Extending the Tropical Cyclone Genesis Index to Global Ensemble Forecasts

Alan Brammer (CIRA/CSU)
Dustin Grogan (U. Albany, SUNY), Andrea Schumacher (CIRA/CSU)
Jason Dunion (U Miami/RSMAS/CIMAS)
Project Background

• Tropical Cyclogenesis Index (TCGI) is a statistical model to provide guidance on 0-48, 0-120 hour genesis probabilities. Developed under JHT funding, and transitioned to NHC operations.

• The model utilizes deterministic GFS track and environmental conditions for Atlantic and Northeast Pacific designated invests.

• Deterministic model is operational and maintained by TSB (NHC) since 2017.
**Project Background**

- **Update & Retrain Existing Genesis Model**
  - This project has updated the existing TCGI model (last updated 2015), adding 6 years of training data to existing basins (AL, EP).

- **Provide Global Probabilistic Genesis Forecasts**
  - Expand the model to cover global systems adding statistical genesis guidance to WP, IO and SH.

- **Develop hybrid Ensemble + Statistical Genesis Guidance**
  - The model will be developed to run across ensemble inputs and provide forecasters guidance on uncertainty and improved guidance on genesis probability.

- **Develop, Evaluate & Share Ensemble Genesis Tools.**
Project Background

• Expand the existing deterministic Tropical Cyclogenesis Index (TCGI), a TC genesis statistical model developed under JHT funding, to utilize ensemble inputs and run for all basins.

• Deterministic model is operational and maintained by TSB (NHC) since 2017.

• Ensemble based version aims to improve uncertainty information available to forecasters. Improve reliability of probabilities through ensemble forecasts.
- Improved a prototype version of the ensemble TCGI model and transitioned into a Docker container (representative of WCOSS / NHC servers; RL4 -> RL7).
  - This allows us to run the full ensemble suite on any docker capable machine or on AWS to leverage the NOAA Big Data Project for easy to access archived GEFS. Also effectively reduces a lot of transition cost, as NHC could run the container with no extra installation required.

- Obtained global invest data from JTWC and expanded the pre-genesis dataset to cover 2001-2009 for AL, EP and 2009-2019 globally.

- Setup website and pipeline to run the "old" model in ensemble mode in real time for 2021.
  
  [https://rammb2.cira.colostate.edu/tc_research/ens-tcgi/](https://rammb2.cira.colostate.edu/tc_research/ens-tcgi/)
Work Performed

Update and expand pre-genesis training database.

Includes pregenesis disturbances for all basins from 2009 onwards.

Increases sample size by 4x from ~7k time periods, to ~27.5k time periods.

Novel dataset for project but also pre-genesis research in general. Developing automated invest criteria based on subjective invest characteristics.
**Updates to training algorithm.**

- Previous training code from JHT project was shared with us.
- We replicated the original linear discriminant analysis algorithm in Python (w/ pandas and scikit-learn).
- Experimented with removing dvorak dependency and alternate models.
- Initial results show that T-number (subjective dvorak classification) can be removed without degrading the model skill.
- Extra skill can be gained through alternative models.

Thanks to Jason Dunion and John Kaplan for the original Fortran LDA code.
Work Performed

Real-time displays

Generate areas of potential genesis from the ensemble spread. Areas include 1 standard deviation of track spread.

Working on verification tools for area of genesis as well as probability of genesis.

Comparing verification against NHC TWO genesis areas as well.
Work Performed

Real-time run Reliability
Cleaned up previous prototype code to utilize 30 GEFS members.

Pulls data from AWS S3 due to nomads rate limiting preventing ensemble file downloads in useable timeframe.
Improve code to adjust for mix of file path convention in previous years, expanded to run on any initialization from 2017-01-01 using S3 storage.
Running at Albany and CIRA to test stability and version updates.

Both 2020 and 2021 show reasonable reliability. Some noise due to low sample sizes in verification. Once Dvorak is removed sample size will increase substantially. Improving various metrics as well as utility for forecasters.
Transition Activities

- Updated backend training code (shared with operational model) to run on commonly used python libraries instead of licenced Fortran libraries.

- Updates to existing Fortran code base.

- Shared progress with JTWC and other parties at the 2021 USINDOPACOM Tropical Cyclone Conference.

- 2 abstracts submitting for AMS Annual 2022:
  - “Tropical Cyclone Genesis Index and Global Ensemble Forecasts: Prediction for the Atlantic and East Pacific Basins from 2018-2020”
  - “Tropical Cyclone Genesis Index and Global Ensemble Forecasts: Expanding to Global Basins and Evaluation of Real-Time 2021 forecasts.”
Transition Activities

Realtime output will be made publically available through a project website at CIRA:

rammb2.cira.colostate.edu/tc_research/ens-tcgi/

The webpage and output will be shared with NHC through the JHT based real-time demonstration portal at NHC.
Current and Future Plans

• Retraining statistical model on expanded training dataset.

• Present results thus far at AMS annual and tropical conferences in 2022.

• Finalize ML/AI/ stats method evaluation and determine optimum model for future use.

• Rerun forecast over recent years to evaluate stability and skill.
Project Information and Highlights

Deliverables

Leads:
Alan Brammer (CIRA/CSU), Dustin Grogan (U Albany)

Scope:
Expand existing Tropical Cyclone Genesis statistical model to run off GEFS members. Capturing uncertainty in the probability of genesis and pre-genesis track uncertainty and thus location of potential genesis. Also expand model to WP, IO, SH basins where there is no current TCGI guidance.

NOAA collaborator:
NHC

Expected Benefits:
- Additional guidance on genesis probabilities and uncertainty.
- New genesis guidance for WP, IO, SH basins.
- Improved pre-genesis TC guidance, and thus improved genesis forecasts, helping to achieve HFIP goal of improving pre-genesis forecasts by 20%.

Changes in the project:
Travel and Presentations delayed from year 1.

Challenges/Problems:
Rate limiting on nomads prevents GEFS files being obtained in a timely manner. AWS is being used instead.

Milestones

<table>
<thead>
<tr>
<th>Milestones</th>
<th>Completion Quarter</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Achieve RL7 for a real-time deterministic and ensemble TCGI forecasts for eastern North Pacific / North Atlantic utilizing the current version of TCGI</td>
<td>'21 Q2</td>
<td>Completed</td>
</tr>
<tr>
<td>2. Finished development of a pre-genesis training dataset.</td>
<td>'21 Q2</td>
<td>Completed</td>
</tr>
<tr>
<td>3. Manuscript in preparation on the updates to the training database and algorithm.</td>
<td>'21 Q4</td>
<td>Ongoing</td>
</tr>
<tr>
<td>Year 2:</td>
<td></td>
<td>----------</td>
</tr>
<tr>
<td>4. Achieve RL 7 for a real-time automated global ensemble-based TCGI probability forecasts for days 2-5.</td>
<td>'22 Q1</td>
<td>On track</td>
</tr>
<tr>
<td>5. Evaluate the use of AI techniques for genesis probability forecasts.</td>
<td>'22 Q2</td>
<td>On track</td>
</tr>
<tr>
<td>6. Collaborate with NHC and JTWC forecasters on the utility of the real-time forecasts.</td>
<td>'22 Q2</td>
<td>On track</td>
</tr>
<tr>
<td>7. Manuscript in preparation with respect to statistical/dynamical hybrid genesis skill in the global ensembles.</td>
<td>'22 Q4</td>
<td>On track</td>
</tr>
</tbody>
</table>