COMP 767 Final Project Proposal

Michael Huang, Gabriela Moisescu-Pareja and Kangrui Ren

March 16, 2021

1 Introduction

The effects of climate change are anticipated to be far-reaching in the coming decades and centuries. One aspect of this can be appreciated when looking at the impact on specific biomes, whose characteristics are critically dependent on the climate. Climate classification systems (CCS) include a variety of ways to categorize the world's climates by grouping climates that are similar according to certain variables. These climate categories may exhibit a strong correlation with a biome category due to the climate playing a large role in the make-up of life in a given region. Using historical data, one can classify these climates to visualize how they have evolved through time. Knowing the climate type of a region can inform the possible vegetation that can grow there. Moreover, being able to model this evolution through time (predicting future scenarios) would be a valuable aid in decision-making with regards to conservation efforts and raising awareness about how we can expect our local regions to change. Planning for the future is an important part of adaptation.

2 Proposed Methodology

Gardner et al. proposed a climate classification system (CCS) using physiological variables most relevant to plant biology rather than the simpler criteria used in classical methods [3]. Classical methods, such as Koppen-Geiger (KG) and Koppen-Trewartha (KT) use fewer variables and simpler measures, for example, KT classifies Boreal as a region with 1-3 months with mean temperatures above 10°C.

The data. For their CCS, Gardner et al. use 10 physiologically relevant variables: (a) soil water content during the growing season; (b) mean growing season temperature; (c) growing season precipitation; (d) total summer precipitation; (e) total annual precipitation; (f) growing season length; (g) maximum temperature during the growing season; (h) mean annual temperature; (i) mean summer temperature and (j) summer soil water content. They construct the 10 variables themselves using climate data gathered from National Oceanic and Atmospheric Association (NOAA). The data from the NOAA was gathered at a 6-hourly temporal resolution and 2.5 spatial resolution (2000-2017) for the following variables: (a) surface skin temperature; (b) air temperature; (c) relative humidity; (d) net short-wave radiation; (e) downward short-wave radiation; (f) net long-wave radiation; (g) wind speed; (h) volumetric soil moisture (0–10cm below ground level) and (i) water equivalent of snow depth. Finally, they also used NOAA precipitation data.

The machine learning method. Gardner et al.'s CCS method was to use Principal Components Analysis (PCA) on the 10 physiologically relevant variables, and performed K-means clustering using the top two principal components [3]. We propose using a *Variational Autoencoder* [7] (VAE) to perform unsupervised clustering instead of the PCA and K-means clustering combination as performed by Gardner et al. The general idea of our method is to take climate variables that are physiologically relevant to plants, such as those used by Gardner et al., as the input vector to our VAE, which we wish to reconstruct. The encoding layer would include a softmax function, so that we can treat the embedding similar to categorical variables, or as topics as described in Embedded Topic Modeling from the NLP domain [2].

Climate classification into the future. The above method achieves the goal of creating a VAE-based CCS for the present day. An additional objective of this project would be to visualize changes in CCS clusters in the future. This has already been done for KG classification [1], [9], where they used projected future conditions under climate change to visualize changes in biome classifications. However, there may be some physiologically relevant variables that are not projected into the future, that need to be excluded from the intial VAE or imputed using other linear methods. Ideally, we would want to find future projected data for all the variables gathered from the NOAA, which is necessary to construct the 10 physiologically relevant variables. It might be possible to find some of this data projected into the future. At least one possible variable for which we could possible gather future predictions includes relative humidity [11], though more research would need to be done to find data relating to the other variables.

3 Additional Considerations

Availability of high-quality historical data in all countries In the most basic aspect, the acquisition of high-quality data should be taken into consideration. The meaning of high quality here is basic and credible (even from an official authority). There are quite a few organizations that can provide such data, such as [5] stated that gridded data are generated and used to derive time series of national and regional averages. Phenological observations and radiosonde data are also part of the data base. These data might be helpful for constructing the model. However, not all regions could provide effective and sufficient data as these organizations do. Quite a few regions fail to set up high-level research institutions due to economic reasons, which leads to lack of data and low credibility. At the same time, these regions have very abundant natural data, so this is a severe problem for data collection.

Uncertainties in predicting future biome changes The uncertainties in predicting future biome changes is also a question that need to be carefully considered. Not only the model parameters but also the predicted results can cause uncertainty. [10] stated that Imperfect process knowledge and limited observational data restrict the possibility to parameterize model adequately and potentially contribute to uncertainty in model results. [8] indicated that simulated biome changes lacked consistent large-scale geographical patterns of change across scenarios. Because of the large uncertainties in future projections, adaptation strategies must be highly flexible, sometimes small factors need to be carefully and deeply studied (stated in next paragraph). Additionally, a large portion of this avenue of inquiry relies on the advancement of accurate climate models at fine spacial and temporal resolutions (see climate downscaling, microclimate models [6]).

Existence of microscopic factors that are easily ignored Normally, we only consider factors that are apparent in climate change. However, there might be some invisible factors or indirect factors caused by precipitation, illumination, temperature, etc. that affect biome changes in microscopic ways. [4] indicated that a large number of studies document substantial impact of solar UV radiation on individual species, yet considerable uncertainty remains with respect to assessing impacts on ecosystems. Factors such as ultraviolet radiation are often not considered, and it is difficult to include them even in the collected data. As a result, survey should be carefully conducted in choosing factors in the model.

Impact This work falls under the broad umbrella of aiming to predict the various secondary effects of climate change. For another problem along these lines, consider predicting sea level rises in various regions of the world. Trying to anticipate the effect of climate change on vegetation is another such problem, as it has deep effects on the make-up of biomes (whose full evolution would be too complex to predict), as well as having possible effects in agriculture (due to the climate dictating what can possibly grow in any given area). The former would relate to conservation efforts, and the latter would relate to issues of food security. This information would be valuable to decision-makers.

References

[1] Hylke E Beck et al. "Present and future Köppen-Geiger climate classification maps at 1-km resolution". In: *Scientific Data* 5.1 (2018), pp. 1–12.

[2] Adji B Dieng, Francisco JR Ruiz, and David M Blei. "Topic modeling in embedding spaces". In: *Transactions of the Association for Computational Linguistics* 8 (2020), pp. 439–453.

[3] Alexandra S Gardner, Ilya MD Maclean, and Kevin J Gaston. "A new system to classify global climate zones based on plant physiology and using high temporal resolution climate data". In: *Journal of Biogeography* 47.10 (2020), pp. 2091–2101.

[4] Donat-Peter Häder et al. "Aquatic ecosystems: effects of solar ultraviolet radiation and interactions with other climatic change factors". In: *Photochemical & Photobiological Sciences* 2.1 (2003), pp. 39–50.

[5] F Kaspar et al. "Monitoring of climate change in Germany–data, products and services of Germany's National Climate Data Centre". In: *Advances in Science and Research* 10.1 (2013), pp. 99–106.

[6] Michael R. Kearney et al. "A method for computing hourly, historical, terrain-corrected microclimate anywhere on earth". In: *Methods in Ecology and Evolution* 11.1 (2020), pp. 38–43. DOI: https://doi.org/10.1111/2041-210X.13330.

[7] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes". In: *arXiv preprint arXiv:1312.6114* (2013).

[8] Carola Martens et al. "Large uncertainties in future biome changes in Africa call for flexible climate adaptation strategies". In: *Global Change Biology* 27.2 (2021), pp. 340–358.

[9] Franz Ruben and Markus Kotter. "Observed and projected climate shifts 1901-2100 depicted by world maps of the Köppen-Geiger climate classification". In: *Meteorologische Zeitschrift* 19.2 (2010), pp. 135–141.

[10] Anna Wramneby et al. "Parameter uncertainties in the modelling of vegetation dynamics—effects on tree community structure and ecosystem functioning in European forest biomes". In: *Ecological Modelling* 216.3-4 (2008), pp. 277–290.

[11] Lei Zhao et al. "Global multi-model projections of local urban climates". In: *Nature Climate Change* 11 (2021), pp. 152–157.