National Oceanic and Atmospheric Administration Joint Hurricane Testbed Program

NOAA Award No. NA17OAR4590140

Title – Transition of Machine Learning Based Rapid Intensification Forecasts to Operations

Andrew Mercer, Principal Investigator
Associate Professor, Department of Geosciences, Mississippi State University
108 Hilbun Hall, PO Box 5448
Mississippi State, MS 39762-5448
mercer@hpc.msstate.edu
662-268-1032 x231

Brandy Akers, Organizational Representative
HPC^2 Director of Operations, Mississippi State University
Mailstop 9627
2 Research Blvd.
Starkville, MS 39759
bakers@hpc.msstate.edu
662-325-8814

Submission Date – January 30, 2018

Recipient Organization (Name and Address)

Project/Grant Period (July 1, 2017 – June 30, 2019)

Valid for Period ending January 30, 2018

Report Term or Frequency (semi-annual)

Final Annual Report? No

1. ACCOMPLISHMENTS

*What were the major proposed **goals, objectives, and tasks** of this project, and what was accomplished this period under each task? (a table of planned vs. actuals is recommended as a function of each task identified in the funded proposal)*

Table 1 below outlines the planned objectives for the first 6 months of the project, as well as the outcomes.

Proposed Outcome	Anticipated Completion Date	Actual Outcome
Port R-based AI codes to Python scripts	28-Feb-18	Codes are still being ported, but much of the porting work is complete
Conduct independent verification of AI-based model on 2017 Hurricane Season	31-Jan-18	Verification results are complete (see Fig. 1 below)
Develop Training Codes to allow AIs to be updated prior to each Hurricane Season	31-Jan-18	Training codes are developed in R and being ported to Python presently

*Are the proposed project tasks **on schedule**? What is the cumulative percent toward completion of each task and the due dates? (table recommended)*

As seen in the results presented in Table 1 above, all tasks anticipated to be completed by 31 January 2018 have already been completed, and work that is expected to be completed by 28 February 2018 has largely been finalized.

*What were the major completed **milestones** this period, and how do they compare to your proposed milestones? (planned vs. actuals table recommended)*

The primary milestones completed during this phase included a porting of a majority of the artificial intelligence (AI) ensemble code to Python language for easier transition to operations, a completion of a full verification of the 2017 Hurricane Season using the AI ensemble (which was done in real-time as forecast observations arrived), and the development of code required to train the AI ensemble. At this stage, the AI ensemble training code is provided in R, but will be ported to Python by the end of February to ensure all codes are ready for testbed experiments in June.

What opportunities for training and professional development has the project provided?

At this stage, any professional development has come about through interactions with NHC personnel, particularly our primary NOAA POC Chris Landsea who provided important feedback to our product and our Testbed and Research to Operations plan. Additional training opportunities will be available once we attend the Interdepartmental Hurricane Forum in March.

How were the results disseminated to communities of interest?

At this point in the project, results have been disseminated only to the investigators and the NOAA points of contact through email. Going forward, a website devoted to showing the resulting AI ensemble output will be available to forecasters and the general public.

What do you plan to do during the next reporting period to accomplish the goals and objectives?

During this next reporting period, we will finalize all porting of R codes to Python, develop training materials for Hurricane Maria to help forecasters understand how to use and interpret the AI ensemble, and begin participating in the Hurricane Testbed experiments in 2018. This participation will include development of a prototype webpage to support the project. We will also attend the Interdepartmental Hurricane Forum to gain in-person experience working with National Hurricane Center forecasters, which will inform our product development as the project progresses.

2. PRODUCTS

*What were the major completed **products or deliverables** this period, and how do they compare to your proposed deliverables? (planned vs. actuals table recommended)*

At this stage, the only proposed deliverables were the formulation of the 2017 Atlantic Hurricane Season prediction results using the AI ensemble, which were completed. Verification was completed in a variety of ways but can be summarized by individual 2017 tropical cyclone event (Fig. 1).

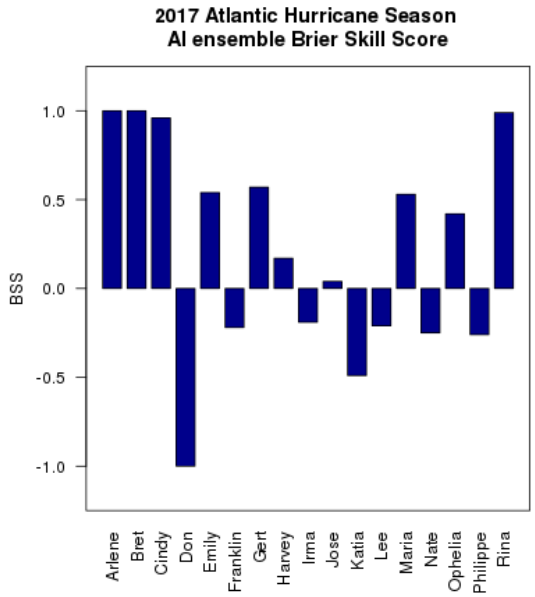

Fig. 1. Brier skill score performance for AI-ensemble probabilistic forecasts for 2017 all Atlantic tropical cyclone events.

In general, it was clear that the performance for most events was strong, with high positive skill particularly for cases which never achieved hurricane intensity (though some hurricanes were forecast exceptionally well). Interestingly, the AI ensemble performance seemed to degrade with events that existed for long periods of time over the central and eastern Atlantic (e.g., Hurricane Irma) or storms originating in the southern Caribbean near Central America, while the method performed better for events originating in the western Atlantic (e.g., Maria) and events which never achieved hurricane status. In general, the AI ensemble had an underprediction bias for rapid intensification (RI), favoring non-RI scenarios, suggesting another feature selection procedure may be needed.

These results demonstrate the viability of the model as a forecast tool and its readiness for entrance into the Hurricane Testbed experiments for 2018, meaning we have exceeded our time expectations for preparation. It is important to note that the NOAA POC stated the importance of the R to Python translation is not critical prior to the testbed as the product will be run at Mississippi State during that time. Regardless, progress towards the AI ensemble code porting to Python was substantial and should meet the February 28 deadline as stated in the original JHT proposal.

What has the project produced?

We have presented the work at one conference at this point in the project:

Grimes, A., and A. Mercer, 2017: Diagnosing Significant Magnitude and Spatial Field Differences Governing Rapid Intensification of Tropical Cyclones. *Southeast Division of the Amer. Assoc. of Geographers*, Starkville, MS.

We also have an abstract accepted to the 33rd AMS Conference on Hurricanes and Tropical Meteorology:

Grimes, A., A. Mercer, and K. Wood, 2018: Evaluation of Machine-Learning Based Rapid Intensification Forecast Performance During the 2017 Atlantic Hurricane Season. *AMS 33rd Conference on Hurricanes and Tropical Meteorology*, Ponta Vedra, FL.

3. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on this project?
Andrew Mercer – Principal Investigator
Kimberly Wood – co-Principal Investigator
Alexandria Grimes – Ph.D. candidate

We have also had NOAA points of contact that we have interacted with:
Chris Landsea – lead point of contact, research to operations point of contact
Erik Blake – testbed point of contact
Stacy Stewart – testbed point of contact

Other than learning the official NOAA points of contact, during this reporting period there have been no changes to personnel.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

At this stage, the most important impact is the demonstration of machine learning methods as viable alternatives to currently implemented statistical methods across the field of meteorology. Additionally, some novel feature selection methods have been introduced in the development of the AI ensemble, which should benefit the discipline as a whole.

What was the impact on other disciplines?

There has been no direct impact on other disciplines, though results presented at the Complex Adaptive Systems Conference in November 2017 have offered some novel methodological changes to a variety of scientific fields that utilize machine learning.

What was the impact on the development of human resources?

There has been no impact on the development of human resources.

What was the impact on teaching and educational experiences?

This project has been funding and training one Ph.D. candidate. Additionally, the project has garnered interest from outside faculty at Mississippi State who have used its content to help increase their knowledge of the machine learning field.

What was the impact on physical, institutional, and information resources that form infrastructure?

There has been no impact on physical, institutional, or information resources that form infrastructure at this point. Once the AI ensemble is ready to enter the testbed experiments, it will provide new informational resources to forecasters issuing intensification forecasts for tropical cyclones.

What was the impact on technology transfer?

There was no impact on technology transfer at this stage in the project.

What was the impact on society beyond science and technology?

Once finished, the potential improved predictability of RI within tropical cyclones will assist infrastructure preparations prior to hurricane landfall and provide forecasters much needed intensification guidance for tropical cyclones. The results could potentially save lives by

facilitating earlier evacuations (as needed) and enabling earlier preparations for landfalling major hurricanes.

What percentage of the award's budget was spent in a foreign country(ies)?

No funds were spent in foreign countries.

5. CHANGES/PROBLEMS

At this stage, other than the realization of some biases in the AI ensemble as a part of the 2017 Hurricane season performance results (e.g., Fig. 1), there are no notable issues with the project. The biases have led the PI to investigate some other feature selection approaches, though if those feature selection methods are not completed prior to the 2018 testbed the product will still be fully ready for operational testing. No changes in site location or costs have been incurred at this stage in the project.

6. SPECIAL REPORTING REQUIREMENTS

As this is a Joint Hurricane Testbed project, there are a few special reporting requirements. Outcomes for those requirements are provided after their questions below.

What is your assessment of the project's Readiness Level?

After completing real-time forecasts for the 2017 Atlantic Hurricane Season, the proposed AI ensemble officially reached readiness level of 5, meaning it is ready to enter the testbed experiments in June 2018.

Project Test Plan

The project test plan is provided as a separate document to this progress report.

Transition to operations activities in the last 6 months

There have been no official transition to operations activities in the last 6 months. However, the Program Manager requested a Research to Operations transition plan, which has been provided as a separate document to this progress report.

Summary of testbed-related collaborations, activities, and outcomes

At this point, we have been introduced to our NOAA points of contact (outlined in section 3) and have provided them with example AI ensemble output, a draft of our testbed plan, and a draft of our research to operations plan. Their feedback has helped us finalize those documents to ready the project for the testbed. Further NOAA POC engagement is expected with the upcoming testbed experiments.

Has the project been approved for testbed testing yet?

As of this progress report, the testbed plan has just been submitted, so official approval to participate in the testbed experiment has not yet been attained.

What was transitioned to NOAA?

Nothing has been transitioned to NOAA as of this progress report.

7. BUDGETARY INFORMATION

No major budgetary issues have arisen at this stage. The only budgeted funds to be spent by the end of the first quarter of the project were salary and benefits to the Ph.D. student, as well as her graduate tuition waiver. Additional funds will be encumbered with upcoming budgeted travel in spring 2018. The project is on-budget.

8. PROJECT OUTCOMES

What are the outcomes of the award?

No formal outcomes have been attained yet. Once the project finishes, the resulting AI ensemble should produce an outcome of improved RI forecasts, which should provide tremendous benefit to society in terms of early warning times for major hurricane landfalls and associated infrastructure preparations.

Are performance measures defined in the proposal being achieved and to what extent?

At this stage, performance metrics of the AI ensemble have not entered into the proposed work, but the objective of the AI ensemble outperforming the currently available RI products has been realized. According to the results of the 2017 hurricane season, the AI ensemble forecast skill for RI forecasts outperformed the SHIPS-RII model for 13 of the 17 named storms (76%), yielding much needed improvement in RI forecasts.

Test Plan for NOAA/OAR/OWAQ Testbed Projects

I. What major **concepts/techniques** will be tested? What is the scope of testing (what will be tested, what won't be tested)?

The primary outcome from the testbed will be an assessment of the AI-based ensemble of RI predictions in a true forecast mode. Each individual AI ensemble member will be run for each 6 hourly set of observations from the GFS and SHIPS data, and resulting performance of the model will be assessed by forecasters. Additionally, feedback on data presentation and format will be requested.

II. **How** will they be tested? What **tasks** (processes and procedures) and activities will be performed, what preparatory work has to happen to make it ready for NOAA testing, and what will occur during the experimental testing in the testbed?

The software will be tested by automating the running of a Python code that will do most of the legwork for the forecasters. The Python code will obtain the most recent GFS file, the most recent SHIPS file, and the most recent b-deck track information (mainly needing storm position and peak speed and current intensity). Once those files are obtained, the code will scale those observations to the previously developed training database and formulate predictions of RI (binary 1/0 predictions) for each of the AI ensemble members. Once completed, the code will compute a net probability of RI as a final forecast.

An ideal but not necessary goal is to have all implemented codes (currently provided in R) transitioned to Python for ease of transition into operations prior to the start of the testbed. If this is not completed, the testbed will still proceed using the R versions of the code which are already functional. After this aspect is finalized, the only step prior to the testbed that will be required is training the AI ensemble on the historical database (which we already have). Once the AI ensemble is trained, running it on new cases will simply be a matter of automatically running the code for a given new case. Forecaster overhead on the "how" side of things will be minimal.

III. **When** will it be tested in coordination with the NOAA testbed? What are **schedules and milestones** for all tasks described in section II that need to occur leading up to testing, during testing, and after testing?

I am a bit unsure about the project timing. I personally would like the timing to span the full 2018 Hurricane Season (1 June – 30 November) so that the ensemble is available to forecasters the entire time. The idea is that each new TC will have an associated RI forecast once at least two observation timesteps

are available for that event (it has existed for at least 6 hours). These two observation timesteps are needed for persistence and storm trajectory information. Once obtained, the ensemble will be run for each Atlantic Basin TC every 6 hours when new GFS output is available.

A brief timeline:

- **Convert currently implemented AI ensemble to Python code, including two primary modules**

    - **Code to train the AI ensemble, which is run once at the beginning of each Hurricane Season**
    - **Code to forecast RI for new timesteps in the current Hurricane Season**

  This conversion will likely take 6 months to complete. Once finished, two Python scripts, access to current SHIPS/NHC/GFS data, and the training database will be all that is required to run the ensemble. Note that if this is not able to be finished, the current R versions of the code are already fully functional and able to be used in the testbed, so completing the Python transition is an ideal outcome. *If possible, completed February 2018.*
- **Brief pre-testbed testing of the AI ensemble Python code using the 2017 Hurricane Season to assess AI ensemble performance in a "quasi-operational" forecast mode. Note these forecasts are being generated presently, but may change slightly using Python's AI libraries instead of R libraries. These results are already preliminary completed and ready to present at the Interdepartmental Hurricane Conference. *Presented March 2018***
- **Development and dissemination of a training/tutorial guide for the product for forecasters on two example cases (Hurricanes Irma and Harvey from 2017). Can be disseminated to forecasters in-person during the Interdepartmental Hurricane Conference. *Completed concurrently with the 2017 assessment by March 2018.***
- **Participation in the testbed by beginning to generate operational RI forecasts on all Atlantic Basin TCs in the 2018 Hurricane Season. Results posted every 6 hours for existing TCs to a local Mississippi State Website that forecasters will have full access to for forecasting purposes. *Will span the 2018 Hurricane Season***
- **Once the 2018 season ends, forecaster feedback on output formatting, output information, etc., will be obtained so that the ensemble may be further modified and improved for eventual transition into operations. *Finalization component ends June 2019.***

- **Complete final report and ideally begin the transition phase into operational forecasts.**

IV. **Where** will it be tested? Will it be done at the PI location or at a NOAA testbed location?

As of now we plan to conduct the simulations on a local server at Mississippi State and provide the updated output on a website. However, that is certainly flexible as the AI ensemble can be run on any modern computer with little to no trouble.

V. Who are the key **stakeholders** involved in testbed testing (PIs, testbed support staff, testbed manager, forecasters, etc.)? Briefly what are their **roles and responsibilities**?

The only major stakeholders in the testbed phase should be the operational forecasters using the product. The forecasters will have two roles. First, they will be responsible for working through and evaluating a tutorial product on how the AI ensemble works for two past cases, Hurricanes Harvey and Irma. This will allow them an opportunity to interact with the product prior to the 2018 Hurricane Season. Once the Hurricane Season begins, their role will be to complete their forecast as they usually would, but to consider the AI ensemble as one of the products they analyze prior to issuing intensification forecasts for any ongoing Atlantic Basin TCs. The idea is that the AI ensemble would complement the methods they are currently using. The only responsibilities of the forecaster would be to include the AI ensemble in their forecast reasoning and to make notes regarding areas where the ensemble may be improved, either from a formatting/aesthetic perspective or a technical/forecasting perspective.

VI. What **testing resources** will be needed from each of the above participants (hardware, software, data flow, internet connectivity, office space, video teleconferencing, etc.), and who will provide them?

During the testbed, the only resources a forecaster will need will be a computer with internet access and a web browser, as well as some mechanism by which the forecaster may take notes. The products will be disseminated to forecasters via a web site hosted at Mississippi State, where we will supervise the product's dissemination to minimize forecaster overhead. Occasional phone meetings with the NOAA POC may be needed to assess how well the product is working and other technical issues that may

arise. Those will not be a consistent requirement of the testbed though and will not be needed for forecasters.

VII. What are the **test goals, performance measures, and success criteria** that will need to be achieved at the end of testing to measure and demonstrate success to advance to higher Readiness Levels and to proceed to full transition to NOAA operations (Readiness Level 8)?

Success will be measured in two key areas:

- **Forecaster feedback: Forecaster feedback will be used to improve the usability of the product, including how the product is disseminated to the forecaster (and which information is not disseminated) and what information may be missing that the forecasters would want. The testbed is critical for this step as it ensures forecasters have input into the product they would eventually be working with and helps identify problem areas that were not previously known in the AI ensemble methodology. It will be determined "successful" if forecasters find the ensemble useful in preparing their work and the ensemble can be improved to fit in the framework of expectations for operational RI forecasts.**
- **Ensemble performance: The ensemble performance will be continuously measured throughout the 2018 Hurricane Season in two ways. First, the Heidke Skill Score for each individual ensemble member will be retained real-time such that the forecasters have AI ensemble performance information for the previous forecasts of the season. This will likely include both a full season HSS value and an HSS value for the current storm of interest. The second performance measure will be a Brier Skill score for the probabilistic output of the AI ensemble as a whole. All of these values will be provided to the forecasters as a component of the main output for each run of the AI ensemble, giving forecasters a real-time glimpse into the ensemble's performance for each event. Ideally, for the program to be successful, performance metrics will exceed those in the current SHIPS-RII model, such that value is added by including the AI ensemble system.**

Assuming the model is successful in both of these areas, we feel the final product will be at readiness level of 7 and will be ready for consideration to transition into an operational product.

VIII. How will testing **results** be documented? Describe what information will be included in the **test results final report**.

The final report will certainly include the performance metrics for the individual ensemble members and the product as a whole. These values will

be compared against the SHIPS-RII baseline performance for the 2018 hurricane season. The AI ensemble will be found "successful" if it outperforms the SHIPS-RII implementation currently used operationally, and if forecasters find its input valuable. A survey at the end of the project given to forecasters will give good qualitative information about the positives and negatives of the AI ensemble, all of which will be provided in the final report.