Gmail

Elizabeth Barnes <eabarnes.earth@gmail.com>

uncertainty for neural network regression tasks

Mark DeMaria - NOAA Affiliate <mark.demaria@noaa.gov> Wed, Oct 20, 2021 at 8:55 PM
To: eabarnes@colostate.edu

Libby:

Attached is a new data file similar to the previous one, but contains the track and intensity information along the lines of what we discussed in your office last week. The first 16 columns are the same as in the file I sent before, but with the headings changed a little so they are more consistent with the additional track columns. Below are details on the file and some experiments it would be nice to run for the proposal. Let me know if you have questions. Sorry this took so long. Matching up the track and intensity cases and doing all of the displacement calculations was trickier than I thought.

Thanks,

Mark

Intensity section: First 16 columns, through the DTL column.

ATCF - Storm ID, where first 2 letters are the basin ID (Atlantic, east Pacific, central Pacific)
Name - storm name
Date/Time - Starting date/time of the forecast
ftime - Forecast hour. The data are included every 12 hr through 168 hr.
NCI - Number of intensity models included in the consensus (max of 4)
OBDV - The difference in the observed max wind from the consensus max wind in kt
This is what we are trying to predict for intensity.

DSDV, LGDV, HWDV, AVDV - Difference between the model predicted intensity and the model consensus intensity. Models include DSHIPS, LGEM, GFS and HWRF (DS, LG, AV, HW). These are inputs to the prediction of OBDV.

VMXC - The max wind of the consensus forecast (kt). This can also be used as input to the OBDV forecast.

DV12, SLAT, SSTN, SHDC, DTL - Storm environmental variables (previous 12 hr intensity change, storm latitude from the NHC official forecast, SST, vertical shear and distance to land (km). These can also be used as input to the intensity forecast.

Track section

NCT - Number of track models in the consensus
OBDX - east-west difference (km) of the observed storm position minus that of the track model consensus. This is what we are trying to predict for the east-west displacement.
AVDX, EMDX, EGDX, HWDX - east-west difference each of the four track models from that of the track model consensus. These can be used to predict OBDX. The 4 track models are the GFS (AV), ECMWF (EM), UKMet global model (EG) and HWRF (HW).

LONC - Longitude of the consensus forecast.

OBDY, AVDY, EMDY, EGDY, HWDY and LATC -are analogous to the previous 6 columns for the north-south displacement.

Note that not every intensity model forecast had a track forecast. Missing track values are indicated by -9999. and need to be removed from the training set.

The simplest version of the track models would be to predict OBDX from AVDX, EMDX, EGDX, HWDX, LONC, and similarly for prediction OBDY from AVDX, EMDX, EGDX ad LATC. However, LATC and LONC could be used in both the OBDX and OBDY predictions and NCT would probably help with the uncertainty estimation (cases with fewer models in the consensus probably have larger track errors). The storm environmental variables VMXC, DV12, SHDC and DTL could

also be included as input to the OBDX and OBDY predictions. We should not use SLAT to predict either of the track variables though, since that is the latitude from the NHC forecast and would be confusing to interpret.

Once you get the prediction models for OBDV, OBDX and OBDY working with the new file we can decide what plots we want to use for the proposal, and how we want to do the partition between training and testing cases. Also, in the LOI we mentioned that we would try this on 2021 cases. I can put together an input file with those in exactly the same format, which I think would make that pretty easy to do.

[Quoted text hidden]

nnfit_vlist_20-Oct-2021.dat
7837K