Weekly Journal

Leila Uy

June 21, 2021

1 Literature Review

There is a growing part of research creating trajectories and looking at the effects of climate change using past and present data. An IPCC emissions scenario report created four storylines and 40 scenarios. One assessment looked at the combined impacts of climate change and acidification of agricultural crops in Asia. They found that the Indian subcontinent would have lower agricultural productivity due to projected temperature and precipitation. In contrast, regional climate change would benefit China, increasing their agricultural output, but high levels of acidic deposition would offset any possible beneficial impacts [8]. Due to this natural relationship between climatic variables and vegetation growth, the agricultural sector is extremely vulnerable to any changes witnessed by future climate change scenarios whether directly or indirectly.

Ideally, if the correlation between temperature and additional climatic variables were fixed over time, predicting the effects of climate change would be simple, but for example, the relationship between temperature and humidity varies over time due to the distribution, intensity, and frequency of precipitation [15]. These variables, in turn, rearrange the ecological suitability of the area for species including agricultural crops and invasive plant species (IPS). Studies suggest that IPS have a large potential to result in landscape homogeneity at ecoregional scales and compete with native plant species [14]. It is predicted that many exotic insect, weed, and pathogen species were responsible for a large portion of the \$130 billion losses estimated to be caused by pests each year in the United States' agricultural sector [1]. In order to predict the impacts of climate change on agriculture, there must be a way to explore the shifts in suitability for different species' ranges.

Geographical ecoregions are the identification of different regions that are similar in ecological and environmental conditions. They are important when estimating the extent of suitable habitat for particular species including agricultural crops and competing IPS. A simple search through academic papers on ecoregions will give you over a hundred results because there is no established classification system. A 2001 study by Olson et al. [9] used human experts to create a detailed terrestrial ecoregion classification that was better suited to identify areas of outstanding biodiversity. Another 2008 study by Ellis et al. [4] introduced eighteen anthropogenic biomes because previous global patterns of biodiversity either ignore or simplify the direct interaction between humans and ecosystems. A 2014 study that included a collaboration between Esri, USGS, AAG, and GEO [11], approached classification by stratifying the Earth into physically distinct areas with their associated land covers. Traditional methods of classification often delineated ecoregions by hand and as a consequence, results are often difficult to replicate, and introducing more variables and large data sets would take a long time.

As we grow into a new age with more accessible and higher resolution spatial data sets, it is important to create an accurate, quick, and flexible classification system for determining ecoregions. K-means clustering was independently discovered in different scientific fields as far back as the 1950s and remained popular to this day due to its simple implementation which has introduced several extensions many years after its introduction [2, 3, 6, 10, 12, 13]. A recent implementation of this multivariate statistical clustering algorithm is the integration of parallel programming and computational optimization to speed up the process. Hargrove and Hoffman [5] developed a parallel multivariate geographic clustering algorithm in C using the Message Passing Interface (MPI). Rather than relying on expertise for classification, individual raster cells were placed in a data space where each dimension represented an environmental characteristic. So, if you were to look at precipitation, mean temperature, and wind speeds, each individual raster cell would have a coordinate based on the three variables (e.g. $u_1 = (0, 1, 0)$). This

implementation was later optimized by Kumar et al. [7] by using the triangular inequality to remove unnecessary calculations.

The focus of the work reported here is the design and implementation of a parallel K-means clustering algorithm in R that is easily adjusted for n variables, based on previous literature in parallel clustering and ecoregions. Utilizing the computational powers of AWS EC2 machines and large historical and future climate data provided by WorldClim, we are testing the accuracy, speed, and flexibility of the algorithm as a classification system for predicting shifts in ecoregions as a result of climate change.

References

[1] Raymond Carruthers. Invasive species research in the united states department of agriculture-agricultural research service. *Pest Management Science*, 55(6–7):827–834, 2003.

[2] Salvatore Cuomo, Vincenzo De Angelis, Gennaro Farina, Livia Marcellino, and Gerardo Toraldo. A gpu-accelerated parallel k-means algorithm. *Computers & electrical engineering*, 75:262–274, 2019.

[3] James Churchill Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of cybernetics*, 3(3):32–57, 1973.

[4] Erle C Ellis and Navin Ramankutty. Putting people in the map: anthropogenic biomes of the world. *Frontiers in Ecology and the Environment*, 6(8):439–447, 2008.

[5] William Hargrove and Forrest Hoffman. Using multivariate clustering to characterize ecoregion borders. *Computing in science & engineering*, 1(4):18–25, 1999.

[6] Anil Jain. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8):651–666, 2010.

[7] Jitendra Kumar, Richard Mills, Forrest Hoffman, and William Hargrove. Parallel k-means clustering for quantitative ecoregion delineation using large data sets. *Procedia computer science*, 4:1602–1611, 2011.

[8] Nebojsa Nakicenovic and Rob Swart. *Emissions Scenerio*. Cambridge University Press, 2000.

[9] David M Olson, Eric Dinerstein, Eric D Wikramanayake, Neil D Burgess, George VN Powell, Emma C Underwood, Jennifer A D'amico, Illanga Itoua, Holly E Strand, John C Morrison, et al. Terrestrial ecoregions of the world: A new map of life on earth. *BioScience*, 51(11):933–938, 2001.

[10] Dan Pelleg and Andrew Moore. Accelerating exact k -means algorithms with geometric reasoning. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 277–281. Association for Computing Machinery, 1999.

[11] Roger Sayre, Jack Dangermond, Charlie Frye, Randy Vaughan, Peter Aniello, Sean Breyer, Douglas Cribbs, Dabney Hopkins, Richard Nauman, William Derrenbacher, Dawn Wright, Clint Brown, Charles Convis, Jonathan Smith, Laurence Benson, D. Paco VanSistine, Harumi Warner, Jill Cress, Jeffery Danielson, Sharon Hamann, Tom Cecere, Ashwan Reddy, Devon Burton, Andrea Grosse, Diane True, Marc Metzger, Jens Hartmann, Nils Moosdorf, Hans Dürr, Marc Paganini, Pierre DeFourny, Olivier Arino, Simone Maynard, Mark Anderson, and Patrick Comer. *A New Map of Global Ecological Land Units — An Ecophysiographic Stratification Approach*. Association of American Geographers, 2014.

[12] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

[13] Zhuo Tang, Kunkun Liu, Jinbo Xiao, Li Yang, and Zheng Xiao. A parallel k-means clustering algorithm based on redundance elimination and extreme points optimization employing mapreduce. *Concurrency and computation*, 29(20), 2017.

[14] Chun-Jing Wang, Qiang-Feng Li, and Ji-Zhong Wan. Potential invasive plant expansion in global ecoregions under climate change. *PeerJ*, 7:e6479, 2019.

[15] Peng Zhang, Junjie Zhang, and Minpeng Chen. Economic impacts of climate change on agriculture: The importance of additional climatic variables other than temperature and precipitation. *Journal of environmental economics and management*, 83:8–31, 2017.