IDENTIFYING FAVORABLE SPATIO-TEMPORAL CONDITIONS FOR WEST NILE VIRUS OUTBREAKS BY CO-CLUSTERING OF MODIS LST INDICES TIME SERIES

Verónica Andreo[1], Emma Izquierdo-Verdiguier[1,2], Raúl Zurita-Milla[1], Roberto Rosà[3], Annapaola Rizzoli[3], Anna Papa[4]

[1]Faculty of Geo-Information Science and Earth Observation (ITC). University of Twente. The Netherlands.
[2]Image Processing Laboratory (IPL). University of Valencia. Spain.
[3]Center of Research and Innovation. Fondazione Edmund Mach. Italy.
[5]Arboviruses Reference Laboratory. Aristotle University of Thessaloniki. Greece.

ABSTRACT

This study presents the first results of the use of co-clustering to identify potential spatial and temporal concurrences of favourable conditions for the emergence and maintenance of West Nile Virus (WNV) in Greece. We applied the Bregman block average co-clustering algorithm with I-divergence to various time series (from 2003 to 2016) of indices derived from Land Surface Temperature (LST) reconstructed from MODIS products. The results show that the combination of two temporal and three spatial groups performs best in identifying times and areas with and without WNV human cases, yielding smaller standard deviations in co-clusters. Among the indices that appeared to perform better we found: number of summer days, annual average of mean and maximum LST, potential number of mosquito and virus cycles (EIP) and mean LST of the WNV transmission season. These variables are consistent with known effects of temperature over mosquito development and reproduction as well as virus amplification. Further research will be carried out to identify groups of variables that cluster both in space and time.

Index Terms— Time series, MODIS, Land Surface Temperature, Data Mining, Geo-Health

1. INTRODUCTION

The West Nile virus (WNV) is one of the mosquito-borne flavivirus most widely distributed in the world. It causes a variety of symptoms to humans: from an asymptomatic infection to severe and even fatal encephalitis. Wild birds are the major reservoirs of the virus and the main transmission route is through *Culex* mosquito-vectors. The frequency of reported outbreaks with severe symptoms has shown an increase over the last $15-20$ years in Europe and the USA. In 2010, Greece reported the largest number of WNV human cases in Europe [1], most of them concentrated in the region of central Macedonia. In the four subsequent years, the disease further spread both southwards and eastwards, and more than 600 confirmed cases and 73 deaths were reported [2].

Several climatic and environmental conditions might determine the onset of zoonotic diseases' outbreaks. In the case of mosquito-borne pathogens such as WNV, temperature plays a central role. For example, higher than usual temperatures are known to influence vector competence, to accelerate virus replication within mosquitoes, to boost mosquitoes reproduction rates, and to prolong their breeding season. As such, temperature is one of the main environmental factors addressed when studying vector-borne viruses carried by mosquitoes [3].

Nowadays, most of these relevant variables or their proxies can be derived from remote sensing time series, for example Land Surface Temperature (LST). These variables can be analyzed by unsupervised methods to identify regions (group of pixels) and periods (time series time stamps) that behave similarly and imply favourable conditions for the onset and development of disease outbreaks. Within unsupervised approaches, *clustering* is a fundamental tool in data analysis as it allows the exploration of complex datasets. Co-clustering algorithms are able to find blocks of similar data in a matrix by simultaneously considering information along the rows and columns [4]. In the last decade, co-clustering methods were used for pattern analysis in disciplines such as text and document relation and local pattern of gene expression. However, these methods were not used in remote sensing applications until recent years. In this field, the co-clustering allows to analyze hyper-spectral images (spectral-spatial groups) [4] or to detect space-time groups in time series data of meteorological stations and gridded phenological indices [5, 6].

This paper presents the first exploratory analysis of climatic indices derived from LST time series using the co-clustering method in the context of Geo-Health research. The emergence of WNV in Greece in 2010 has been related to positive anomalies in temperature that year [1]. Therefore, in this study we attempt to identify the potential spatial and temporal concurrence of favorable conditions that might relate to the emergence and maintenance of WNV in Greece. The rest of the paper is outlined as follows. Section 2 reviews materials and methods used in this work. Section 3 presents the experimental results and its discussion. Finally, Section 4 concludes this paper.

2. MATERIALS AND METHODS

2.1. Data

We used a series of biologically meaningful indices known or hypothesized to affect mosquito populations, WNV transmission or disease risk on an annual basis (See [7] and references therein). The indices were derived from a time series of daily reconstructed LST products from MODIS sensor for the period $2003 - 2016$ [8] at 1 km spatial resolution, for an area covered by 805×889 pixels. In total, we estimated 23 indices including annual averages of minimum, maximum and average LST (LST_minimum, LST_maximum and LST_average, respectively), standardized anomalies, number of summer days (Days_Tmax_high25), number of days with average LST $\in [20, 30]°$C (Days_Tmean_h20_l30), average LST of mosquito growing season (Tmean_mosq_season), average LST of WNV transmission season (Tmean_WNV_season), length of mosquito growing season (Mosq_season_length) and of WNV transmission season, number of potential mosquito cycles per year (Mosq_cycles) and number of potential Extrinsic Incubation Periods (EIP) per year (WNV_EIP). Details about the estimation of the indices can be found in [7].

2.2. Analysis

We applied the Bregman block average co-clustering algorithm with I-divergence (BBAC_I) to each of the time series of LST derived indices in order to identify similar observations along both spatial and temporal dimensions. In essence, co-clustering looks for blocks of rows and columns that permit to recompose the original matrix by minimizing the distance between the recomposed matrix and the original one. We chose the I-divergence metric based on previous experiences with time series co-clustering [5]. We run the co-clustering for different combinations of space and time groups and selected the best set according to the proportion of WNV cases during the period $2010 - 2014$ that overlapped with single co-clusters. We also estimated mean and standard deviation for each co-cluster in all combinations tested and for each of the 23 variables studied.

We used GRASS GIS for the processing of MODIS LST time series and extraction of derived indices. R implementation of the BBAC_I [1] was used for the co-clustering. The co-clustering was run in a virtual machine with 8 cores and 32 GB of RAM provided by ESA RSS Cloudtoolbox[2].

[1]https://github.com/fnyanez/bbac

Table 1. Mean and standard deviation (SD) of variables that correctly split areas with and without WNV reported cases in the best co-cluster (CC) runs: 2-2, 2-3 and 3-2 (space and time, S/T).

CC	Variable	S/T	Mean	SD
2-2	Days_Tmax_high25	2-2	164.27	28.80
	LST_average	2-1	17.09	1.67
	LST_maximum	2-1	24.30	2.19
	Mosq_cycles	2-2	13.63	1.78
	Tmean_WNV_season	1-2	23.63	1.88
	WNV_EIP	1-2	15.81	3.24
	Mosq_season_length	1-1	259.78	24.27
	Tmean_mosq_season	2-2	22.39	1.79
2-3	Days_Tmax_high25	1-1	159.84	27.11
	LST_average	1-2	17.02	1.83
	LST_maximum	2-3	24.22	2.22
	Mosq_cycles	1-1	13.63	1.78
	Tmean_WNV_season	1-2	23.84	1.91
	WNV_EIP	1-3	15.67	3.19
	Mosq_season_length	1-1	262.94	22.92
	Tmean_h20_l30	2-3	110.30	19.28
	LST_minimum	2-2	10.44	1.83
3-2	Days_Tmax_high25	1-1	176.65	21.54
	LST_average	2-2	18.16	1.22
	LST_maximum	1-2	25.50	1.69
	Mosq_cycles	3-1	14.48	1.40
	Tmean_WNV_season	3-1	24.69	1.46
	WNV_EIP	2-1	17.46	2.58
	Days_Tmean_20_30	3-1	119.75	16.39
	Tmean_mosq_season	2-2	23.42	1.37

3. RESULTS

Different number of spatial and temporal groups were used in the co-clustering, from 2 to 4 in the spatial and from 2 to 3 in the temporal clusters. Since the main objective was the identification of environmental differences in time, before and after the outbreak of WNV and spatial differences in areas with and without reported human cases, we started with 2 spatial and 2 temporal groups.

We found that the combinations $2 - 2$, $2 - 3$ and $3 - 2$ (space and time) yielded the best clusters in terms of separation of areas with and without WNV reported cases and years that were grouped together. In spatial terms, there were 6 variables that appeared in all those best combinations: number of summer days, annual average of mean and maximum LST, potential number of mosquito cycles and potential num-

[2]http://eogrid.esrin.esa.int/cloudtoolbox/

ber of EIP and mean temperature of the WNV transmission season (June-October). Figure 1 shows four of these variables for the co-cluster with 3 spatial and 2 temporal groups.

Mosquito season length and mean LST of mosquito season also appeared to differentiate areas with and without WNV human cases. In all co-clustering combinations run, the spatial co-clusters obtained for these variables included more than 87% of the WNV reported cases.

In the temporal dimension, several variables were found to identify the years surrounding the onset of WNV (2009, 2010 and 2011) in the same temporal group. However, no co-clustering was able to group all 5 years with reported cases ($2010 - 2014$) in the same co-cluster. Only the annual mean LST as well as the annual average minimum LST showed groups that split the series into meaningful groups: $2003 - 2006$ and $2007 - 2016$, though years 2009 and 2011 were included in the first group.

Mean and standard deviations were estimated for each co-cluster and variable that correctly separated areas with and without WNV reported cases in the period $2010-2014$ (Table 1). The temporal cluster reported in each case, was the one including the years surrounding the onset ($2009-2011$). The lowest standard deviations were obtained for the co-cluster including 2 temporal groups and 3 spatial groups. The variables that better identified the areas with WNV cases, showed higher mean values as compared to the other co-clustering combinations. These variables and values are clearly related to favorable conditions for mosquitoes and virus development, i.e., more summer days, higher mean and maximum LST, higher number of mosquito cycles and virus EIP, as well as higher mean LST during the transmission season (June-October).

4. DISCUSSION

This study is to the best of our knowledge the first attempt to use co-clustering in the Geo-Health domain. Preliminary results seem promising, especially in terms of variables and spatial groups identified. It seems likely that an environmental signature characterizes places where WNV cases were first reported and further spread. Rather frequently the years 2009, 2010 and 2011 appeared together in temporal clusters. Human cases were first reported in 2010 in Greece, but in 2009 the virus was already detected in host birds [9]. It is then highly likely that some favorable conditions concurred those years to trigger the outbreak.

The spatial clusters obtained in our co-clustering runs seem to better separate areas with and without WNV human cases than the temporal clusters separate years with and without cases. This is because the temporal groups included years without WNV cases. Such results might be explained by the number of groups selected (i.e., 2 or 3 might not be enough), other environmental factors not considered in our study, or just by the fact that conditions leading to outbreaks were already met in Greece, but the virus was not yet established. Once it did, plus (or because of) some anomalous temperatures those years and in certain places, i.e., higher number of summer days, potential for higher number of mosquito and virus cycles, higher mean LST during mosquito growing season, the outbreak occurred given the susceptibility of population. These are all conditions known to influence vector competence, accelerate virus replication within mosquitoes, boost mosquitoes reproduction rates, and prolong their breeding season [10].

This study was an exploratory research in which we attempted to uncover variables that allow us to characterize areas where WNV have occurred and new areas behaving similarly in space and time that could potentially be at risk for future outbreaks. For example, after two years without reported WNV human cases (2015 and 2016), Greece reported WNV cases again in 2017 in ”risky” areas according to our analysis, i.e., parts of West Greece, Peloponese and Crete island [3]. The next step in this research is to use tri-clustering techniques to analyze space-time groups using the information of all LST derived indices together. This would allow us to identify groups of variables that cluster both in space and time [11].

5. REFERENCES

[1] K. Danis, A. Papa, G. Theocharopoulos, G. Dougas, M. Athanasiou, M. Detsis, A. Baka, T. Lytras, K. Mellou, S. Bonovas, and T. Panagiotopoulos, “Outbreak of West Nile Virus Infection in Greece, 2010,” *Emerging Infectious Diseases*, vol. 17, pp. 1868–1872, 2011.

[2] D. Pervanidou, M. Detsis, K. Danis, K. Mellou, E. Papanikolaou, I. Terzaki, A. Baka, L. Veneti, A. Vakali, G. Dougas, and others, “West Nile virus outbreak in humans, Greece, 2012: third consecutive year of local transmission,” *Euro Surveill*, vol. 19, pp. pii20758, 2014.

[3] R. Rosà, G. Marini, L. Bolzoni, M. Neteler, M. Metz, L. Delucchi, E. Chadwick, L. Balbo, A. Mosca, M. Giacobini, and others, “Early warning of West Nile virus mosquito vector: climate and land use models successfully explain phenology and abundance of Culex pipiens mosquitoes in north-western Italy,” *Parasites & vectors*, vol. 7, pp. 269, 2014.

[4] E. Izquierdo-Verdiguier and R. Zurita-Milla, “Introducing co-clustering for hyperspectral image analysis,” in *2015 IEEE - IGARSS*, July 2015, pp. 465–468.

[5] X. Wu, R. Zurita-Milla, and M. Kraak, “Co-clustering geo-referenced time series: exploring spatio-temporal patterns in

[3] https://ecdc.europa.eu/en/west-nile-fever/surveillance-and-disease-data/disease-data-ecdc

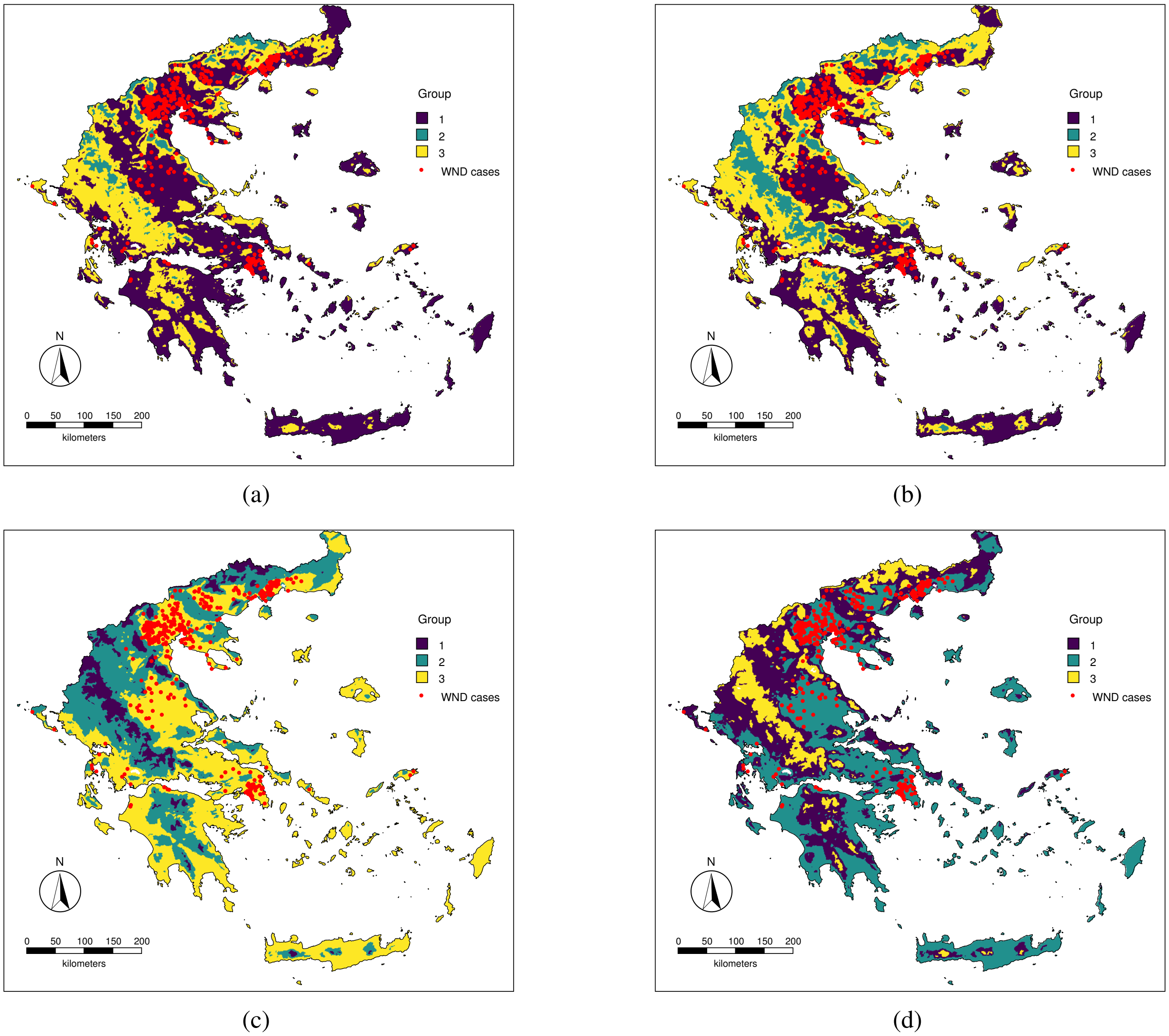

Fig. 1. Spatial co-clusters for the variables (a) number of summer days, (b) average maximum LST, (c) number of potential mosquito cycles and (d) number of potential WNV extrinsic incubation periods (EIP).

Dutch temperature data," *Int. J. of Geographical Information Science*, vol. 29, no. 4, pp. 624–642, Apr. 2015.

[6] X. Wu, R. Zurita-Milla, and M. Kraak, "A novel analysis of spring phenological patterns over Europe based on co-clustering: Co-Clustering European Spring Phenology," *J. of Geophysical Research: Biogeosciences*, vol. 121, no. 6, pp. 1434–1448, June 2016.

[7] V. Andreo, M. Metz, M. Neteler, R. Rosà, M. Marcantonio, C. Billinis, A. Rizzoli, and A. Papa, "Can reconstructed land surface temperature data from space predict a west nile virus outbreak?," *ISPRS - Int. Arc. of the Photogrammetry, Remote Sensing and Spatial Inf. Sc.*, vol. XLII-4/W2, pp. 19–26, 2017.

[8] M. Metz, V. Andreo, and M. Neteler, "A new fully gap-free time series of land surface temperature from modis lst data," *Remote Sensing*, vol. 9, no. 12, pp. 1333, 2017.

[9] G. Valiakos, A. Touloudi, L. V. Athanasiou, A. Giannakopoulos, C. Iacovakis, P. Birtsas, V. Spyrou, Z. Dalabiras, L. Petrovska, and C. Billinis, "Serological and molecular investigation into the role of wild birds in the epidemiology of West Nile virus in Greece," *Virology journal*, vol. 9, pp. 266, 2012.

[10] T. Groen, G. LAmbert, R. Bellini, A. Chaskopoulou, D. Petric, M. Zgomba, L. Marrama, and D. Bicout, "Ecology of West Nile virus across four European countries: empirical modelling of the Culex pipiens abundance dynamics as a function of weather," *Parasites & Vectors*, vol. 10, pp. 524, 2017.

[11] X. Wu, R. Zurita-Milla, E. Izquierdo Verdiguier, and M. Kraak, "Triclustering Georeferenced Time Series for Analyzing Patterns of Intra-Annual Variability in Temperature," *Annals of the American Association of Geographers*, pp. 1–17, June 2017.