal Decomposition:The Wavelet Representation, IEEE Transaction on Pattern
342 Analysis and Machine Intelligence, 11, 674–693, 1989.

343 Martinez, C., Campins, J., Jansa, A., and Genoves, A.: Heavy rain events in the Western Mediterranean: an atmospheric
344 pattern classification, Adv. Sci. Res, 2, 61–64, 2008.

345 Massei, N. and Fournier, M.: Assessing the expression of large-scale climatic fluctuations in the hydrological variability of
346 daily Seine river flow (France) between 1950 and 2008 using Hilbert Huang Transform, Journal of Hydrology, 448-449,
347 119–128, 2012.

348 Massei, N., Laignel, B., Deloffre, J., Dupont, J., Durand, A., and Valdes, D.: Investigating possible links between the North
349 Atlantic Oscillation and rainfall variability in northwestern France over the past 35 years, Journal of Geophysical research,
350 112, D09121, doi: 10.1029/2005JD007000, 2007.

351 Massei, N., Laignel, B., Deloffre, J., Mesquita, J., Motelay, A., Lafite, R., and Durand, A.: Long term hydrological changes
352 of the Seine River flow (France) and their relation to the NorthAtlantic Oscillation over the period 1950-2008, Int.
353 J.Climatol., 30, 2146–2154, 2010.

354 Palus, M.: Multiscale atmospheric dynamics:Cross-frequency phase–amplitude coupling in the air temperature, Physical
355 Research Letters, 112, 2014.

356 Percival, D. and Walden, A.: Wavelet Methods for Time Series Analysis, Cambri, 2000.

357   Pinault, J.: Global warming and rainfall oscillation in the 5–10 yr band in Western Europe and Eastern North America,
358   Climate Change, 114, 621–6, 2012.

359   Renard, B., Lang, M., and Bois, P.: Statistical analysis of extreme events in a non-stationary context via a Bayesian
360   framework: case study with peak-over-threshold data, Stoch. Env.Res. Risk A, 21, 97–112, 2006.

361   Sutton, R. and Dong, B.: Atlantic Ocean influence on a shift in European climate in the 1990s, Nature Geoscience, 5, 788–
362   792, 2012.

363   Sutton, R. and Hodson, D.: Atlantic Ocean Forcing of North American and European summer Climate, Science, 2005.

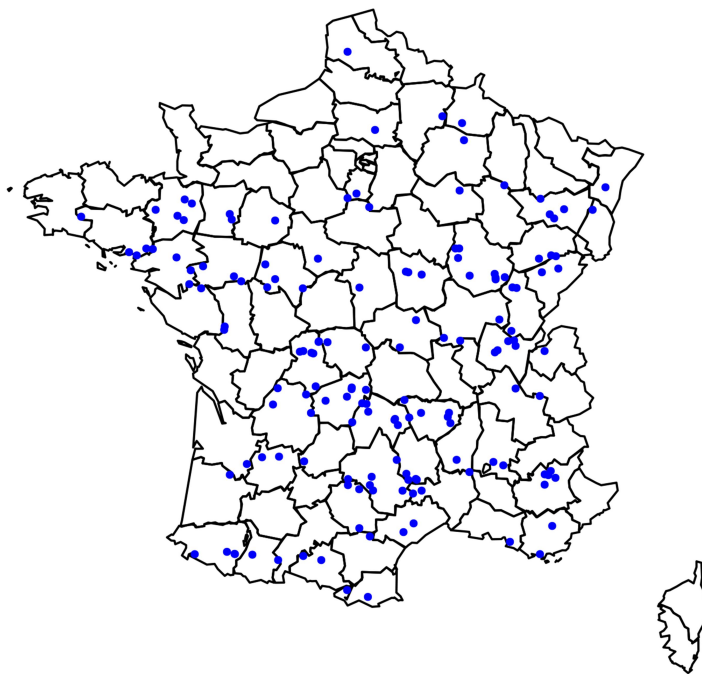
364   Teweles, J. and Wobus, H.: Verification of prognosis charts, Bull. Am. Meteorol. Soc., 35(10), 455–463, 1954.

365   Wang, W., Vrijling, J., Van Gelder, P., and Ma, J.: Testing and modelling autoregressive conditional heteroskedasticity of
366   streamflow processes, Nonlinear Processes in Geophysics, 12, 55–66, 2005.

367   Wilson, D., Hannah, D., and McGregor, G.: A large-scale hydroclimatological perspective on western European river flow
368   regimes, Hydrology Research, 44.5, 809–833, 2013.

369   **Figures**

370

371

372

373   **Fig. 1. Streamflow time series stations map; Stations chosen are those with little anthropogenic impact.**

374

375

376

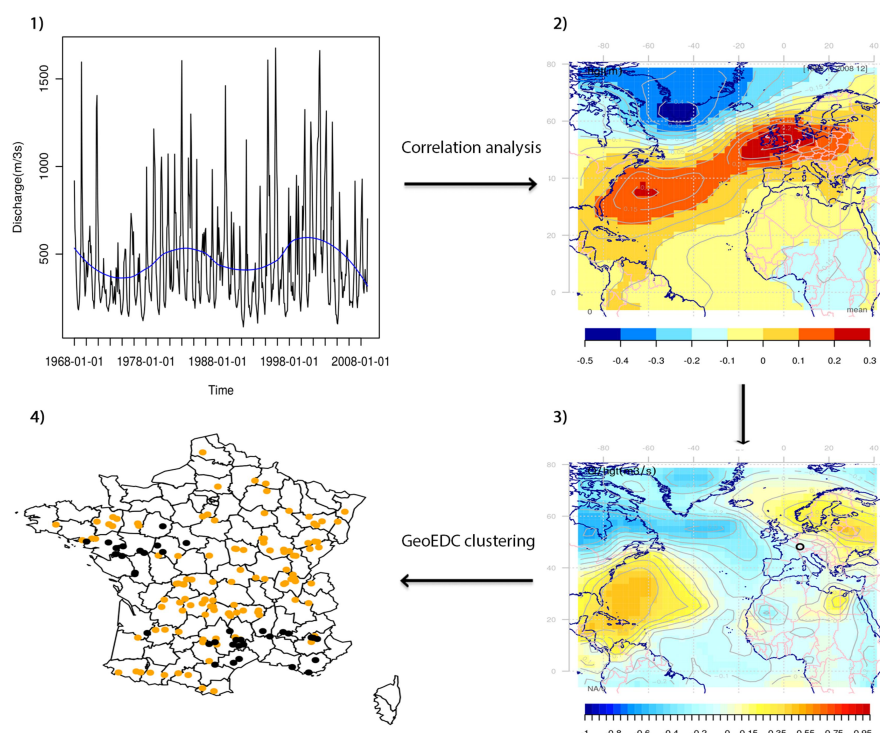



Fig. 2. Workflow of the study: (1) Streamflow and climate fields 6, 10 and 21 years TSV are extracted by multiresolution analysis;
(2) A point wise correlation analysis is performed; (3) A climate correlation field (CCF) is obtained for each of the 152 streamflow
gauging stations; (4) GeoEDC clustering is applied to each set of CCFs for given climate field and TSV allowing to establish a
cluster map.

381

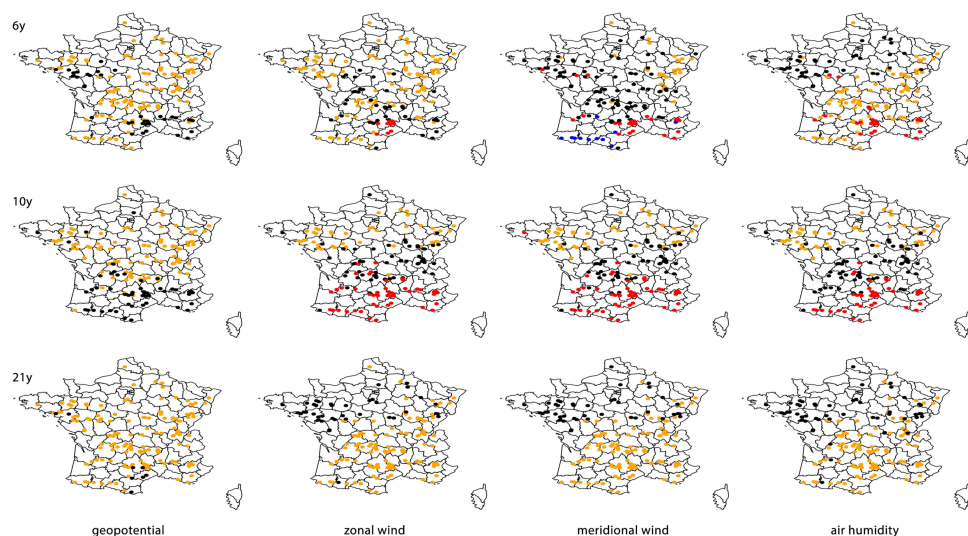
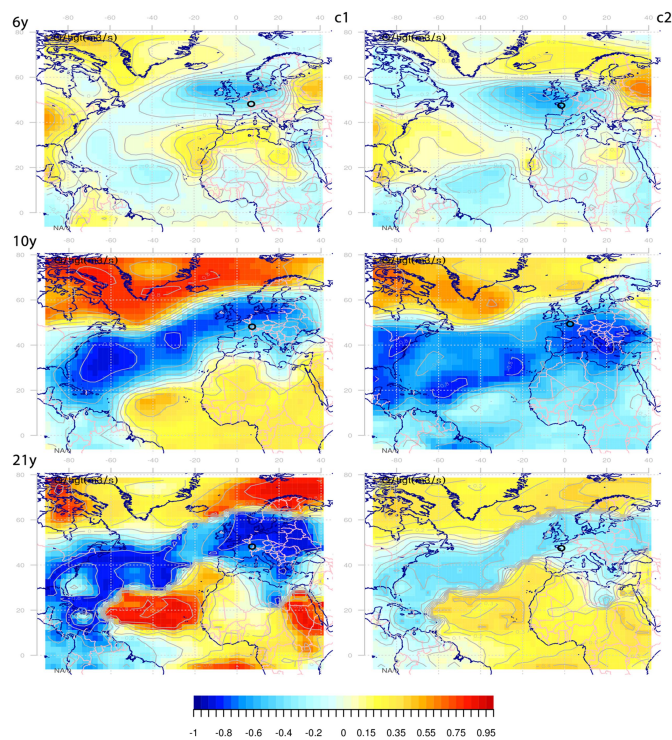
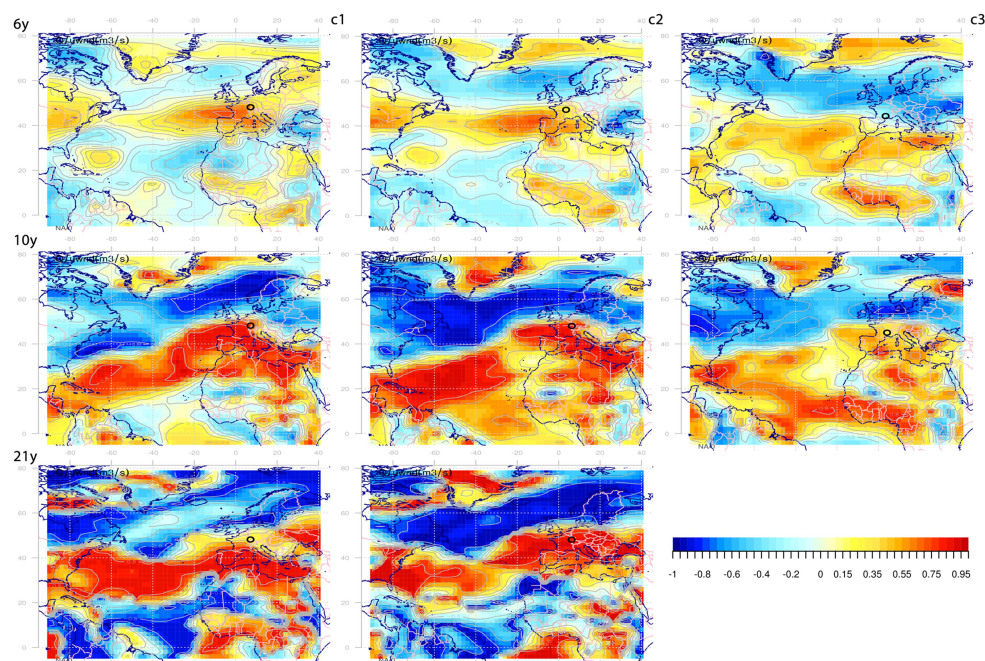
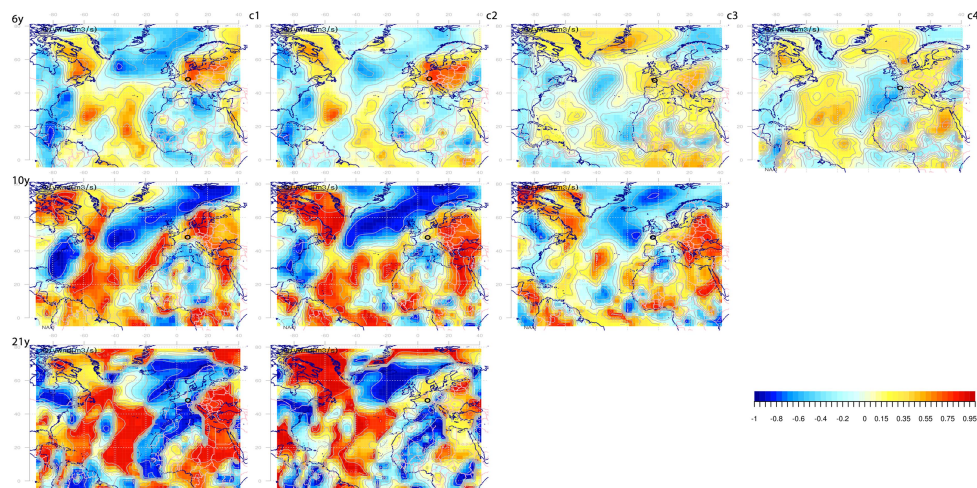
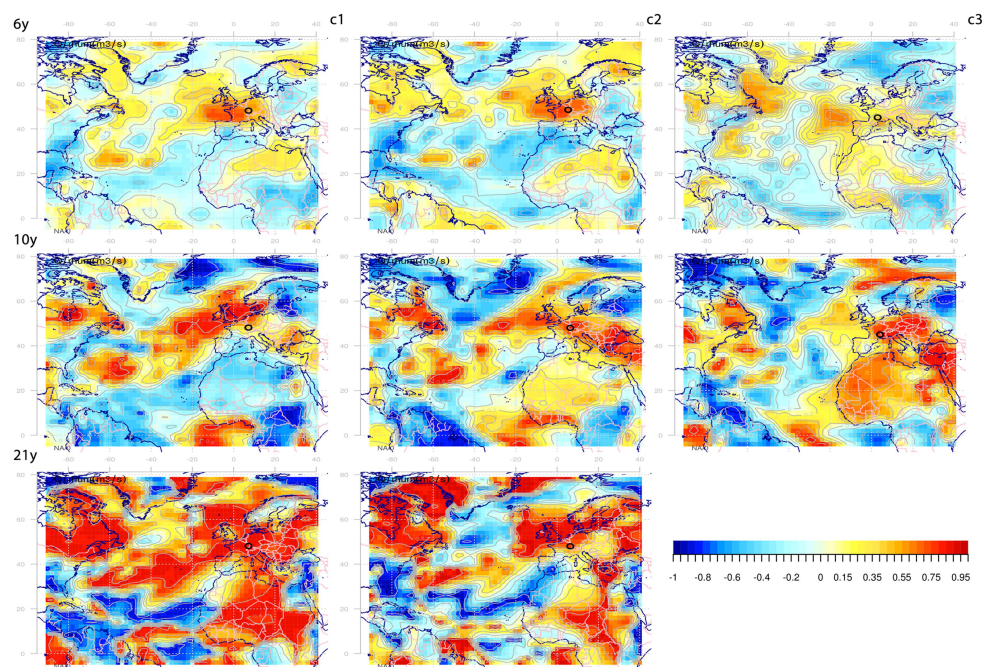



Fig. 3. Cluster map of climate drivers correlated with streamflow over France: rows are the different TSVs : 6 year (1st row), 10 year (2nd row) and 21 year (3rd row) time scales

382
383

384

385   **Fig. 4(a). Average climate correlation fields for geopotential; TSVs in rows, clusters in columns.**

386

387   **Fig. 4(b). 1Average climate correlation fields for zonal wind; TSVs in rows, clusters in columns.**

389   **Fig. 4(c). Average climate correlation fields for meridional wind; TSVs in rows, clusters in columns.**

390

391   **Fig. 4(d). Average climate correlation fields for humidity; TSVs in rows, clusters in columns.**

392

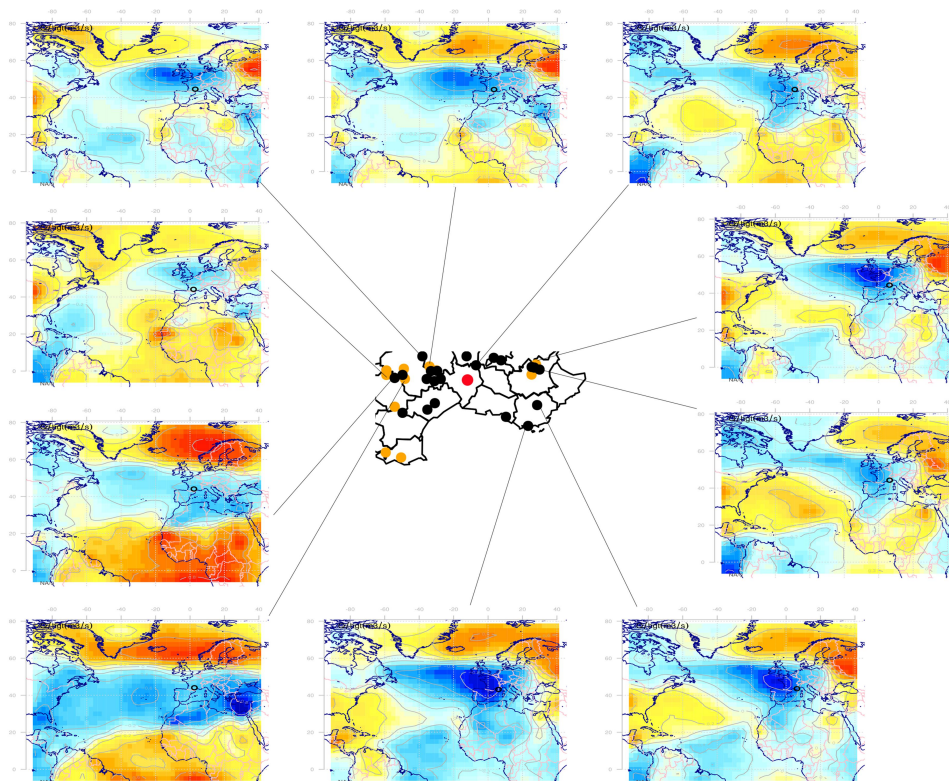
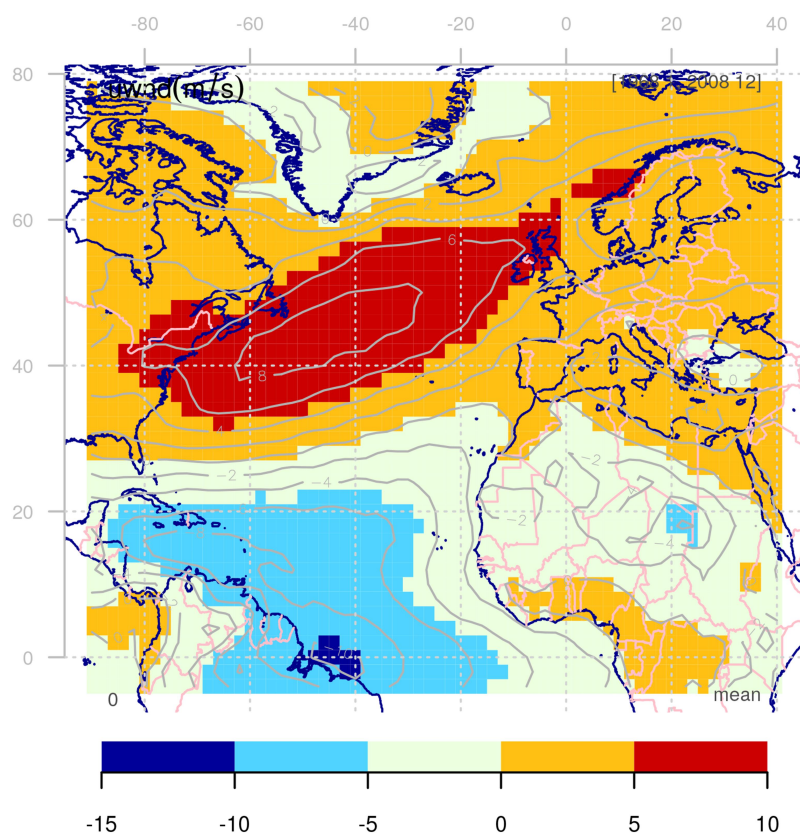



Fig. 5. map of the 6 year TSV geopotential CCFs for some selected stations that are some distance away from the geometric center (red point) of cluster 2 (the set of black points). Sampled stations are station at the interface between cluster 1 and 2 (left and upper left CCFs), the closest station from the geometric center (upper right CCF) and stations at the administrative boundaries of the cluster (lower right and bottom CCFs)

397

398 **Fig. 6. Average Zonal wind Field over the Euro Atlantic Area**

399

400

401

402

403