Regionalizing & Partitioning Africa's Coronavirus (COVID-19) Fatalities Using Environmental Factors and Underlying Health Conditions for Social-economic Impacts

Alaba Boluwade
Dept. of Soil, Water & Agricultural Engineering
Sultan Qaboos University
Muscat, Oman
alaba@squ.edu.om

***Abstract*— The COVID-19 event was unexpected and has had shocking impacts such as widespread economic losses and tens of thousands of deaths. The COVID-19 infection rate is relatively low in Africa compared to other continents, but the number of cases is rising. As of July 12, 2020, in Africa, there are a total of 13,194 deaths and 591,153 reported cases. The dynamics of this pandemic spread are relatively unknown; however, previous studies have established a relationship between poor air quality standards due to nitrogen dioxide (NO_2) and fine particulate matter (PM2.5) and COVID-19 deaths and cases. Meanwhile, other studies have linked pre-existing health conditions from cardiovascular diseases with COVID-19 fatalities. However, none of these studies have examined these indicators from socio-economic and strategic planning perspectives. The primary aim of this paper is to combine and cluster these two air qualities indicators, pre-existing heart conditions due to morbidity and mortality from cardiovascular disease (MMDC), the probability from dying from four main (cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes) non-communicable diseases (NCDs) using a self-organizing map (SOM) and the hierarchical clustering method (HCM). Using SOM and HCM, all the variables mentioned above were partitioned into five clusters that did not follow the geographical boundaries of five regions in Africa. The results show that the countries with the highest COVID-19 deaths and cases as of 12 July 2020 are Egypt (3769 and 81,158) and South Africa (3971 and 264,184). The SOM technique was successfully used to combine these two countries into a single cluster. Notably, these two countries also have high rates of pre-existing health conditions (MMDC, NCDs), poor air quality indicators (NO2 and PM2.5) and pollution levels. Since no single country can manage this pandemic alone, a concerted effort is needed to mitigate and combat this virus. Therefore, relating these indicators together at the continental level would help improve state-of-the-art planning and management of the COVID-19 pandemic in Africa.**

Keywords*— PM2.5, Nitrogen Dioxide (NO_2), Clustering, Self-organizing maps; COVID-19, Heart Disease, Africa Union

I. INTRODUCTION

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), 2019 (hereafter: COVID-19) first appeared around December, 2019 Wuhan City, China, and the first reported case in Africa was in Egypt on 14 February 2020 [1]. Since then, there has been a daily rise in both the number of cases and associated fatalities. As of July 12, 2020, the total numbers of cases and deaths in Africa are 591,153 and 13,194 respectively. While the dynamics of this pandemic are not fully clear, studies have associated poor air quality, such as air with high levels of fine particulate matter (PM2.5) and tropospheric nitrogen dioxide (NO_2), with COVID-19 fatalities [2,3,4,5,6]. What all these studies have in common is their conclusion that these air pollutants can exacerbate COVID-19 fatalities. Prior studies have also associated pre-existing underlying health conditions with COVID-19 fatalities. Li et al. [7] reported that cardiovascular metabolic comorbidities have a positive association with COVID-19 fatalities, thereby making the patient more susceptible to the effects of the virus.

In Africa, air quality indicators are often above acceptable standards, and pre-existing health conditions such as coronary artery disease are growing [9]. According to [9], in 2013, there were approximately 1 million deaths due to cardiovascular disease in sub-Saharan Africa. The African continent is a heterogeneous in terms of landscape, race, language, and population perspectives. One common denominator, however, is that it is a growing continent from an economic perspective. Therefore, no single country can combat this pandemic alone. Information and knowledge sharing, especially among member states that have similar patterns of both poor air quality and pre-existing health issues are paramount. This is possible through the regionalization, partitioning, and clustering of these several important indicators associated with COVID-19. The self-organizing map (SOM) technique [10, 11] is a neural network process that can be used to visualize relationships and similarities in high-dimension multivariate datasets using a 2-dimensional grid. This process will help partition Africa into several classes or clusters (using the hierarchical algorithm) while optimizing the sum of square deviation (SSD, i.e. heterogeneity) in each cluster. The SOM and clustering approaches have been used in several disease clustering-related applications such as in [12, 13, 14].

Regionalizing and partitioning Africa into a number of clusters and classes using NO_2, PM2.5, pre-existing heart diseases, and COVID-19 cases and deaths will provide a better understanding of the spatial pattern, association and heterogeneity (homogeneity) of these indicators in Africa. Therefore, with all these important explanatory variables, this study aims to answer a pertinent question: How many contiguous clusters can the African continent be divided into from a COVID-19 management perspective? The results of this study will be useful for national, regional, and international organizations such as the African Union (AU) and the World Health Organization (WHO) to improve the state-of-the-art management of the COVID-19 pandemic in Africa. This result will help decision makers such as the WHO, AU and regional organizations such as the East African Community (EAC), Southern African Development Community (SADC), Economic Community of West African States (ECOWAS), Intergovernmental Authority on Development (IGAD), Economic Community of Central African States (ECCAS), and Economic Community of Central African States (ECCAS), which share similar customs, cultures and geopolitical interests.

II. METHODOLOGY

A. Input Data

The input variables used for this study are extracted from various sources. The number of COVID-19 cases and deaths are extracted from the WHO dashboard (https://covid19.who.int/) as of July 12, 2020. For the entire African continent, the total number of cumulative deaths and cases are 13,194, and 591,153 respectively. The premature non-communicable disease (NCD) mortality rate is also extracted from the WHO portal (https://aho.afro.who.int/) (the most recent year is 2016). NCD is the probability of dying (%) from four main diseases: cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes. The mortality and morbidity due to cardiovascular diseases (MMCD) is also extracted from the [15] web portal (https://ourworldindata.org/burden-of-disease). MMCD is measured in "Disability Adjusted Life Years" (DALYs). For instance, one DALYs will equal to the loss of one year in good health due to disease, disability or premature death [15].

All the environmental variables were obtained from the National Aeronautics and Space Administration (NASA) web portal. PM2.5 is the monthly (January–April 2020) average dust surface column mass concentration extracted from MERRA-2 global data, which has a spatial resolution of 0.5^ox0.625^o [16]. NO_2 is the tropospheric NO_2 column (30% cloud screened), which is also obtained from the NASA portal. This can be defined as the indication of air pollution from fossil fuel burned due to the transportation and generation of electricity in January–April 2020. The version of this study is version 3: *OMI/Aura NO_2 Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25^o x 0.25^o* [17].

B. SOM and Hierarchical Clustering

The SOM method is an unsupervised neural network technique developed by [10]. It can be used to analyze the topology of multivariate data. In other words, the primary goal of the SOM method is to reduce higher-dimensional vector Z datasets (i.e. NO_2, PM2.5, MMCD, NCD, and COVID-19 deaths and cases) into a low 2-D grid discrete representation. This grid consists of an array of neurons or nodes arranged either hexagonally or in a rectangular shape. The dimensional reduction to the 2-D space is done without knowing (unsupervised) the structure of the input variables mentioned above. The SOM is a competitive and iterative learning procedure where, at each iteration of the process, a sample (sometimes called the training data) of vector Z of (n dimension) is chosen randomly, and the Euclidean distances between Z and all the weight of the vectors of the SOM's nodes (n dimension) are estimated. In other words, the "winning unit", also called the "Best Matching Unit" (BMU, i.e. the node with the weight vector) with the shortest distance to the input vector Z is selected [18], and within a certain radius, all the neighboring nodes' positions are adjusted simultaneously. The weight of the nodes is updated, and this process is repeated until all the nodes in the grid have been visited and completed. The final result will have all the homogenous nodes contiguously grouped, and non-resembling nodes will be in a different area. More information about this process can be found in [10, 11]. The hierarchical clustering technique can be used to cluster the derived node weight vectors. The process starts by treating each node weight as a separate cluster and then merging two clusters that are close together, which continues until all the clusters have been merged. The SOM and hierarchical clustering algorithms have been implemented in R statistical software under different packages. The Kohonen package [19, 20] in R software was used in this study.

III. RESULTS

Table 1 shows the cross-correlation between the variables. There is a correlation of ~ 0.7 and 0.42 between MMCD and COVID-19 deaths and between MMCD and COVID-19 cases respectively. This shows that, throughout Africa, there are relationships between COVID-19 deaths, morbidity and mortality due to cardiovascular disease and the probability of dying from four main non-communicable diseases. This shows that these underlying health conditions are contributing to COVID-19 deaths. There is also a correlation between 0.45 between NO_2 and NCDs. It shows that in countries where the air quality is poor due to fossil fuel burned for transportation purposes and the generation of electricity, there is a high probability of dying from the four main NCDs. NO_2 also has weak positive correlations of 0.15, 0.30, and 0.30 with MMCD and COVID-19 deaths and cases respectively. These relationships are very important, as they indicate that the increase in NO2 could also lead to an increase in both COVID-19 cases and deaths. There is also a weak positive correlation (0.14) between PM2.5 and MMCD. Notably, the table also shows that in areas where there are high COVID-19 cases, there are also high COVID-19 deaths.

Figure 1 shows the SOM in a 2-D grid for all the variables used in this study. These are heatmaps showing the variability and distribution of each variable across the entire grid. Figure 1 shows that (a) MMCD, (b) NO_2, (c) NCDs and (d) PM2.5 have a systematic pattern. On the other hand, it is clear in Figure 1 that (e) COVID-19 deaths and (f) COVID-19 cases have the same distribution. The distribution of the air quality parameters having a slightly similar pattern is expected since they are governed by the same atmospheric processes. It is also expected to see that MMDC and NCDs show a slightly similar system pattern with air quality indicators because of the weak positive cross-correlations mentioned above. It is also not surprising to see that COVID-19 deaths and cases have a similar distribution. This is also due to their high positive correlation and association. It is interesting to see that the nodes have high (low) values in a similar position in both Figures (e) and (f).

TABLE I. CORRELATION OF COVID-19 DEATHS & ACTIVE CASES, ENVIRONMENTAL VARIABLES, AND PRE-EXISTING HEART DISEASES.

	MMCD	*NCDs*	*COVID-19 DEATHS*	*COVID-19 CASES*	*NO_2*	*PM2.5*
MMCD	1.00					
NCDs	0.04	1.00				
DEATHS	0.70	0.21	1.00			
CASES	0.4	0.22	0.9	1.00		
NO_2	0.20	0.45	0.30	0.30	1.00	
PM2.5	0.14	0.08	0.02	-0.06	0.08	1.00

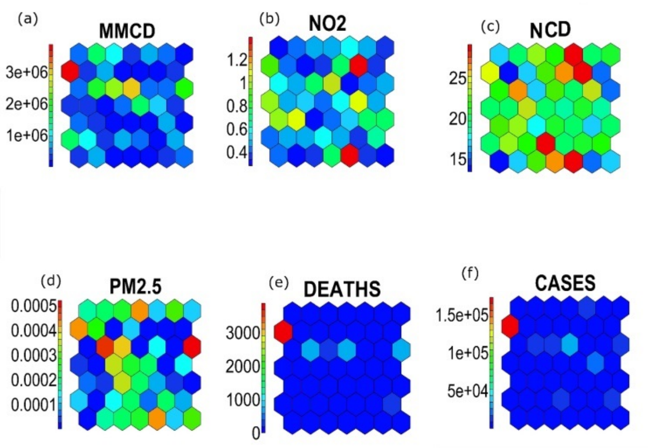

Figure 1: Self-organizing heat maps showing (a) mortality and morbidity cardiovascular disease (MMCD), (b) tropospheric nitrogen dioxide (NO2), (c) probability of dying from a non-communicable disease (NCD), (d) fine particulate matter (PM2.5) and (e) cardiovascular disease (mortality and morbidity)

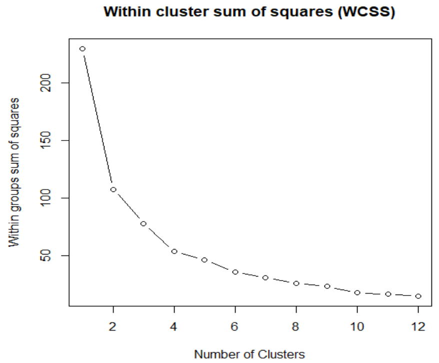

Figure 2: Number of potential clusters using a hierarchical clustering method.

Using the elbow method, Figure 2 shows that 5 clusters as "optimal" number of clusters. Figure 3 shows the locations of the five obtained partitions or clusters and the fan diagram showing the normalized magnitude of the input vectors in each node and cluster. These clusters were then mapped to the original data space. Figure 4 shows the derived clusters for each African member state.

Based on all these explanatory variables and COVID-19 cases and fatalities, the following interpretation can be derived from the derived clusters:

Cluster 1: (Egypt and South Africa: dominated by MMCD, NCDs, COVID-19 deaths and cases, NO_2). This cluster has all the dominant components of poor air quality indicators, pre-existing health conditions, and COVD-19 deaths and cases. This is the cluster to watch in Africa from a risk management perspective. As of July 16, Egypt has 3769 (COVID-19 deaths) and 81,158 (COVID-19 cases), while South Africa has 3971 (COVID-19 deaths) and 264,184 (COVID-19 cases). The dominant presence of MMCD, NCDs, NO_2, and PM2.5 in this cluster makes this an important hotspot in Africa. Apart from organizations such as the WHO or AU, this cluster would be of high relevance to Africa's regional organizations such as the SADC and IGAD.

Cluster 2: This cluster includes countries from the West Africa region, and it includes Nigeria, Niger, Burkina Faso, Mali, Senegal, Mauritania, and other countries such as Chad, Algeria, and West Sahara. This region has high air quality pollution due to PM2.5. In some nodes, there are also high MMCD, NCDs, and certain recorded COVID-19 deaths. The high MMDC in these nodes is a potential risk because of the positive correlation between MMDC and COVID-19 deaths. Therefore, the management of pre-existing heart disease and maintaining good atmospheric quality (i.e. reducing PM2.5) is paramount. This cluster would also concern regional institutions such as ECOWAS, IGAD, and ECCAS.

Cluster 3: This is mainly the eastern and southern parts of Africa. This region also has low COVID-19 deaths and cases. However, the PM2.5, MMDC, and NCD indicators dominate the southern, northern, and south-eastern parts of the cluster, respectively. These footprints show that this cluster is not immune to fatalities from the pandemic. The highest numbers of COVID-19 deaths and cases have been reported in the Democratic Republic of Congo (188 and 7979) and Kenya (184 and 9729). Other countries within this cluster have low COVID-19 (deaths and cases) such as Uganda (0 and 1013), Burundi (1 and 250), Rwanda (3 and 1299), etc. Regional institutions such as the SADC and EAC would find data on this cluster and its heterogeneity most useful.

Cluster 4: This cluster includes countries in West Africa such as Togo, Benin, Ghana, Guinea, and others such as Sudan, South Sudan, and Cameroon. This cluster has high levels of NCDs and NO_2. The COVID-19 cases and deaths are relatively low. For instance, Benin has 26 and 1,379 COVID-19 deaths and cases respectively, while Togo 15 and 719 COVID-19 deaths and cases respectively. ECOWAS, the EAC, and ECCAS are regional institutions affected by this cluster.

Cluster 5: This cluster (Lesotho, Swaziland, Central Africa Republic, Côte d'Ivoire, and Sierra Leone) is predominately dominated by NCDs and NO_2 and a small amount of PM2.5. This shows that poor quality indicators and the probability of dying from underlying diseases such as cardiovascular diseases, cancer, chronic respiratory diseases, and diabetes are high in this cluster. As of 12 July 2020, the COVID-19 deaths and cases in this cluster have been: Lesotho (1 and 184), Swaziland (18 and 1311), Central Africa Republic (53 and 4288), Côte d'Ivoire (81 and 12052), and Sierra Leone (63 and 1618). The SADC, ECOWAS, and ECCAS are the concerned regional institutions affected.

In recent decades, there have been both epidemic and pandemic incidents such as the outbreaks of severe acute respiratory syndrome-related coronavirus (SARS), highly pathogenic Asian avian influenza A (H5N1), influenza A virus subtype (A/H1N1), Middle East respiratory syndrome coronavirus (MERS-CoV), and Zaire ebolavirus (Ebola), all of which have had different timelines and fatalities. This shows that these outbreaks are likely to continue to occur in the future and no country or continent is immune in preventing this from happening. Therefore, the procedure developed in this study is a step in the right direction, especially for Africa, where other challenges could exacerbate the impacts of any pandemic. For further studies, it would be interesting to have a common platform in Africa where this result could be hosted and displayed in real time. In other words, a platform or web portal that could, in real time, extract air quality indicators and integrate them with recent pre-existing health conditions and COVID-19 deaths and cases would be of high interest to both regional and international institutions.

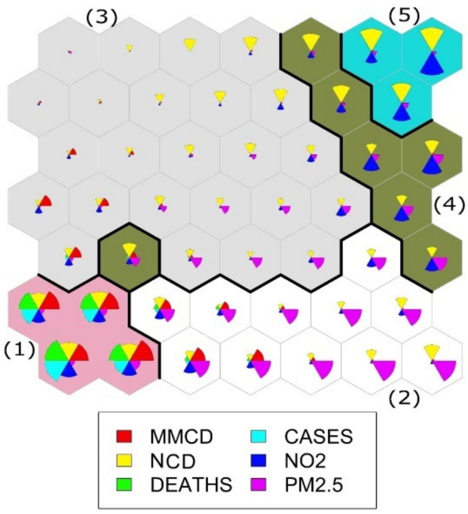

Figure 3: Star charts visualization of the normalized values of the original variables in a 2-dimensional grid. The height of each sector in a node signifies intensity. The black border line signifies clusters and segmentation using the self-organising map and hierarchical clustering methods.

IV. Conclusion

Past decades have witnessed epidemics and pandemics such as SARS, H5N1, A/H1N1, ERS-CoV, and Ebola; however, the COVID-19 pandemic and the associated widespread economic loss and high number of fatalities is an event that was not expected. The spread, death, and control of this coronavirus are not fully understood. Therefore, there is a need for pragmatic and proactive actions in combating this virus. Some studies have linked poor air quality indicators such as NO_2 and PM2.5 extracted from MERRA-2 remote sensing images. Other studies have looked at the impact of pre-existing underlying health conditions such as cardiovascular disease. However, no studies have been able to combine and quantify these factors from the socio-economic and strategic planning perspectives. No single country can fight this pandemic alone. Therefore, there is a need for collaboration and information sharing across various regions of Africa and other places in the world. This study will help improve the state-of-the-art planning and management of the COVID-19 pandemic in Africa. Using the SOM technique, this study partitioned the African continent into five administrative clusters based on air quality indicators such as NO_2, PM2.5, pre-existing health conditions (MMDC and NCDs), and COVID-19 deaths and cases. It is interesting to note that these five clusters did not follow the five geographical boundaries in Africa, meaning that a cross-regional approach should be taken in COVID-19 mitigation and control in Africa

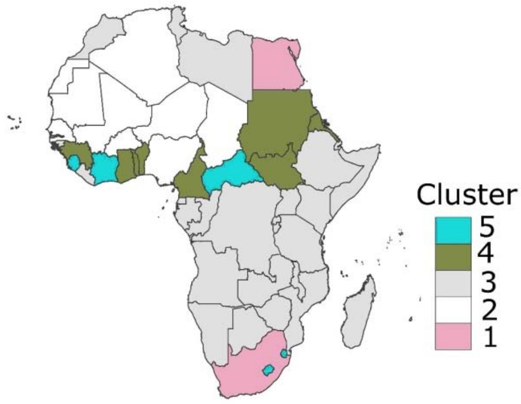

Figure 4: Map of Africa showing the five clusters

. The strength of the SOM technique lies in its ability to visualize high-dimensional multivariate datasets in a 2-D grid platform. The countries with the highest COVID-19 (deaths and cases) as of 12 July 2020 are Egypt (3769 and 81,158) and South Africa (3971 and 264,184). The SOM technique successfully combined these two countries into a single cluster. Notably, these two countries also have high pre-existing health conditions (MMDC and NCDs), poor air quality indicators (NO_2 and PM2.5), and high pollution levels.

References

[1] WHO, 2020 Timeline of WHO's response to COVID-19. https://www.who.int/news-room/detail/29-06-2020-covidtimeline. Accessed on July 12, 2020

[2] Ogen, Y. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) Fatality. Science of the Total Environment. 2020. 726: 138605

[3] Wu, C, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA. Intern. Med. 2020, 10.1001/jamainternmed.2020.0994

[4] Contini, D.; Costabile, F. Does Air Pollution Influence COVID-19 Outbreaks? Atmosphere 2020, 11, 377.

[5] Chan, K.H. J. S. Malik Peiris, S. Y. Lam, L. L. M. Poon,1 K. Y. Yuen, and W. H. Seto. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. 2011 https://doi.org/10.1155/2011/734690

[6] Sharma S, Zhang M, Anshika GJ, Zhang H, Kota SH. Effect of restricted emissions during COVID-19 on air quality in India. Sci Tot Environ 728:138878. https://doi.org/10.1016/j.scitotenv.2020.138878

[7] Wichmann, D et al. Autopsy Findings and Venous Thromboembolism in Patients With COVID-19. A Prospective Cohort Study. Annals of Internal Medicine. 2020. https://doi.org/10.7326/M20-2003

[8] Li, B., Yang, J., Zhao, F. et al. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res

Cardiol 109, 531–538 (2020). https://doi.org/10.1007/s00392-020-01626-9

[9] Keates AK, Mocumbi AO, Ntsekhe M, Sliwa K, Stewart S. Cardiovascular disease in Africa: epidemiological profile and challenges. Nat Rev Cardiol. 2017;14(5):273-293. doi:10.1038/nrcardio.2017.19

[10] Kohonen, T. Self-Organization and Associative Memory. Berlin: Springer-Verlag.

[11] Kohonen, T.. Self-Organizing Maps. Berlin: Springer-Verlag, 2011

[12] Lequerica-Fernández, P., Peña, I., Iglesias-Rodríguez, F.J. et al. Outcome prediction for salivary gland cancer using multivariate adaptative regression splines (MARS) and self-organizing maps (SOM). Neural Comput & Applic 32, 1219–1229 (2020). https://doi.org/10.1007/s00521-018-3473-6

[13] Crespo, R., Alvarez, C., Hernandez, I. et al. A spatially explicit analysis of chronic diseases in small areas: a case study of diabetes in Santiago, Chile. Int J Health Geogr 19, 24 (2020). https://doi.org/10.1186/s12942-020-00217-1

[14] Bulut, H., Onan, A. & Korukoğlu, S. An improved ant-based algorithm based on heaps merging and fuzzy c-means for clustering cancer gene expression data. Sādhanā 45, 160 (2020). https://doi.org/10.1007/s12046-020-01399-x

[15] Roser, M and H Ritchie 2016 - Burden of Disease. Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/burden-of-disease' [Online Resource]. (Accessed on 20 July 2020)

[16] Global Modeling and Assimilation Office (GMAO), MERRA-2 tavgM_2d_lnd_Nx: 2d,Monthly mean,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), 2015. Accessed on July 12, 2020, 10.5067/8S35XF81C28F

[17] Nickolay et al., 2019 Nickolay A. Krotkov, Lok N. Lamsal, Sergey V. Marchenko, Edward A. Celarier, Eric J.Bucsela, William H. Swartz, Joanna Joiner and the OMI core team (2019), OMI/Aura NO2 Cloud-Screened Total and Tropospheric Column L3 Global Gridded 0.25 degree x 0.25 degree V3, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed on July 12, 2020, 10.5067/Aura/OMI/DATA3007

[18] Ferentinou, Maria & Karymbalis, Efthimios & Charou, E. & Sakellariou, Michael. Using Self Organising Maps in Applied Geomorphology. 2011. 10.5772/13265.

[19] Wehrens R, Buydens LMC (2007). "Self- and Super-Organizing Maps in R: The kohonen Package." Journal of Statistical Software, 2007.21(5), 1–19. doi: 10.18637/jss.v021.i05

[20] Wehrens R, Kruisselbrink J. "Flexible Self-Organizing Maps in kohonen 3.0." Journal of Statistical Software, 87(7), 2018. 1–18. doi: 10.18637/jss.v087.i07.