Hi Quantum Black partners,

Thank you for collaborating with us on our drought prediction project. As promised in our meeting on Apr 20, I am sending this email on behalf of our team to provide a summary report on our current progress and our roadmap for the remainder of the project. We welcome your feedback and any additional ideas you may have. Additionally, we have attached our research paper, with the introduction and literature review sections written up, for your reference.

In the past several weeks, we have implemented different types of models.

combined available datasets and now are ready to proceed to modeling step. We finished feature extraction from the satellite images, including the change of the water reservoir and vegetation band. We started to build a set of simple, auto-regressive models and discovered the regional variability in predictability is an exciting direction to delve deeper. After all, this approach fits our story of discussing the change in predictability of droughts when we add layers of knowledge (i.e. features). Our immediate next step is to systematically apply OLS model with different combinations of features, for our nine climate regions.

We have received extensive support from our teaching fellow, Zona, and our head instructor, Dr. Tanner. They provided tremendous help in forming our project goal. We are looking forward to presenting you our updated modeling results next week. If you were to continue reading, the following two sections walk through our progress and immediate future plans one by one.

**Current progress:**

* *Data Aggregation.* We successfully aggregate available datasets including SET (soil evaporation and runoff), SMAP (soil moisture), NOAA (temperature and precipitation), Sentinel L2 (water reservoir and vegetation band), and USDM (drought index). The full dataset is now available in the weekly and county levels. However, the imbalance in the dataset is worth attention. We acknowledged that these subsets are not available throughout the whole time range, and we plan to address the predictability in relation to the imbalance.
* *Modeling and Improvements.* We implemented a baseline, autoregressive model only using the USDM variable (response variable) to understand the predictability issues. We first started with an ensembling model, aggregating the data for the whole country and making predictions. The results are not promising per our selected metrics MSE. We then tried the prediction with more granularity. That is, we fit one model for each climate region in the CONUS. We observed variability in MSE in both spatial and temporal senses, and we believed these variabilities are explainable given the county size and data/feature availability. Therefore, we continue the exploration following this regional perspective.

**Next steps:**

* *Modeling.* The first set of models we will implement is OLS. In sum, we have 4 endogenous variables and will result in 10 different combinations. There are 9 regions in the CONUS, in total, we will have 90 OLS models. Since OLS is easy to code, we won’t expect the job to be a hard one but we will put more emphasis on the analysis of the results. Another thing we will try to explore is the choice of using a categorical target or a numerical target. Until now we are using a pre-processed, numerical drought index computed from 5 drought levels. We are curious about whether using the original, categorical response variable will significantly ease or improve the prediction task. The above task shall be carried out before the next group meeting (Apr 5). Finally, we will discuss based on the preliminary result if a more advanced ML model shall be used to enhance predictability for some models.
* *Documentation.* It is good merit to take notes while we are developing the project. From now on, we will keep a log and update the documentation regularly.

Again, we wanted to thank you for your enthusiasm and support for the project! Please let us know if you have any suggestions or comments regarding the summary above, and we look forward to keeping in contact with you throughout the remainder of the semester.

Sincerely yours,

Yujie, Jim, Michael, Thee