ADVANCES IN FOREST FIRE RESEARCH 2022

Edited by
DOMINGOS XAVIER VIEGAS
LUÍS MÁRIO RIBEIRO

Clusters analysis applied to drought and forest fires in mainland Portugal (NUT III regions) from 1980 to 2019

Edna Cardoso*[1,2]; Ilda Novo[2]; Nuno Moreira[2]; Pedro Silva[2]; Álvaro Silva[2]; Vanda Pires[2]

[1]*University of Coimbra, FCTUC, Department of Mechanical Engineering. Pólo II*
Rua Luís Reis Santos, 3030-788 Coimbra, Portugal, *{uc2004118633@student.uc.pt}*
[2]*IPMA. Rua C do Aeroporto, 1749-077 Lisboa, Portugal,*
{edna.cardoso, nuno.moreira, ilda.novo, pedro.silva, alvaro.silva, vanda.cabrinha}@ipma.pt

** Corresponding author*

Keywords

Drought, Forest Fires, Clusters Analysis, NUT III

Abstract

In order to better understand Extreme Fire Event conditions at the regional level in Portugal, a preliminary analysis on drought conditions and forest fires considering NUT III administrative organization was performed and is presented in this work. This analysis will support a more comprehensive study on the relation between weather-climate conditions (drought, heat, wind) and forest fires (occurrences and burned areas). To classify the drought, two drought indices were considered – PDSI (Palmer Drought Severity Index) and SPI (Standardized Precipitation Index), the last one for 3- and 6 months time scales. The computation was based on weather station observations for the period 1981 - 2019 in mainland Portugal, with aggregation in 23 NUT III on a monthly basis. To analyse the forest fires distribution per NUT III a monthly dataset of fire occurrences and burned areas during the period was used. The exploratory analysis was carried out in the STATISTICA software, using two clustering methods (K-Means and tree clustering). Since the results of those methods are subjective, before applying them, in the case of drought, the ideal number of clusters was calculated, making a sensibility analysis using the R tool NbClust. The results from clustering are presented through maps in which the groups of NUT III regions with the same behaviour are identified.

1. Introduction

Droughts are a natural occurring phenomenon, essentially associated with deficient precipitation values, over an extended period of time, resulting in low soil moisture and influencing water resources availability (Riley et al. 2013). Portugal is a country recurrently affected by drought, which has implications in fire occurrences and burned areas, with dry years usually being related with the greatest severity wildfires (Viegas et al. 1999). For a better understanding of the spatial distribution, an exploratory analysis of drought and forest fires was made separately. The final goal of the study is to analyse the regional distribution of the drought and its relation to the fires. The characterization of the droughts for mainland Portugal (in 23 regions of NUT III shown in Figure 1) is done using the Palmer Drought Severity Index (PDSI) and 3- and 6-months Standardized Precipitation Index (IPMA 2021), both presented in Table 1. This data is available on a monthly basis from January 1981 to December 2019. Likewise, for the characterization of the distribution of forest fires for mainland Portugal (in 23 NUT III regions), fires occurrences and burned areas on a monthly basis, from January 1980 to December 2018, were used. The first step was to develop a preliminary analysis of drought conditions calculating frequency and temporal evolution per NUT III. The second step aimed to understand how each NUT III is related to its neighbours considering the drought, through the use of correlation and applying two cluster analysis methods (K-Means clustering and Tree clustering) to identify the spatial drought and spatial distribution of forest fires in NUT III regions. Since this analysis has a subjective nature, to estimate the best number of NUT III clusters according to drought, a sensibility analysis based on the R tool NbClust was performed. The next chapters are organized as follows: i) preliminary analysis of the drought in Portugal, ii) detailed analysis of both drought and forest fires on a NUT III basis, and iii) conclusions drawn from presented data.

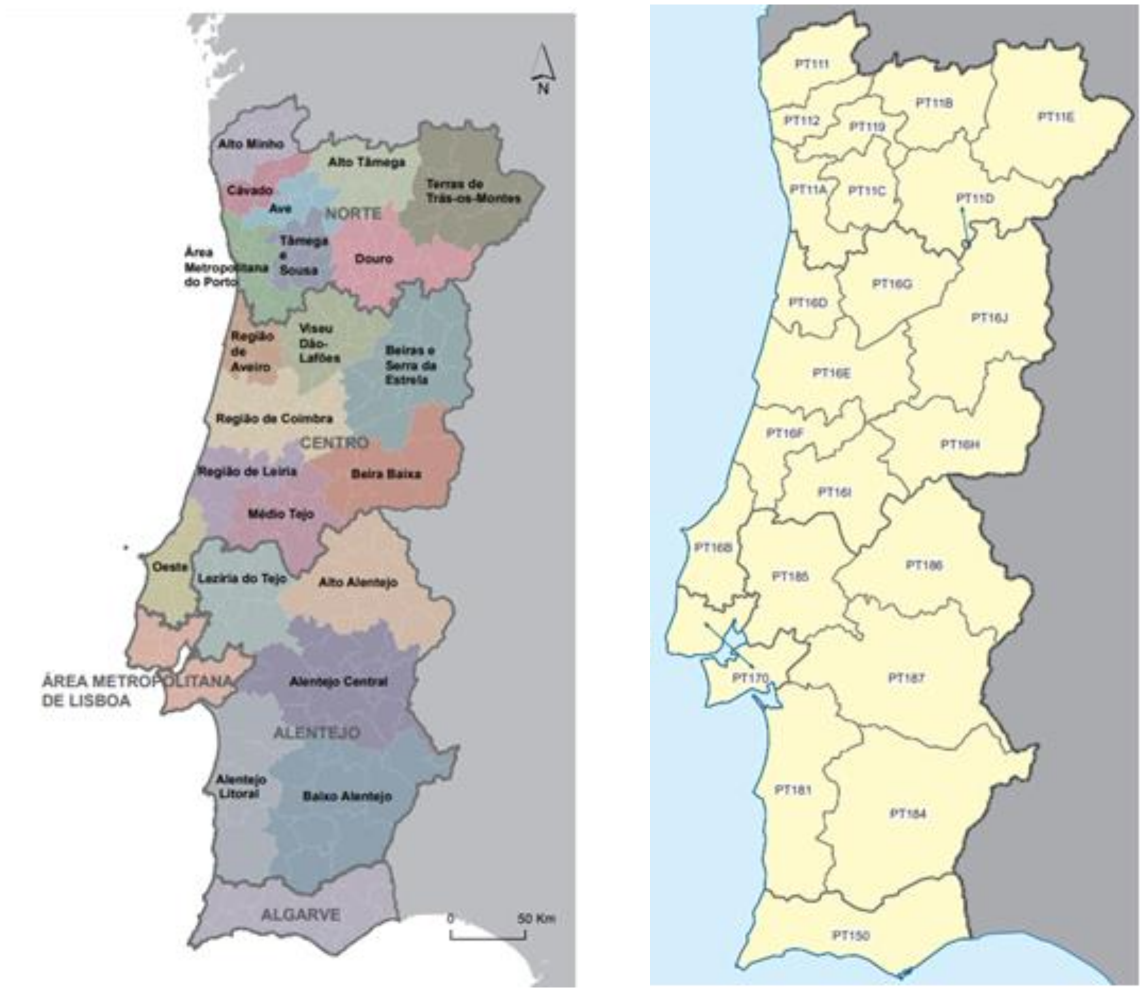

Figure 1 – Name of each regional division (NUTS 2013) applied in May 2015 (on the left) (the image is adapted from (INE 2015)) and the respective regional codes (on the right) (the image is adapted from (EUROSTAT 2016)).

Table 1 - PDSI and SPI index category with their respective value ranges and colour code. (the Table is adapted from (Portal do Estado de Ambiente 2021)).

PDSI	Drought classes	SPI
>= 4.00	extremely wet	>= 2.00
3.00 a 3.99	very wet	1.50 a 1.99
2.00 a 2.99	moderately wet	1.00 a 1.49
0.50 a 1.99	slight wet	0.99 a 0.50
0.49 a -0.49	near normal	0.49 a -0.49
-0.50 a -1.99	mild drought	-0.50 a -0.99
-2.00 a -2.99	moderate drought	- 1.00 a -1.49
-3.00 a -3.99	severe drought	- 1.50 a -1.99
<= - 4.00	extreme drought	<= - 2.00

2. Preliminary drought analysis

The first step was a frequency analysis of drought conditions per NUT III (see an example for PDSI in Figure 2). The analysis was followed by an evaluation of the temporal evolution of drought in the period 1981 to 2019, taking into consideration the spatial prevalence of drought on NUT III basis (see an example for PDSI in Figure 3).

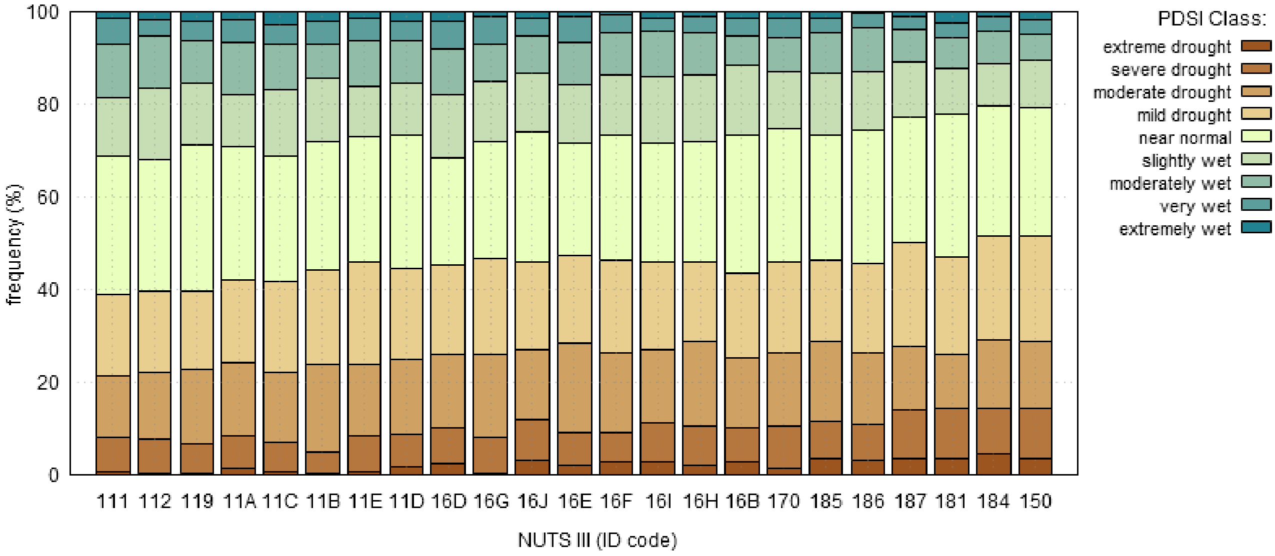

Figure 2 - Frequency of monthly PDSI classes (%) per NUT III in the period 1981-2019 for Portugal.

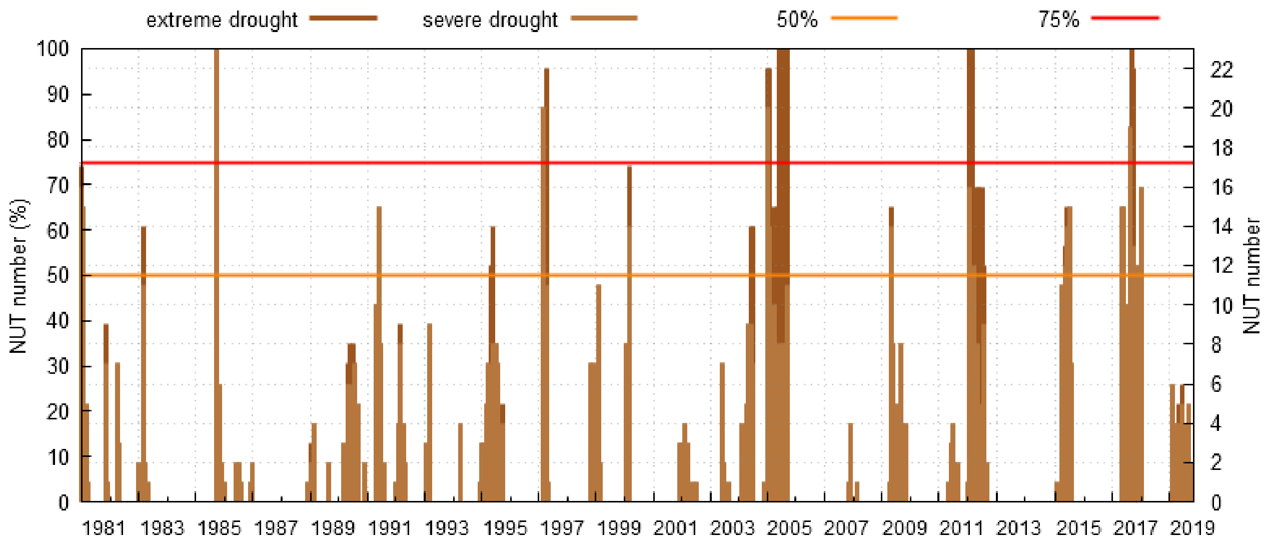

Figure 3 - Number of NUT III (%) in the 2 higher classes of PDSI index (severe and extreme) in the period 1981-2019 for Portugal.

3. Drought on NUT III basis

A second step aimed to understand how each NUT III is linked to its neighbours in relation to drought, through the use of correlation and clustering methods. Regarding correlation, simple correlation matrixes were used for a preliminary inspection of the data (see the example for PDSI considering all months of the year in Table 2).

Table 2 - Correlation matrix for PDSI in NUT III, considering all months within the year, from 1981 to 2019.

	111	112	119	11A	11C	11B	11E	11D	16D	16G	16J	16E	16F	16I	16H	16B	170	185	186	187	181	184	150
111	1,000	0,958	0,930	0,944	0,904	0,908	0,837	0,866	0,872	0,884	0,844	0,888	0,854	0,767	0,767	0,745	0,721	0,696	0,733	0,660	0,599	0,560	0,460
112	0,958	1,000	0,975	0,930	0,892	0,921	0,820	0,866	0,859	0,875	0,846	0,867	0,839	0,765	0,767	0,746	0,708	0,702	0,736	0,661	0,588	0,551	0,438
119	0,930	0,975	1,000	0,924	0,917	0,952	0,819	0,873	0,857	0,867	0,831	0,856	0,829	0,754	0,748	0,737	0,689	0,694	0,724	0,644	0,560	0,516	0,402
11A	0,944	0,930	0,924	1,000	0,966	0,911	0,864	0,891	0,929	0,905	0,853	0,914	0,876	0,774	0,788	0,754	0,718	0,700	0,744	0,657	0,593	0,537	0,423
11C	0,904	0,892	0,917	0,966	1,000	0,909	0,843	0,876	0,885	0,878	0,816	0,869	0,835	0,741	0,749	0,727	0,689	0,684	0,717	0,636	0,565	0,496	0,392
11B	0,908	0,921	0,952	0,911	0,909	1,000	0,910	0,929	0,880	0,904	0,879	0,887	0,873	0,810	0,814	0,801	0,770	0,757	0,785	0,710	0,629	0,583	0,472
11E	0,837	0,820	0,819	0,864	0,843	0,910	1,000	0,941	0,869	0,893	0,906	0,909	0,913	0,868	0,885	0,857	0,843	0,817	0,862	0,792	0,719	0,691	0,576
11D	0,866	0,866	0,873	0,891	0,876	0,929	0,941	1,000	0,893	0,904	0,918	0,914	0,909	0,857	0,874	0,845	0,814	0,811	0,838	0,765	0,684	0,649	0,529
16D	0,872	0,859	0,857	0,929	0,885	0,880	0,869	0,893	1,000	0,902	0,878	0,937	0,905	0,815	0,827	0,795	0,748	0,747	0,779	0,695	0,637	0,587	0,462
16G	0,884	0,875	0,867	0,905	0,878	0,904	0,893	0,904	0,902	1,000	0,931	0,929	0,911	0,857	0,879	0,838	0,808	0,803	0,840	0,768	0,711	0,662	0,568
16J	0,844	0,846	0,831	0,853	0,816	0,879	0,906	0,918	0,878	0,931	1,000	0,919	0,917	0,890	0,935	0,881	0,860	0,854	0,884	0,839	0,779	0,751	0,643
16E	0,888	0,867	0,856	0,914	0,869	0,887	0,909	0,914	0,937	0,929	0,919	1,000	0,983	0,892	0,887	0,854	0,804	0,812	0,851	0,772	0,710	0,676	0,543
16F	0,854	0,839	0,829	0,876	0,835	0,873	0,913	0,909	0,905	0,911	0,917	0,983	1,000	0,948	0,921	0,925	0,865	0,895	0,908	0,844	0,788	0,750	0,618
16I	0,767	0,765	0,754	0,774	0,741	0,810	0,868	0,857	0,815	0,857	0,890	0,892	0,948	1,000	0,937	0,941	0,892	0,935	0,952	0,895	0,838	0,808	0,691
16H	0,767	0,767	0,748	0,788	0,749	0,814	0,885	0,874	0,827	0,879	0,935	0,887	0,921	0,937	1,000	0,922	0,918	0,908	0,950	0,901	0,849	0,816	0,712
16B	0,745	0,746	0,737	0,754	0,727	0,801	0,857	0,845	0,795	0,838	0,881	0,854	0,925	0,941	0,922	1,000	0,949	0,984	0,934	0,924	0,894	0,839	0,738
170	0,721	0,708	0,689	0,718	0,689	0,770	0,843	0,814	0,748	0,808	0,860	0,804	0,865	0,892	0,918	0,949	1,000	0,926	0,923	0,927	0,904	0,856	0,766
185	0,696	0,702	0,694	0,700	0,684	0,757	0,817	0,811	0,747	0,803	0,854	0,812	0,895	0,935	0,908	0,984	0,926	1,000	0,940	0,939	0,907	0,853	0,761
186	0,733	0,736	0,724	0,744	0,717	0,785	0,862	0,838	0,779	0,840	0,884	0,851	0,908	0,952	0,950	0,934	0,923	0,940	1,000	0,957	0,900	0,875	0,773
187	0,660	0,661	0,644	0,657	0,636	0,710	0,792	0,765	0,695	0,768	0,839	0,772	0,844	0,895	0,901	0,924	0,927	0,939	0,957	1,000	0,960	0,946	0,852
181	0,599	0,588	0,560	0,593	0,565	0,629	0,719	0,684	0,637	0,711	0,779	0,710	0,788	0,838	0,849	0,894	0,904	0,907	0,900	0,960	1,000	0,955	0,895
184	0,560	0,551	0,516	0,537	0,496	0,583	0,691	0,649	0,587	0,662	0,751	0,676	0,750	0,808	0,816	0,839	0,856	0,853	0,875	0,946	0,955	1,000	0,925
150	0,460	0,438	0,402	0,423	0,392	0,472	0,576	0,529	0,462	0,568	0,643	0,543	0,618	0,691	0,712	0,738	0,766	0,761	0,773	0,852	0,895	0,925	1,000

In what concerns clustering, two complementary approaches were taken. On one side, a sensibility analysis based on the R tool NbClust (Ghazzali 2014) was performed to find the best number of NUT III clusters according to drought. Most of the aggregation methods indicated 2 or 3 NUT III clusters to be the ideal number of clusters for Portugal, also with a high number of methods to indicate 8 NUT III clusters as ideal (see the example for PDSI considering all months of the year in Table 3).

Table 3 - Number of indexes (per aggregation method and for all methods – last column) that indicate the best number of NUT III clusters, using the R Tool NBClust, for PDSI considering all months.

Number of Clusters	Aggregation Method									All methods
	Ward.D	Ward.D2	Single	Complete	Average	McQuitty	Median	Centroid	Kmeans	
2	6	**11**	**12**	**10**	7	6	8	11	**13**	84
3	**11**	6	8	4	**8**	**11**	2	3	8	61
4	1	1	2	3	2	4	2	**12**	0	27
5	1	1	0	0	1	3	2	0	1	9
6	1	1	0	0	3	0	**12**	0	2	19
7	0	2	2	2	0	0	0	0	0	6
8	4	3	4	6	2	4	1	1	4	29

On the other side, final clustering results were obtained for K-Means (with 3 different initial cluster centers settings, named A (choose observations to maximize initial between-cluster distances), B (sort distances and take observations at constant intervals) and C (choose the first N (number of clusters) observations) and tree clustering methods (Ward's method for 2 metric distances, Euclidean and Pearson r) in STATISTICA software (TIBCO 2022). The example for PDSI considering all months of the year is shown in Figure 4. Four solutions were considered for addressing season sensibility: i) all months of the year; ii) summer months; iii) spring months; iv) non-winter months.

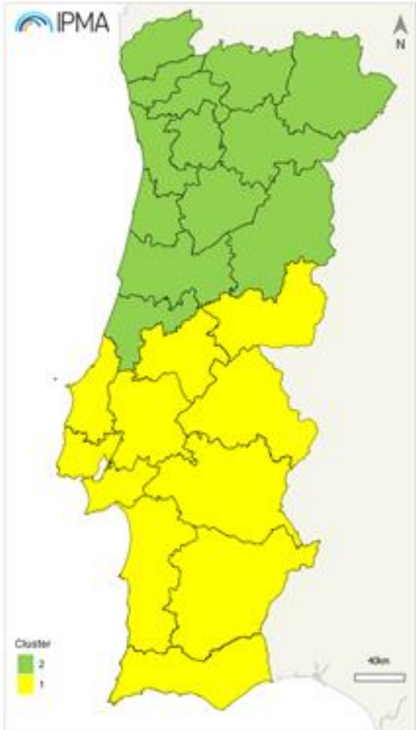

Figure 4 – NUT III aggregation in 2 clusters considering monthly values for 3 considered drought indexes PDSI, SPI3 and SPI6 for all months of the year between January 1981 and December 2019, using 3 different settings for K-Means and 2 metric distances, Euclidean and Pearson r for Ward's method.

4. Preliminary Results for Drought

The results suggest that the longer the period considered throughout the year (inherent to the drought index itself or the number of months considered for drought analysis), the greater the tendency for the drought behaviour to be dominated by the Montejunto-Estrela mountain system. For shorter periods within the year, a finer clustering should be the most appropriate.

Thus, for all months of the year or for the non-winter period (March to November) the drought presents a differentiated behaviour for two main sets of NUTS III (see Figure 4):

- Oeste, Lezíria do Tejo, Médio Tejo, Beira Baixa and other NUT III to the south;
- Região de Coimbra, Região de Leiria, Beiras e Serra da Estrela and remaining NUT III to the north.

In a finer aggregation as a function of drought, the sets of NUT III that show the same behaviour for the meteorological drought are identified below for each of the 4 periods of the year (see Figure 5):

Every month of the year:
- Terras de Trás-os-Montes + Douro;
- Região de Coimbra + Região de Leiria.

Summer:
- Terras de Trás-os-Montes + Douro;
- Região de Coimbra + Região de Leiria;
- Oeste + Lezíria do Tejo.

Spring:
- Terras de Trás-os-Montes + Douro;
- Alto Minho + Cávado;
- Oeste + Lezíria do Tejo.

Non-winter:
- Viseu Dão Lafões + Beiras e Serra da Estrela;
- Região de Coimbra + Região de Leiria;
- Beira Baixa + Alto Alentejo;
- Oeste + Lezíria do Tejo.

In summary, it can be considered that there are 3 sets of NUT III presenting the same behaviour in terms of drought, almost regardless of the period of the year considered:

- **Terras de Trás-os-Montes and Douro**, which just do not belong to the same set in the non-winter period;
- **Região de Coimbra and Região de Leiria**, which do not belong to the same set in the spring period;
- **Oeste and Lezíria do Tejo**, which just do not belong to the same set considering all the months of the year.

There are also other NUT III with drought similarities, but to which this analysis does not allow to attribute the same level of confidence, since there was one combination of drought index (PDSI, SPI3 or SPI6) and a clustering solution (in these cases the solutions "K-Means B", "K-Means C" or "Ward - Pearson") in which aggregation was not verified. In this situation, for 3 of the 4 periods of the year considered, the following sets of NUT III are identified:

Every month of the year:
- Alto Minho and Cávado;
- Viseu Dão Lafões and Beiras e Serra da Estrela;
- Beira Baixa and Alto Alentejo;
- Oeste and Lezíria do Tejo.

Spring:
- Alto Minho, Cávado and Ave;
- Alentejo Litoral and Baixo Alentejo.

Summer:
- Oeste, Área Metropolitana de Lisboa and Lezíria do Tejo.

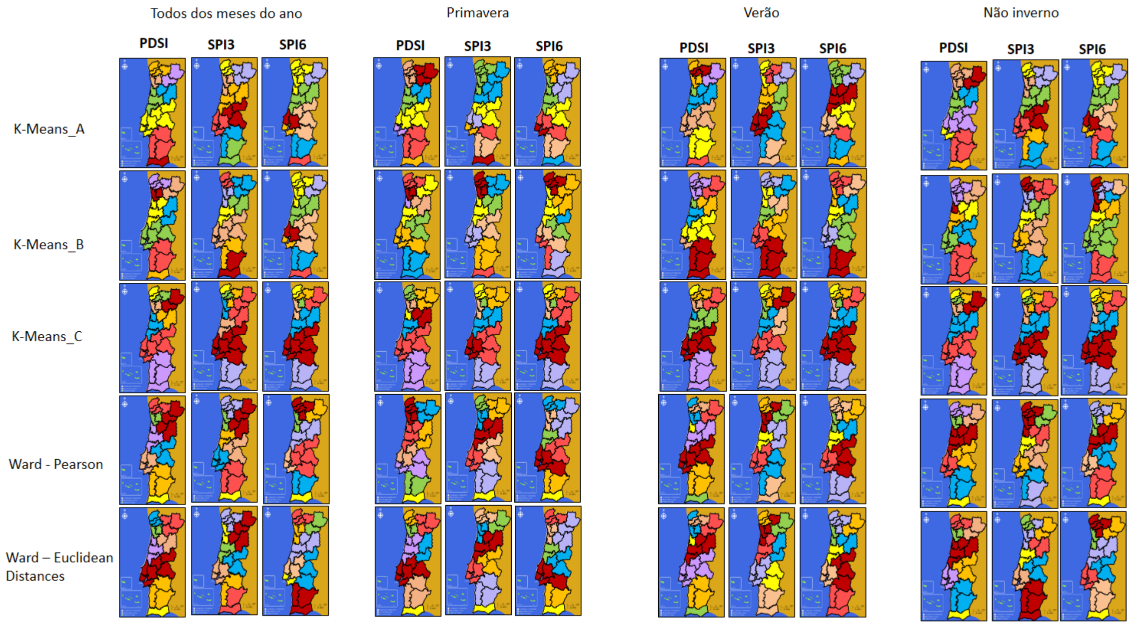

Figure 5 - Spatial aggregation of NUT III into 8 clusters for drought behavior between January 1981 and December 2019 as a function of; i) drought index (PDSI, SPI3 and SPI6); (ii) the time of year (All months of the year, spring, summer and non-winter) and iii) clustering solutions ("K-Means A", "K-Means B", "K-Means C", "Ward – Pearson", "Ward – Euclidean").

5. Forest fires distribution in NUT III

The third step was towards understanding NUT III spatial distribution of fire occurrences number and burned areas, considering both all months of the year (January to December) and forest fire seasons (June to October) for 39 years (1980 to 2018). The results obtained for 8 clusters by applying a tree clustering analysis - Ward's method with Pearson's metric distance r - are shown in the maps in Figure 6. It should be noted that the same results are obtained whether original or normalized data (division by the area of each NUT III) are used.

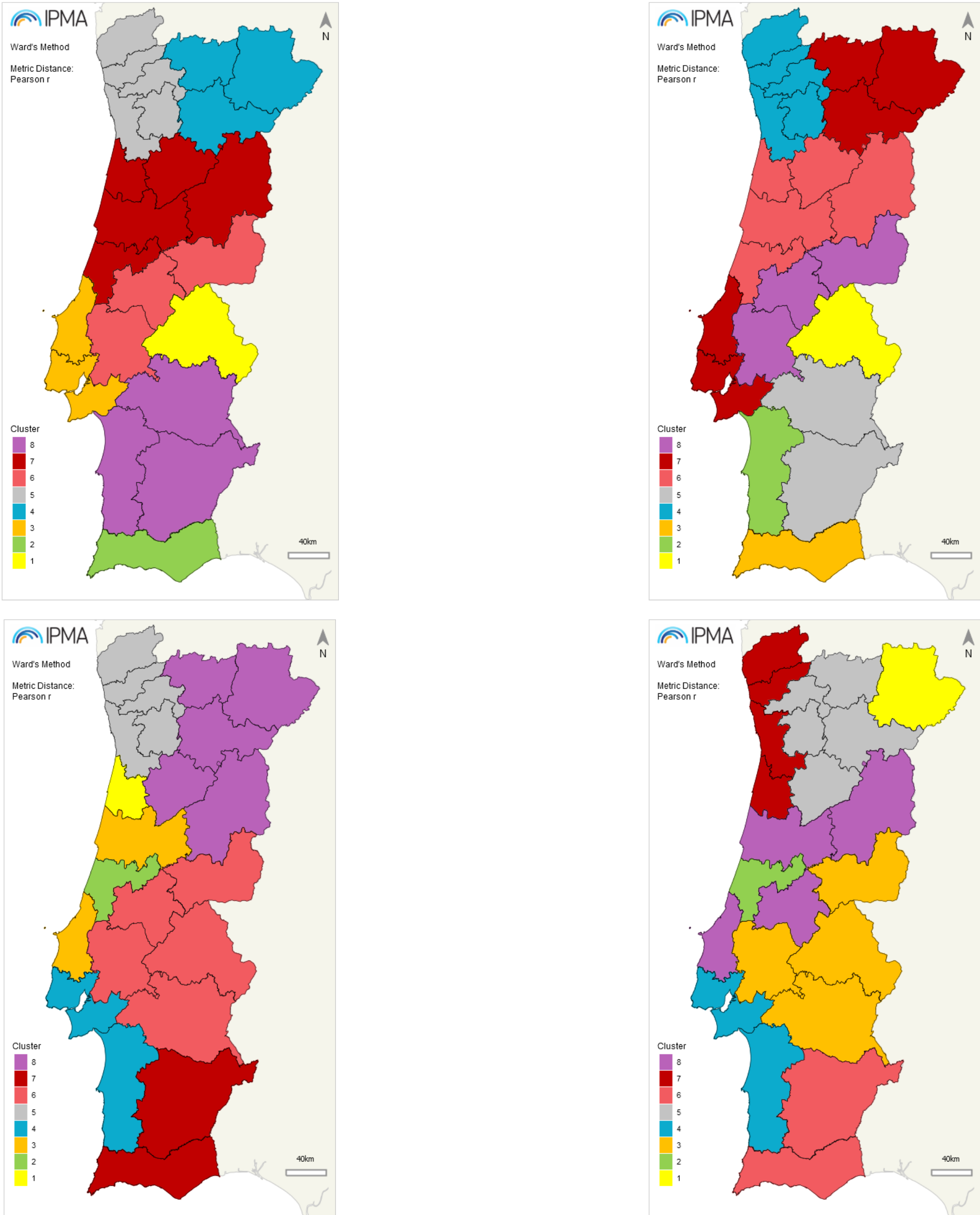

Figure 6 – NUT III aggregation in 8 clusters considering monthly values of fire occurrences (top) and burned areas (bottom) considering all months of the year (in the left) and the fire season months between June and October (in the right), applying a tree clusters analysis with Ward's method using the metric distance of Pearson r. Data from 1980 to 2018.

6. Preliminary Results for Forest Fires

Analysing Figure 6, it can be seen that for the case of forest fire occurrences there is a more homogeneous grouping in terms of NUT III spatial proximity considering both all months of the year and fire season months. The only differences between all months of the year (at the top on the left) and fire season months (at the top on the right) lies in the fact that, in the fire season months the Oeste and Área Metropolitana de Lisboa are in the same cluster as Alto Tâmega, Terra de Trás-os-Montes and Douro, and that there is an individualization of Alentejo Litoral in a single cluster, no longer belonging to Alentejo Central and Baixo Alentejo cluster (as for all months of the year).

The map for burned areas (bottom maps in Figure 6) shows a more homogeneous grouping of NUT III regions in the north and interior of the country, while for the coastal part of the country this is not the case. Anyway, it seems that grouping follows north-south guidance.

Comparing the maps in Figure 6 of the fire occurrences (at the top) with those of the burned areas (at the bottom), it can be seen that in the case of the burned areas there is less spatial consistency distribution of clusters. This suggests that there may be additional factors besides the regional proximity (on NUT III basis) that have to be considered to better understand the regional distribution of burned areas in Portugal.

7. References

Riley, Karin L, John T Abatzoglou, Isaac C Grenfell, Anna E Klene, and Faith Ann Heinsch. 2013. "The Relationship of Large Fire Occurrence with Drought and Fire Danger Indices in the Western USA, 1984 – 2008 : The Role of Temporal Scale," 894–909.

Viegas, D Xavier, Giovanni Bovio, Antonio Nosenzo, and Bernard Sol. 1999. "Comparative Study of Various Methods of Fire Danger Evaluation in Southern Europe" 9 (4): 235–46.

IPMA. 2021. "Definição Dos Índices de Seca Meteorológica." Lisboa. 2021. https://www.ipma.pt/pt/oclima/observatorio.secas/spi/apresentacao/definicao/.

INE. 2015. "NUTS 2013 AS NOVAS UNIDADES TERRITORIAIS PARA FINS ESTATÍSTICOS." 2015.

EUROSTAT. 2016. "NUTS-3-Map-PT." 2016. https://ec.europa.eu/eurostat/documents/345175/7451602/2016-NUTS-3-map-PT.pdf.

Portal do Estado de Ambiente. 2021. "Riscos Ambientais - Seca." 2021. https://rea.apambiente.pt/content/seca.

Ghazzali, Nadia. 2014. "NbClust : An R Package for Determining the Relevant Number of Clusters in a Data Set" 61 (6).

TIBCO. 2022. "Guide of Statistica Software - Data Science Textbook." 2022. https://docs.tibco.com/data-science/GUID-E79CEA1A-02B7-4C9B-9328-980D1DF20FEB.html.