Here all the descriptions for all our code files:

1: Initial Preprocessing: Each file executes the preprocessing done on each of the original IMF datasets, where we isolated variables of interest, renaming columns uniformly, and sometimes changed the layout of the database to be consistent. The IMF datasets were abbreviated with the following letters:

T: ‘Annual Surface Temperature Change’

C:‘World Monthly Atmospheric Carbon Dioxide Concentrations’

F: ‘Forest and Carbon’

L: ‘Land Cover and Land Cover Altering Indicator’

D: ‘Climate-related Disasters Frequency’

2: Creating Data for Training and Validation: These files are where we created our datasets to test the performance of each model on each feature set.

* “T\_no\_missing” cleans the preprocessed T dataset of missing values, making a baseline for the temperature feature set
* “T\_missing” creates a database equal to T\_no\_missing except with some values randomly removed; this is the dataset that will be given to the models to impute values into, with the results being compared to the actual data values held in T\_no\_missing
* “merged\_no\_missing” and “merged\_missing” function the same as above but for the extended feature set

3: Implementing Methodology: These files include all of our implementations of both ML-based and non-ML-based methods for imputing data into the datasets created above

* The Non-ML-Based Imputation Methods folder contains all our implementations of non-ML-based methods, including “normal” imputation with SimpleImputer (done row-wise and column-wise, on both feature sets) and completely random imputation
* ML\_Based\_Imputations\_from\_Packages contains the implementations of all ML-based methods we used from existing packages on both feature sets
* Our\_Random\_Forest\_Regressor contains the code where we implemented a random-forest-based imputation method from scratch

4: Analysis and Visualization of Results

* Result Visualizations, where we created all our figures to visualize our results. RMSE and MAE from each model were typed in manually, as they came from many different imputation methods in several different code files. All graphs created here are discussed in the Results and Evaluation section of the paper.
* Imputation into IMF Dataset with Best Model: includes our implementation of the MissForest model to impute missing values into the IMF temperature dataset