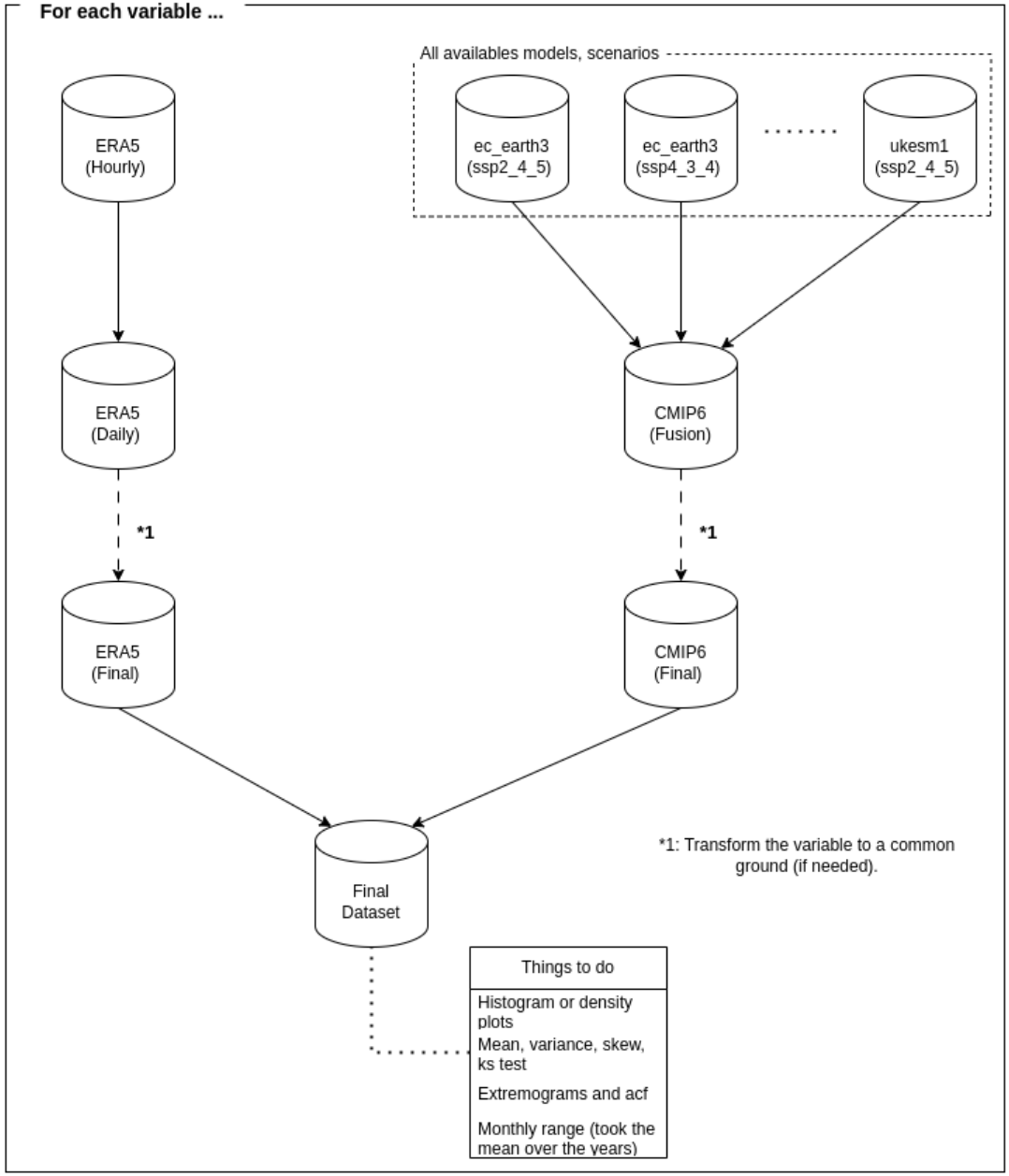
Report 03/05

May 18, 2024

Introduction

This presentation shows a summary of the work done in the past few weeks. It includes the models used and the results obtained.
This document is for internal use, so it may contain some errors.

Validate the models on the period (2020-2023)



- If we choose to use all the variables as a predictor, we need to look for the variables that are available in both models. Also we need to check if are measured in the same way or a transformation is needed. Also we need to define the period, right now we are only working with the data between 2000-2014, but we can expand this period.
- We train the models with the era5 data only with the train/validation/test split. (2000-2013) train, 2014 test.
- With the predicted result for the test, we can measure the errors of the model using our metrics. (We need to define if here we use paired or non paired)
- Right now the models are naive, xgboost and lstm (knn discarded (?)). The lstm is the one that is giving the best results, so we are going to focus on NN models.

- We need to define if we are gonna preserve the cmip values or we are gonna use the predicted values without considering the conservation of the aggregated values.
- With the downscaled data we can use our metrics (non paired) to evaluate the performance (taking into account the work done at the validation part).

- Validation
- Look for the predictors that are available in both models.
- Look for new NN models (Convolutional NN, ...).
- Preservation vs Non preservation