2nd Report of the Scientific Review Committee
for the Hurricane Forecast Improvement Project (HFIP)

Committee Members:
Robert E. Hart, Florida State University
David S. Nolan, University of Miami
James F. Price, Woods Hole Oceanographic Institution
Gary M. Barnes, University of Hawaii
James Hansen, Naval Research Laboratory
Wayne Schubert, Colorado State University

8 October 2014
I. Background

The Hurricane Forecast Improvement Project (HFIP) is a 10-year program supported by the National Oceanographic and Atmospheric Administration (NOAA) with the express purpose of improving operational forecasts of hurricane track, intensity, and related storm surge. Since 2008, HFIP has provided funds to NOAA laboratories, U.S. government laboratories outside of NOAA, NOAA cooperative institutes, and university investigators for purposes ranging from basic research on tropical cyclones (hurricanes), improving computer forecast models and the use of observations, and generally improving the regular forecasts and updates that are produced by the National Hurricane Center (NHC) during hurricane season. A more detailed discussion of the inception and ongoing activities of HFIP are available from the article by Gall et al. (2013) and in the HFIP annual reports.

In summer 2012, Drs. Robert Gall and Frank Marks recruited a number of scientists and university faculty to form a Scientific Review Committee (SRC). In early 2014 two SRC members were replaced at the end of their terms. The current members of the SRC and their affiliations are listed on the front page of this report. The purpose of the SRC is to provide independent evaluation and scientific direction in regards to the various projects that HFIP supports.

II. Latest SRC Meeting

A meeting of the SRC was arranged for July 29 and 30, 2014, at the Atlantic Oceanographic and Marine Laboratory (AOML) in Miami, Florida. Of the SRC members, Drs. Hart, Nolan, Barnes, Schubert, and Price were able to attend. Dr. Hansen participated in all discussions by conference call. Also in attendance were HFIP Development Manager Dr. Robert Gall, NOAA/HRD director and HFIP Research Lead Dr. Frank Marks, HFIP Model Strategy Team Lead Dr. Vijay Tallapragada, Dr. Ivanka Stajner of NOAA/NWS, and NHC Branch Chief of the Hurricane Specialists Unit, Mr. James Franklin, presenting in place of NHC Deputy Director Edward Rappaport.

The SRC members heard reports from each of Drs. Gall, Marks, Franklin, and Tallapragada. HFIP Program Manager Fred Toepfer also reported via teleconference regarding current and future funding for HFIP. The SRC members asked questions that led to extensive discussions throughout the course of the presentations. At the end of the first day, the SRC members met alone to discuss what they had heard and learned that day and to formulate their scientific assessment of the program. The following morning, the SRC presented its initial findings in the form of a two-page summary PowerPoint to Drs. Gall and Marks, which acted as the framework for this report.
The SRC accepted its charge of producing this report, describing its evaluation of the status and progress of HFIP, offering critique of current activities, making suggestions for which activities to increase or reduce in the future, and making recommendations for new activities.

III. SRC Evaluation of HFIP

Overall, HFIP has been a benefit to the hurricane research and forecasting community, at the very least by channeling additional funds and effort to various operational improvements and research projects. HFIP has also sponsored meetings and workshops, which we know from experience often lead to further advances and new collaborations.

In the following, we discuss specifically which aspects of HFIP have had positive impacts, and which aspects of HFIP concern the SRC.

**Positive changes since the prior report:**

1. **The community perception of HWRF and HFIP**
   Informal conversations some SRC members had following the first SRC report and during the 31st AMS Hurricane and Tropical Meteorology Conference in April 2014 suggest that the community perception of HWRF and HFIP has improved. The SRC suggests that this is a result of the combination of the prior release of the first SRC report (which highlighted advances) as well as the HFIP advances noted below, which were apparent to some in the community before the report.

2. **Potential HWRF forecast improvement**
   We are in general agreement that HWRF is performing better in the last two years than in the first three to four. Further, retroactive runs shown to the SRC using the 2014 HWRF configuration changes reveal potential for additional future improvement. However, we temper this optimism with a bit of caution, as we did in the prior SRC report:

   a. When Atlantic intensity forecast skill is normalized by the well-accepted benchmark measure of skill (SHIFOR), as was shown by James Franklin during presentations to the SRC, it is clear that there has been no statistically significant improvement in normalized intensity skill since the start of HFIP.

   b. It is also unclear how much of the track improvement shown during HFIP is due to simultaneous improvement in the GFS global model, that latter of which provides the environmental initial conditions and lateral boundary conditions for HWRF.

Nonetheless, the SRC looks forward with an open mind to the results of the 2014 Atlantic and Eastern Pacific seasons using the new HWRF configuration in an aggregate active
season such that a wide range of TC behaviors can be evaluated. The wide range of intensities of observed tropical cyclones across both basins at the time of writing of this report will hopefully have provided a robust test of the new model configuration.

3. **The focus of recent awarded grants**
   The recent (August 2014) awarded HFIP grants appear to address the data assimilation and other issues noted in the prior SRC report. The SRC was very pleased to see that the recommendations in the first report have been acknowledged through this new funding and very much look forward to the results from these new grants.

4. **The ongoing expansion of the JET computers**
   Two important contributions of HFIP to the TC research and forecasting community are the development and support of the “stream 1.5 models” and also the real-time experimental ensembles (the GFDL and HWRF ensembles). These could not have been possible without the creation and expansion of the so-called “Jet” computers (Njet, Tjet, Ujet, and Sjet) which now make over 22,000 computational cores available for both real-time forecasts, off-season retrospective forecasts, and experimental research.

5. **Wider Use of the HWRF model itself**
   The dissemination of the HWRF model itself has grown significantly and its wider use should lead to further model improvements through case studies, conference presentations, publications, and inevitable bug reports. The proliferation of the model and its improvement has been further facilitated by workshops sponsored by HFIP on the use of HWRF.

6. **NHC’s involvement in vetting models**
   It is a great benefit that NHC is involved in the annual vetting of the models and their determination for inclusion in consensus guidance. Such vetting has led to a shift and refinement in the selected models year to year. The process has identified those models that make the greatest contribution to the forecast process, eliminating those that are redundant guidance or provide no skill.

**HFIP Areas of Concern:**

1. **The further increasing unlikelihood of intensity skill goals being met**
   While there has been progress after 5 years of the HWRF raw TC intensity performance, realistic assessment of the progress of HFIP in its primary goal shows that the 5 year skill goal is not yet being met and the 10-year intensity skill goal is very unlikely to be met. Further and more revealingly, as shown by James Franklin during the SRC meeting,
when intensity forecast skill is normalized by benchmark measures of skill (SHIFOR), it is clear that there has been no statistically significant improvement in intensity skill since the start of HFIP. Thus, the Atlantic raw intensity error improvements attributed to HWRF may in fact be due more to interannual changes in forecast difficulty. It should be added that it is also unclear how much of the raw intensity performance change was due to coincident improvements in the GFS global model, which should be evaluated independently of changes in the HWRF model.

2. The evaluation barrier of metric choice
The choice of verification metrics drives the design and implementation of the entire forecast system. The SRC continues to encourage HFIP to develop and evaluate additional metrics of forecast performance (e.g., cumulative error distributions, storm size and structure metrics). While the SRC understands that there may be resistance to such additions/changes from the operational community in light of the observational infrastructure established toward existing metrics, the choice of the HFIP metric of intensity performance may be obscuring practical and scientifically important improvements within HWRF and HFIP related to structure.

3. The sudden and unexpected shift in HFIP focus in the strategic plan to global modeling
There was a lot of emphasis during presentations to the SRC on current basin-scale and future global model development associated with the strategic plan. The SRC believes that a focus on global model development distracts from the most pressing issues (e.g., data assimilation and an improved regional model leading to improved TC intensity forecasting). In short, the strategic plan that was provided to SRC is “too much EMC and not enough HFIP.” This is not a judgment about the value of regional modeling vs. global modeling – but given the prescribed scope of HFIP and its goals, such new goals would seemingly be outside the scope of HFIP. This redirected scope is more alarming in light of the problems with data assimilation – presumably more key to short-term intensity change – identified during the SRC meeting by those presenting and discussed next.

4. Verification of inner core structure and assimilation of observations therein.
It was shown clearly in the presentations to the SRC that assimilation of observations within the TC largely leads to a degradation of HWRF performance. The solution proposed (for 2014) to the SRC by Dr. Tallapragada was to exclude observations from assimilation within a specified radius of the center. However, this is a solution indicative of capitulation. Inner core data assimilation cannot succeed until simulated inner core structure gets closer to observations, including the correct slope of the eyewall. It is noteworthy that the lack of observational assimilation within the TC in the western
Pacific (where there are no reconnaissance flights) was argued to be one reason why retroactive runs of the 2014 model configuration showed the greatest improvement in that basin.

5. **Extending forecasts beyond 5 days**
   Five to seven day forecasts should be viewed as a low priority in the view of the SRC. A regional model with skillful short term intensity forecasts is much more valuable than a basin scale or global model with better 5-7 day track forecasts.

6. **Forecaster use of HFIP ensembles**
   Ensemble predictions can be used as model diagnostics for improved understanding, for input into data assimilation, and/or as forecast guidance. Regarding forecast guidance, it is not clear that the information being provided by the HFIP ensembles is being used by the NHC forecasters (beyond the ensemble mean). This could be because there is not an effective way to disseminate the ensemble information, the relevant information is not being disseminated, or because effective training is not in place. Does the current use of the ensemble information by NHC forecasters justify their production and large computing expense – especially in light of the pressing and CPU-demanding inner core assimilation (ensemble-independent) issues noted above? Or, perhaps additional emphasis should be placed on product development, dissemination, and training.

**IV. Recommendations of the SRC**

Based on our evaluation of HFIP, its goals, and its activities, this SRC makes the following specific recommendations:

1. **Developing and Embracing Alternative Measures of Success**
   *Note that this recommendation remains verbatim the same as in the first report.*
   The HFIP goals related to reduction of track forecast errors may indeed be achieved by the end of the program. While this is very positive, almost all of this improvement likely comes from ongoing steady improvements in global forecast models, and thus is generally outside the scope of HFIP activities. Indeed, it remains the case that the average track errors of the regional models such as HWRF, GFDL, and the Stream 1.5 models are greater than those of some global models.

As noted above, the intensity and rapid intensification (RI) goals are probably unrealistic and perhaps impossible. While the “intensity” of a hurricane is of course a very important metric, we have come to understand in recent years that it is not the only important metric. Hurricanes Katrina (2005), Ike (2008), and Sandy (2012) were examples of storms that made far more impact due to the sizes and shapes of their wind fields than their intensity alone would have
indicated. There appears to be room for substantial improvements in the capabilities of operational forecast models to predict the size and shape of the hurricane wind field. Along with further improvements in surge predictions, this can also lead to improvements in the sizes of warning and evacuation zones, and the arrival at the coast of tropical storm and hurricane force winds. A continuing commitment to acquiring and using observations of the horizontal extent of the wind and ocean wave fields will provide an opportunity to validate these aspects of the operational forecasts, will lead to improvements in model physics and initialization, and has the potential to lead to improvements in all aspects of the forecasts.

Other measures of model forecast success are worth considering. For example, a reduced frequency of the largest errors (or similarly, a reduction of the value of the 90th percentile of error) may have little impact on the mean but would add confidence to the forecasts by reducing the more “spectacular” failures. Increasing realism and accuracy in the structure of the precipitation field (or alternatively, simulated satellite observations) should also be interpreted as signs of improvement, as they will likely lead to further improvements in the size and shape of the outer wind field and will aid in real-time assessments of model accuracy.

As discussed in the SRC meeting, there may be some reluctance from the National Hurricane Center to support redirection of HFIP toward other metrics. But from the SRC perspective, it is pointless to chase unachievable intensity goals, when other metrics could lead to advances: if skillful, reliable forecasts of hurricane size, structure, etc., can be made available, we believe the emergency response community will make good use of it.

2. Consider normalizing HWRF performance relative to external and independent guidance
Normalization of intensity performance relative to SHIFOR illuminated that much of the implied intensity skill in the past five years may have been independent of HWRF model improvements, given the evolving interannual seasonal difficulty. It is recommended that such normalizations continue with respect to intensity, such that correct degree of attribution of HFIP-specific progress can be made.

Further, while there has been similar improvement in track performance, it is unclear how much of this improvement is due to improvements in the GFS model, which provides the initial conditions for the HWRF outside the TC and also provides the HWRF boundary conditions. The SRC recommends that track performance be normalized not only with respect to CLIPER (consistent with SHIFOR for intensity) but separately also with respect to GFS track forecast error. Such comparisons will help illuminate how much of the improvement in HWRF track skill is due to HWRF itself, and how much is due to changes outside HWRF.
3. Redirecting global modeling resources to data assimilation
The apparent refocusing of long-term HFIP resources toward global modeling is very concerning in light of the deficiencies noted above. It is the strong recommendation of the SRC that the emphasis of HFIP should be to strengthen efforts on data assimilation within the TC itself, and in particular the core.

4. Coupled model emphasis
It is uncontroversial that a coupled, air/sea forecast model should be the goal of HFIP. Our present understanding and assumption here is that SST is the ocean variable of greatest importance. Presently, the operational HWRF and GFDL models for the Atlantic are coupled to a version of the Princeton Ocean Model (POM) that is initialized with a combination of climatological SSTs and currents, the SST from Reynolds 7-day running analyses, and information from the positions and strengths of various features such as the loop current and the Gulf Stream. While this is undoubtedly a large improvement over past, simpler initializations (or no coupling at all), it seems logical that the ocean coupling and its initialization should evolve toward a more comprehensive, perpetually cycled data-assimilated system, such as used by HYCOM for ocean forecasts. The priority in ocean model coupling should be in advancing the ocean model and its initialization, rather than in introducing new physical processes, such as ocean wave modeling or the effects of sea spray in high winds. The SRC made essentially this recommendation in our first report, which was unfortunately misquoted in some later summaries. The SRC is restating it here to ensure clarity.

5. Polling the NHC forecasters on HFIP ensembles
The ensemble results presented to the SRC focused on the skill of forecast guidance products. It is not clear if the ensemble products are of value to the forecasters, or if they are being disseminated to the forecasters in a useful way. The SRC recommends that a cognitive task analysis of the hurricane forecast process be performed with an eye towards identifying the ways in which products based on probabilistic guidance can best be utilized (if at all). Such an analysis will help to elucidate the partitioning of the barriers to the use of ensemble guidance in the context of the existing operational workflow and infrastructure.

Summary Statement
The SRC surmises that concentration on the 12-48 h intensity forecast problem, with keen attention to the effective assimilation of inner core observations, is the best avenue to approach HFIP goals. The adoption of other TC forecast metrics beyond track and intensity endures as a way to more fully exploit the work conducted under the HFIP aegis.