Note: We hold weekly meetings with our active partners, Francesca Sogaro and Vidur Nayyar. Francesca and Vidur are data scientists at Quantum Black. Additionally, we maintain email communication if there is an urgent need for discussion. Still, we want to present this summary report as a rather detailed document, just to guarantee that our partners have an understanding of our progress thus far.

Hi Quantum Black partners,

Thank you for collaborating with us on our drought prediction project. As promised in our meeting on Mach 2nd, 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 examined available datasets and read relevant literature in drought prediction. As we gained more knowledge through exploration, we refined our scope of work and the problem statement we plan to address. We plan to present middle-to-long term drought risk prediction methods in the Contiguous United States using climate features and satellite images in the form of a research paper. Our immediate next step, which we plan to complete before Harvard's spring recess, is to finish the data processing stage and begin the discussion of modeling choices.

We have received extensive support from our teaching fellow, Zona. She suggested possible ways of satellite images processing, and we are looking forward to discussing them with the domain expert you mentioned in next week's meeting. The following two sections walk through our progress and immediate future plans one by one.

Current progress:

- *Problem statement and the scope of work.* Following our exploratory data exploration, we concluded that we shall attempt to predict middle-to-long term drought risk in the CONUS. For our predictions, we will use weekly county-level data wherever possible from 2012 to 2022. We plan to predict drought risk in 1-, 2-, 4-, and 8-week intervals. This interval choice extends into and past average time intervals in the literature we reviewed. We will just focus on the CONUS given the availability of data, consideration of time constraints, and the complex nature of drought prediction tasks.
- *Data exploration.* We have accessed and started to process four datasets in detail. We will use the US Drought Monitor (USDM) as our official measure of drought risk and the response variable dataset in our prediction model. The independent variables we chose are Soil Moisture Active Passive (SMAP), which measures soil moisture, SET Evaporation data, which measures surface and subsurface runoff of water, and Sentinel

L2 satellite images. According to the literature, we believe these variables are related to drought risk, and therefore we chose to include them in our model.

Next steps:

- *Data Aggregation.* Now that we found accessible data, we plan to start aggregating data into a weekly, county-level data frame. There is an exception for Sentinel L2 data due to website request limits, we will find a way to create a paid account or create multiple accounts to download data parallelly. This part is expected to be done before spring recess (3/12).
- *Modeling Direction.* As we continue exploring and assessing possible data aggregations, we will begin better understanding modeling decisions. Our original plan is to use all independent variables in a single model. If the aggregation is more complex, however, we will have each person be responsible for a dataset, and develop four independent models which we will then compare. Depending on our progress, we can then discuss potential combinations of datasets. This part is expected to be done before the spring recess (3/12).
- *Baseline Modeling.* Once we decide on our modeling choice, we will start building our baseline models during spring recess (3/12-3/20). For a time series problem, we will try auto-regressive models initially, and generally add more features into the model over time.
- *External Help.* We plan to consult Quantum Black's external domain expert for satellite images processing. The meeting is scheduled for 3/9.

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