

Notes 2025-04-25

SMAP-HB / WRF-Hydro Project

Table of contents

This week	1
Fix data leakage	1
Updated current unet	1
PCA	2
Notes	4

This week

- ☒ Check distribution of target data before interpolation to look for weirdness
- ☒ Fix the normalized tiles
- ☒ Do PCA
- ☒ Make a unet that adds the dynamic data one layer before the last down block

Fix data leakage

- Fix data leakage from standardizing / normalizing data with respect to whole dataset
- Compute statistics beforehand and then edit data loader so it takes in original data and normalizes them when I run the model

Updated current unet

- Standardized data using statistics calculated from training data (2016-2017)
- Instead of creating a new set of netcdf tiles, just standardize the existing tiles as part of the model script
- Change placement of data fusion to occur before the last down block

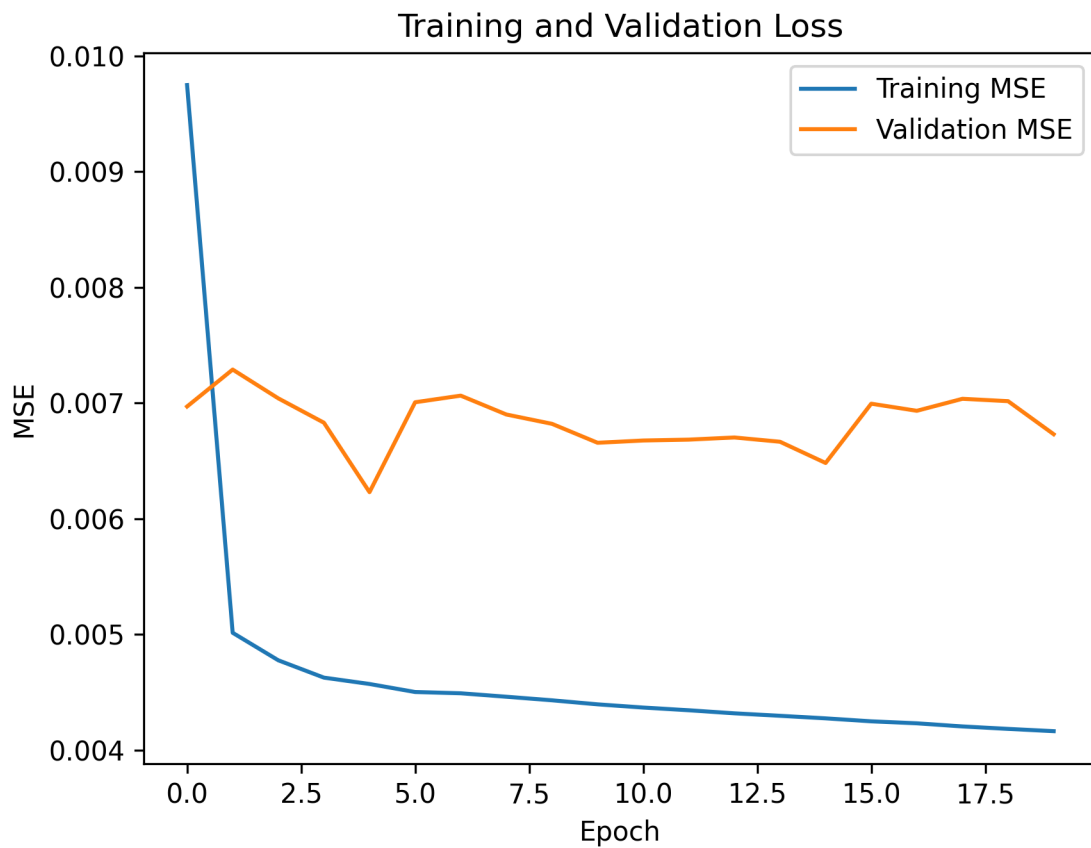


Figure 1: Current error curve

PCA

- Standardize instead of unit scaling
- In order to avoid dividing by zero, remove variables with zero variance beforehand
- Choose first 10 components
- First component captures 58% of variation

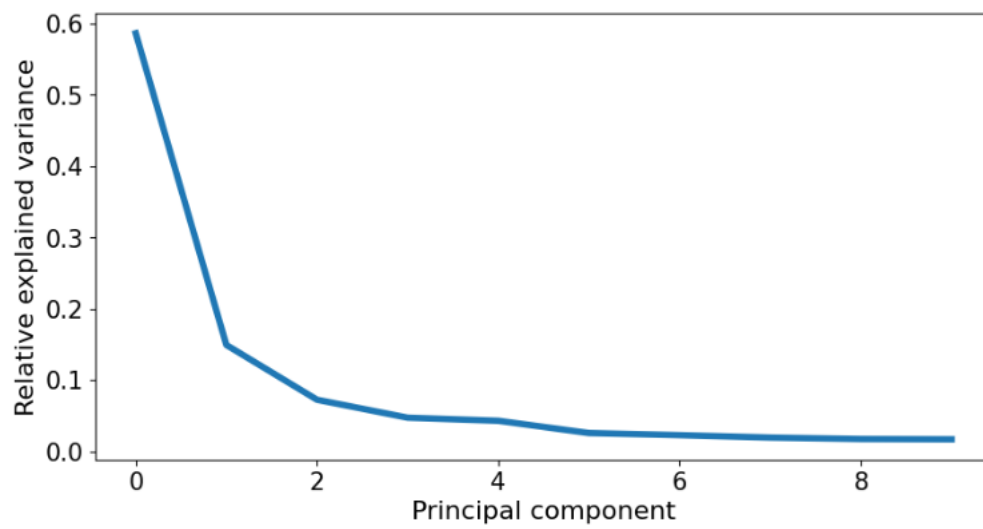


Figure 2: PCA variance explained

Notes

- Can always add more components, but first 5 is enough probably
- Remap static predictors from PCA to see what we're losing
- Save PCA data and use them for unet inputs
- Try standardizing the outputs
- Use 100 epochs with PCA
- PCA helps reduce noise, so maybe will increase accuracy
- Start with 5 and 10 and compare
 - How long it took and loss curve
 - Do it for 100 epochs