Transition Plan

1. Overview of Project
   1. Name of Project and Description:

GPP and RECO Flux Tower Data Optimization

This project involves creating a new calibration process for a climate change model that deals with different plant types and soil moisture. The model gets its test data from eddy flux towers located across the globe, as well as other scientific data. This improved calibration process will allow the user to tweak the model’s inputs and graph outputs to more accurately reflect the tower data. The sponsor of this project is the Numerical Terradynamic Simulation Group (NTSG), a group that researches tech development in certain Earth satellite data. The sponsor’s mission statement is "our primary mission is to develop capabilities to quantitatively describe the structure and function of ecosystems from regional to global scales using emerging technologies in satellite remote sensing, computational modeling and biophysical theory."

This project deals with the current calibration process of this model data, as the current method does not allow for a visual comparison of the model output to the real-life data. In addition to solving this problem, the software will allow the user to optimize the GPP and RECO parameters of the model during this visualization. The improved calibration will help further the ecological understanding of climate change effects on plants and soil moisture. A challenge of creating the software will be fully understanding the entire workflow, as it requires understanding of complex ecological and meteorological concepts. NTSG will benefit by streamlining the workflow and having an easily changeable codebase to improve model calibration for others to learn and continue to grow.

* 1. Major Stakeholders and their roles
* The client (Arthur Endsley) is the main individual, or primary stakeholder, that will be using this software. Currently, the client uses the previously described calibration process at NTSG. The project will improve the job experience by being able to visually compare the model output to actual data and edit the model parameters.
* NTSG employees are the secondary stakeholders or potential primary stakeholders. They will indirectly and directly use this software. The program will improve user experience, by providing a GUI to perform the calibration process. Employees will also have to view and compare the output of the model to the real-life data, which will be calibrated correctly with help of the program.
* NTSG will benefit from this program, because the new calibration process will improve the workflow by improving model data by packaging and improving the outputs. This company is a NASA-funded company so the data will be used by other NASA-funded companies and researchers.
* Other scientists will be dealing with the output of the model. This data can be used by more complicated models that scientists develop. This will further improve the understanding of climate change on soil moisture and plant life which will allow for the prediction of ecological effects.
  1. Status of requirements

All the high, a majority of medium, and all the low priority requirements listed in the Requirements Specifications document, within the “ProjectDocs” folder, are completed. The uncompleted medium priority requirements are the final calculations and reports of RMSE and NEE as well as running the spin-ups from 2000-2019. Although the project is mostly coded, the team was only able to develop through the execution of the numerical spin-ups part of the calibration process.

* 1. Suggested next steps

Apart from needing the three medium priority requirements, there also needs to be an output configuration file of “.cfg” format that needs to be created. This output file includes the updated BPLUT, which holds the updated table values, and the spin-ups from 2000-2019. The optimization of GPP and RECO are coded, but the numbers/calculations seem to be thrown off by NaN values in the input data or other calculations. The base code is there and well commented but the logic may need to be edited to get better results.

* 1. Contact info
* Arthur Endsley (Client/Primary Stakeholder): [endsley@umich.edu](mailto:endsley@umich.edu)
* Mark Matas (Team Lead/Developer): [mark.m.matas@gmail.com](mailto:mark.m.matas@gmail.com)
* Lucas Hamilton (Developer): [dukum2014@gmail.com](mailto:dukum2014@gmail.com)
* Max Thibeau (Developer): [thibeaumax@gmail.com](mailto:thibeaumax@gmail.com)
* Jake Pennington (Developer): [jake.pennington@umontana.edu](mailto:jake.pennington@umontana.edu)

1. Technical information
   1. Source Code Repository Information

The Source Code is owned by the client (Arthur Endsley) in a private repository on GitHub. The link is below but need permission from Arthur to access. The python3 code itself is within the “calibration” folder, data files that are grabbed by get\_data\_files.sh are in the “DataFiles” folder, and all the documentation that was required during this course resides in the “ProjectDocs” folder.

Link: <https://github.com/arthur-e/esm-calibration>

* 1. Assets that may not be in the repositories
     1. Digital assets

The input data files can be found at the link below, but permission may be needed to access these files. Pictures for the GUI are located in “calibration/GUI/images” folder. There are no other digital assets. Link: <http://files.ntsg.umt.edu/data/SKYentists/L4C_calibration_inputs/>

* + 1. Design documents, mockups, etc.

In “ProjectDocs” folder. NTSG and/or the client may have additional information necessary if needed.

* 1. Dev-ops
     1. Backups

The project only exists on the developers’ local computers and in the GitHub repository.

* + 1. Hosting information

None

* + 1. Deploying information

Version Control on GitHub.

* + 1. Test suites

None

* 1. Third party integrations
     1. Services (e.g. Stripe, AWS)

The project does not use any third-party services, unless NASA is counted. NASA funds and shares data with NTSG.

* + 1. Credentials for said services

None

* 1. Technology stack and hardware requirements
     1. Analysis of dependencies on third party libraries (e.g. what would break if you had to stop using this dependency, say, due to a security flaw)
* python-pyqt5: this package is used for the display of the pages within the GUI. If broken, then the GUI would be unusable.
* Matplotlib: this package is used to help create the plots for the GUI and CLI. If broken, the project would not be able to display the graphs (ramp functions, SOC estimation, etc).
* h5py: this package is used for the reading the input files of .h5 format. If broken, then the data needed from these files would not be present and the program may not run.
* Affine: this package is used for calculating tower weights correctly. If broken, the calculation of the dominant PFT may be thrown off.
* python3-gdal: same as affine package
* Scipy: this package is used within calculations for SOC and Flux Tower data, as well as outlier removal. If broken, these calculations would not be able to continue
  + 1. Detailed instructions to install all components necessary to develop and operate (with version numbers used)

Running the setup.sh file will automatically install all dependencies for the project.

* 1. Design Documents

Located in the “ProjectDocs” folder of the GitHub repository.

* 1. Analysis of technical debt

Technical debt exists in this project due to the time pressures with the CS Capstone class concluding on April 24, 2020 and the isolation of team members due to the COVID-19 virus situation. Some of the calculations are thrown off in the creation of ramp functions, optimized parameters, SOC estimation, etc. The logic may be wrong for handling the input data needed for these classes and may need to be reworked. The output file also needs to be developed due to the time constraints. The input data will be changed daily, as tower data updates every day. The isolation caused the team to not be able to meet in person or extensively test the software and interface design with many testers.s

* 1. Security considerations
     1. Privacy of data

Most data files are gathered from or created by NASA or NTSG repositories, which are mostly publicly available. Permission may be needed to access some of the NTSG data files.

* + 1. vulnerability to hacking

No vulnerabilities, as this project will be run locally with accessible flux tower and NASA data.

1. Considerations in a professional environment
   1. NDAs

None/Not needed as there is no private or sensitive information being used.

* 1. Ownership of data and accounts

Arthur Endsley owns the account that allows collaborators on the GitHub repository. NASA and NTSG own the data that is used for input in the program.

* 1. Transition timeline

There will be no “handing-off” of the code since the client owns the repository. The class, and the team member development, concluded on April 17, 2020.

* 1. Risk factor analysis

The NaN values in the input data may cause the data to be thrown off in the project, causing logical errors in the calculations. The output file also needs to correctly output file paths of the needed data.