pipeline-gridded

The scripts for performing Argo profile gridding, mean field estimation, obtaining residuals, and learning the Gaussian process model are primarily in MATLAB. The MATLAB scripts are primarily derived from @mkuusela's scripts in this repository.

The scripts for performing the Argo profile pairing, projection, and downstream analysis are written in Python.

The scripts in `../pipeline-integrated` are run in exactly the same way; use this README as documentation.

Note that in most of the MATLAB files, I have placed macros in the format of `<PY:MACRO>` that are string-replaced by the code in `pipeline.py` prior to running. This makes it easy to change global parameters across scripts without editing individual scripts. However, this may make it difficult to run individuals scripts directly in MATLAB. Thus, it may be useful to clone the repository and replace these macros with literal variables if necessary.

MATLAB (oceanographic analysis)

1. `A00_processDACdata.m`: Read in raw DAC dataset and create Matfiles for later use.
2. `A01_pchipGridding.m`: Create p-chip interpolant for each profile, and read off estimates at a grid of depths.
3. `A02_concatenateArrays.m`: Glue together arrays that were created separately due to memory constraints.
4. `A03_createDataMask.m`: Mask based on amount of data available
5. `A04_filterUsingMasks.m`: Filter data based on aforementioned mask
6. `A05_splitHurricaneProfiles.m`: Split profiles into near-hurricane and non-hurricane profiles, based on the output of `B01` and `B02`.
7. `A06_estimateMeanField.m`: Using non-hurricane profiles, learn a global seasonal mean field.
8. `A07_subtractMean.m`: Remove the value of the mean field from the profile values, obtain residuals.
9. `A08_divideDataToMonths.m`, `A09_extendedData.m`: Prepare data for the Gaussian Process model.
10. `A11_localMLESpaceTime.m`: Fit Gaussian Process model.
11. `A12_fitLocalMLESpaceTime.m`: Evaluate GP model on data.

Python (tropical cycloone analysis)

1. `B00_SlimHurricaneDatabase.py`: Create lightweight database of all Argo profiles.
2. `B01_MarkHurricaneProfiles.py`: Find profiles near hurricane to create the hurricane subset and non-hurricane subset.
3. `B02_CreateHurricaneMask.py`: Takes the marked profiles to create a mask over profiles based on whether or not a profile is near a hurricane.
4. `B03_HurricanePairs.py`: Performs the pairing and projection steps of the paper. Calls a routine defined in `processing.py`.
5. `B04_ProfileDict.py`: Associate Profile IDs with the profile values outputted by the MATLAB scripts
6. `B05_AttachTemps.py`: Put data into Pandas Dataframes
7. `B06` uses kernel regression to lightly smooth the mean field data, to improve their viewability.
8. `B08`, `B09`, `B10` prepare the data structures for the spline estimates.
9. `B32`, `B35`, `B34` perform LOOCV, and then `B36` produces the spline fits at the chosen λ
10. `B21`, `B22`, `B23`, `B24`, `B25`, `B27` produce plots of the final estimates.

Running order

1. `B00`
2. `B01`
3. `A00`

4. A01
5. A02
6. A03
7. A04
8. B02
9. B03
10. A05
11. A06
12. A07
13. A08
14. A09
15. A11
16. A12
17. B04
18. B05
19. B06
20. B08
21. B09
22. B10
23. B32
24. B35
25. B34
26. B36
27. B21
28. B22
29. B23
30. B24
31. B25
32. B27