Final Report

From the

National Unified Operational Prediction Capability (NUOPC)

Interim Committee

On

Technology Transition Processes (TTP)

1 October 2009

Prepared and Reviewed by:
Technology Transition Processes Committee
Maj J. Marc Hidalgo (Air Force Weather Agency, Chairman)
Mike Horner (Air Force Weather Agency)
Ken Smith (Air Force Weather Agency)
Steve Payne (Commander Naval Meteorology and Oceanography Command)
Mark Swenson (Fleet Numerical Meteorology and Oceanography Center)
Patricia Phoebus (Naval Research Laboratory)
CDR Dan Eleuterio (Office of Naval Research)
John Ward (National Oceanic and Atmospheric Administration)
Julian Wang (National Oceanic and Atmospheric Administration)
Bob Gall (National Oceanic and Atmospheric Administration)
Dave Knapp (Army Research Laboratory)

Submitted by: J. MARC HIDALGO, Major, USAF
2nd Weather Group/Technology Exploitation Operations
Air Force Weather Agency

Approved by: FREDERICK TOEPFER
NUOPC Project Manager
National Oceanic and Atmospheric Administration

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY. IV

2. INTRODUCTION. 1

2.1 BACKGROUND. 1
2.2 GOALS OF THE TTP COMMITTEE. 1
2.3 METHODOLOGY. 2
2.4 ORGANIZATION OF THIS DOCUMENT. 2

3. OUTREACH PROCESS. 2

3.1 PURPOSE. 2
3.2 RESEARCH 3
3.3 REACHING OUT TO THE RESEARCH COMMUNITY. 4
3.4 LONG TERM NATIONAL AGENDA AND SHORT TERM MANAGEMENT PLAN. 5
3.5 POTENTIAL RISKS AND ROADBLOCKS. 6
3.6 COSTS. 7
3.7 RECOMMENDATIONS. 7

4. RESEARCH TRANSITION FACILITY (RTF). 8

4.1 PURPOSE. 8
4.2 EXISTING TRI-AGENCY TECHNOLOGY TRANSFER PROCESSES. 8
4.2.1 NOAA. 8
4.2.2 U.S. Navy. 8
4.2.3 U.S. Air Force. 9
4.3 RTF FUNCTIONS. 9
4.3.1 Outward Facing Objectives. 9
4.3.2 Inward Facing Objectives. 11
4.4 DISTRIBUTED TESTBED CENTER (DTC). 14
4.5 POTENTIAL RISKS AND ROADBLOCKS. 19
4.6 COSTS. 19
4.7 RECOMMENDATIONS. 21

5. METRICS PROCESS 22

5.1 PURPOSE. 22
5.2 CURRENT STATUS. 22
5.3 METRICS 23
5.3.1 Performance Metrics. 23
5.3.1.1 Deterministic Forecast Metrics. 23
5.3.1.2 Ensemble Forecast Metrics. 24
5.3.1.3 Application of Metrics. 25
5.3.1.4 Selection and Maintenance of Metrics. 27
5.3.2 System Metrics 28
5.3.3 Strategic Metrics 29
5.3.3.1 Metrics to Determine Accelerated Technology Transfer. 30
5.3.3.2 Metrics to Demonstrate Reduced Duplication of Effort. 31
5.3.3.3 Metrics to Assess a Superior National GNWP Capability. 31
5.4 POTENTIAL RISKS AND ROADBLOCKS. 31
5.5 COSTS. 32

5.6 RECOMMENDATIONS. 33

6. SUMMARY 35

APPENDIX 1 – LIST OF ACRONYMS AND DEFINITIONS 38

APPENDIX 2 – GLOSSARY 40

APPENDIX 3 – DTC PROPOSAL 42

APPENDIX 4 – NOGAPS SCORECARD 52

APPENDIX 5 – JEFS DESCRIPTION OF ENSEMBLE VERIFICATION CHARTS. 54

List of Figures

Figure 1. RTF Model. 14

Figure 2. TTP Committee Schedule and Costs 37

List of Tables

Table 1. DTC Advantages and Disadvantages, 17

Table 2. RTF Advantages and Disadvantages, 18

1. EXECUTIVE SUMMARY.

The overarching goal of the interim Technology Transition Processes (TTP) Committee is to facilitate a common or collaborative technology transition process to accelerate the development and transition of new and improved capabilities into National Unified Operational Prediction Capability (NUOPC) operations. We propose several processes which will allow NUOPC to quickly transition research advances into operations.

First, the interim TTP Committee recommends the establishment of a standing TTP Committee. Much like the interim committee has done up to this point, a standing committee is needed to manage and coordinate NUOPC's technology transfer processes. Central coordination is required to prioritize Tri-Agency research needs, manage/coordinate on the technology transfer functions collocated with the operational centers, and develop and update (as required) a metrics program for measurement of technology transition processes. Additionally, the standing TTP Committee will encourage and stimulate research community involvement in NUOPC in order to facilitate the rapid transition of new improvements or technologies into operations. Outreach will necessitate providing stimuli in the form of visitor programs, funding, and access to computing resources, in order to promote interest and involvement from the research community.

Second, the interim TTP committee recommends establishing a Fellowship Program to elicit researcher assistance in tackling a prioritized Tri-Agency research requirements list. Ideally, each production center would host up to three fellows, and a Program Manager (PM) would be required to effectively manage these fellows, resulting in a total of 10 full time equivalent (FTE) positions. Calls for fellowship opportunities will be made via Broad Agency Announcements (BAA) or other programs, such as the National Research Council (NRC) Fellowship Program. Estimating $200K per year for one FTE, and assuming that each fellow would be funded to work full time on NUOPC projects, yields a total of $2M per year for 10 FTEs, beginning in 2013, when NUOPC's technology transfer functions stand up. This dollar figure is scalable based upon the amount of research the Tri-Agency wants performed, the number of fellowships desired per center, and the total award for each fellow, including travel costs. In addition to working on the coordinated research agenda, fellows would also be expected to visit the production centers for one to two months per year to interact and exchange ideas with operators and visiting scientists to enable effective collaboration, ultimately leading to improved operational capabilities.

Third, the interim TTP committee recommends a Research Transition Facility (RTF) be established at each Tri-Agency production center to address "outward" facing and "inward" facing technology transfer objectives. *Outward* facing objectives refer to improved collaboration and communication with the research community by the production centers. *Inward* facing objectives pertain to those actions taken to assist NUOPC member agencies in improving their operational capabilities. Ideally, each production center would host up to three visiting scientists

for RTF work on-site, beginning in 2013. The number of visiting scientists is scalable based on member agency desires and funding profiles. Calls for research visitors would be made via the BAA process or other programs, such as the University Corporation for Atmospheric Research (UCAR) Visiting Scientist Program (VSP). A PM would be needed to manage the visitors, resulting in a total of 10 FTE positions. (However, the PM for the Fellowship Program may be able to manage the visiting scientists as well, reducing our recommendation for two PM FTEs down to one.) Estimating $200K per year for one FTE yields a total of $2M per year for 10 FTEs. This dollar figure is scalable based upon the amount of operational improvements and Tri-Agency collaboration desired. This estimate doesn't include the cost of computers and people on-site to support the visiting scientists. In order to assist the visitors, some script writing would be required. A few programmers would be needed at the RTFs to set-up and write user-friendly graphical user interface (GUI) programs, much like the European Centre for Medium-range Weather Forecasts (ECMWF) has done. This work will need to be completed prior to the standup of the RTFs to ensure RTF operations can be successful immediately upon NUOPC implementation. This process has been successful at ECMWF because visitors can begin making contributions and perform modeling research in days versus months with an easy to use GUI and pre-developed scripts, shortening the "spin-up" time. The interim TTP committee proposes one $200K FTE position (for one year only) at each RTF to write scripts and standardize GUIs across the RTFs, yielding a onetime cost of $600K, beginning in 2013. This recommendation is an expansion of existing functions at the individual agencies. Maintenance of the GUIs will be absorbed into the operational functions and RTFs of the production centers after the initial one-year period.

A common configuration management (CM) process for technology transition processes is also needed across the Tri-Agency. The amount of code that will be handled in the NUOPC era is massive. Codes will have to be periodically synchronized between centers as well as between the operations and research functions. CM personnel at the RTFs will also maintain large global databases and standardized test cases and other metrics used to support the RTFs. 1.5 FTEs are needed at both the Navy and National Oceanic and Atmospheric Administration (NOAA) RTFs: 0.5 FTEs for data assimilation (DA) work, 0.5 FTEs for global modeling, and 0.5 FTEs for ensembles. The Air Force RTF requires approximately one FTE (0.5 FTEs for post-processing and 0.5 FTEs (potentially) for DA, global modeling, and ensembles to maintain Navy and NOAA baselines and back-up capabilities). Thus, a total of four FTEs are recommended for CM work at the RTFs by 2013. Assuming a cost of approximately $200K per year for one FTE, a total of four FTEs would equal $800K per year. The CM process at each RTF would be an extension of current CM processes and would act to synchronize CM requirements at the operational centers.

The total projected cost of the RTFs is approximately $2.8M per year. However, this dollar figure is scalable based upon the amount of operational improvement and Tri-Agency collaboration desired. However, scalability is primarily tied to the number of RTF visitors,

which is estimated to cost approximately $2M per year; the funding for CM processes (estimated at $800K per year) has less scalability since CM is required regardless of how many visitors work in each RTF. However, total costs should generally decrease (increase) as the number of visitors decrease (increase) below (above) the three visitors this committee recommends per RTF.

Finally, the interim TTP committee recommends the establishment of a NUOPC metrics program to incorporate standardized measures of technical capability. This program will necessarily include the development of metrics to assess how well proposed science and code changes improve NUOPC forecast skill and how those changes might affect the operating efficiency of product delivery at member agencies. Additionally, this committee recommends the NUOPC metrics program also measure how well the entire NUOPC program goals are being achieved from a TTP perspective. That is, this committee recommends the NUOPC metrics program also develop metrics to assess if TTP processes are indeed being streamlined as a result of NUOPC implementation. This program requires a onetime cost of approximately $165K, to formulate initial metrics and cover travel costs for customer visits. The initial metrics program will likely be developed under the supervision of the standing TTP Committee by 2011, and would include key measures of NUOPC forecast skill and operational efficiency, as well as identify standard test cases for proposed technological changes to be measured against. This recommendation is an expansion of existing functions at the individual agencies. After the development of the initial products and processes, it is recommended that the NUOPC metrics program be folded into the routine functions of the production centers and RTFs.

In order to accelerate the transition to new technology, eliminate unnecessary duplication among the production centers, and achieve a superior National global prediction capability, a National process for technology transition must be formulated and applied by the NUOPC Tri-Agency. A standing TTP Committee is needed in order to manage key technology transfer functions. These functions include the establishment and management of outreach programs, technology transfer centers (i.e., RTFs), and a strong metrics program. The annual cost of proposed NUOPC TTP functions, as recommend by the interim TTP Committee, is approximately $4.8M per year, beginning in 2013. However, this cost is mostly scalable based upon the amount of research and operational improvements/collaboration desired by the NUOPC Executive Steering Group (ESG).

2. INTRODUCTION.

This report describes the roles, responsibilities, procedures, and methods for establishing and managing the responsibilities and duties of NUOPC's TTP Committee and makes recommendations for the establishment of a technology transition process that promotes NUOPC goals. Upon approval of the NUOPC ESG or other appropriately designated NUOPC body, the details contained herein will be implemented and/or further studied for implementation during NUOPC Phase II, the initial implementation phase.

2.1 Background.

NUOPC is an agreement to coordinate the activities of the Tri-Agency operational weather partners (National Oceanic and Atmospheric Administration (NOAA), U.S. Air Force (USAF), and U.S. Navy (USN)) to accelerate the transition of new technology into operations, eliminate unnecessary duplication and certification efforts, and achieve a superior National weather forecasting capability. NUOPC is currently focused on next-generation systems for global numerical weather prediction (GNWP), particularly global ensemble forecast systems, allowing for possible later expansion into other areas of numerical prediction. This NUOPC TTP Committee report outlines the processes necessary to implement a coordinated and streamlined technology transition process across the Tri-Agency.

2.2 Goals of the TTP Committee.

The overarching goal of the NUOPC TTP Committee is to facilitate a common or collaborative technology transition process to accelerate the development and transition of new and improved capabilities into NUOPC operations. More specific aims of the committee geared toward achieving this goal include:

2.2.1 Develop a collaborative/coordinated technology transition process that streamlines Tri-Agency processes, reduces duplication of effort, and results in improved tools and/or techniques for exploitation by Tri-Agency operations components.

2.2.2 Develop processes and programs that facilitate outreach to the broader research community to encourage involvement by researchers on problems of operational significance.

2.2.3 Develop a process to generate a mutually agreed upon agenda for research activities to include a prioritized list of research areas/needs for the benefit of Tri-Agency operational activities.

2.2.4 Develop a process to generate mutually agreed-to metrics and measurement standards for pre- and post-implementation evaluation of new tools and techniques generated by NUOPC technology transition activities.

2.2.5 Identify/develop configuration management and change control processes to effectively manage pre-implementation activities for new tools and techniques.

2.3 Methodology.

To achieve the goals set forth above and to address various aspects of the technology transition process, the interim TTP Committee established five sub-committees with the following focus areas:

2.3.1 Research Needs. This sub-committee focused on 1) establishing a process whereby the Tri-Agency members can collaborate on a common, prioritized research agenda and 2) identifying potential programs to fund activities of the research community.

2.*3.2 Outreach.* This sub-committee focused on developing programs to encourage involvement of the research community in activities that improve upon operational forecast tools and techniques.

2.3.3 Metrics. This sub-committee developed processes to establish metrics for evaluating emerging technological advancements. The metrics process focused on establishing procedures to measure the expected impacts of new technologies before they are implemented operationally. It is believed that, armed with this information, the NUOPC Unified Ensemble Operations (UEO) committee (through the processes they identify) would then make operational implementation decisions. Post-implementation evaluations will also be performed.

2.3.4 Configuration Management. This sub-committee focused on the processes by which developmental code repositories (including metrics codes and test cases) would be managed such that version control could be maintained.

2.3.5 Change Control. This sub-committee focused on the processes to effectively manage changes to the developmental code baseline and, in particular, potential conflicts related to those changes.

2.4 Organization of this Document.

During the course of deliberations, it became apparent that the goals of NUOPC would best be served by the interim TTP Committee focusing primarily on outreach (to include research needs identification/prioritization and the establishment of technology transfer facilities with configuration management responsibilities at each operational center) and metrics as the broad focus areas. As a result, the organization of this document is such that chapter 3 discusses the outreach process and includes necessary sub-components. Chapter 4 then addresses our proposed technology transition facilities at each center. Chapter 5 addresses metrics, and Chapter 6 serves as a summary of our activities and proposal.

3. OUTREACH PROCESS.

3.1 Purpose.

The purpose of the NUOPC outreach process is to identify and prioritize Tri-Agency research needs, assist agencies in improving their operational capabilities, and to facilitate the

involvement of the broader scientific research community into the NUOPC technology transition process through:

- Improved efficiency of individual agency Research and Development (R&D) activities
- Development of a Tri-Agency coordinated research needs list that prioritizes research requirements and reduces duplication of effort
- Direct funding of critical enabling capabilities (as required)
- Establishment of a process by which the TTP Committee can be made aware of promising science and technology (S&T) and can entrain critical, emerging S&T into the Tri-Agency operational systems (even if the new technologies are not specifically identified as requirements beforehand through the operational requirements process)

These goals are challenging because, at present, NOAA, USAF, USN and other federal R&D sponsors for research in the atmospheric sciences each have independent procedures regarding tasking, funding, developmental testing, and integration of research into operations. Funding for each of these activities is not consistent across the agencies as some agencies have funding profiles for the full spectrum of research, while others only have funding for applied research and integration into operations. Applied R&D, as well as basic S&T in the atmospheric sciences, is currently conducted at various government agencies and by government funded academia at large. Each member of the Tri-Agency identifies and prioritizes research needs based on operational capability gaps or new opportunities becoming available from the scientific community, but does so independently. Each agency has its unique method for funding decisions based on operational needs and technology maturity, among other things, and has its own methods for communicating those needs to the research community, enticing researchers to work on these particular problems, and managing the resulting efforts. Scientific symposiums, targeted workshops, joint user community and research conferences, targeted requests for proposals and requests for information (RFI), and other meetings and performance reviews are all methods used for this effort. As a result, it will be challenging to coordinate these efforts among the Tri-Agency members in an efficient fashion that reduces duplication of effort and creates a synergistic relationship with the Tri-Agency members and the research community at large.

3.2 Research Agenda.

As a first step to facilitate R&D and S&T coordination among the Tri-Agency partners, it is recommended that NUOPC management, through a standing TTP Committee, release a coordinated, prioritized list of operational needs and research requirements every two years. While it is recommended this requirements list not be considered prescriptive, research sponsors within the NUOPC agencies should take it into consideration when making funding decisions. It is also expected that this list and discussions at the meetings which develop this list will aid in reducing duplication of effort between the agencies and increasing entrainment of the research community towards NUOPC goals. Additionally, it is recommended that a Fellowship Program be created for laboratory and academia performers to propose against and that this program be

collaboratively managed through a joint proposal review process. This program will augment and leverage other sponsored research, rather than solely relying on a "volunteer" collaboration for multi-model ensembles such as is done in The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) or the North American Ensemble Forecast System (NAEFS) efforts. Finally, sponsorship of dedicated workshops or special sessions at larger community conferences, such as American Geophysical Union (AGU) or American Meteorological Society (AMS) conferences, should be considered if the budget permits.

3.3 Reaching Out to the Research Community.

Because most of the basic research community (i.e., universities) and the basic to applied research community (e.g., Federal Laboratories and University Affiliated Research Centers (UARCs), such as university-affiliated Applied Physics Laboratories (APL)) are required to recoup most or all of their salary and expenses through sponsors, it can be challenging to encourage them to work on problems directly relevant to operations. In order to leverage basic research funding from organizations such as the National Science Foundation (NSF), problems need to be both scientifically interesting and relevant to the interests and expertise in the wider basic research community so that positive peer review will result in favorable funding decisions. Otherwise, the operational agencies will have to completely fund the effort which is often not possible given current budget constraints. Alternatively, agencies could fund internal researchers or sole source contractors; however, this does not always guarantee sufficient subject matter expertise for optimal and effective results.

While many researchers at the Federal Labs and UARCs are anxious to see the results of their work used practically, they still need to propose each particular effort to either their internal management sponsors or to other sponsoring agencies. This management structure results in performers often proposing work to the sponsor on the ultimate end-users' behalf with no direct communication between sponsor and end-user. This can result in obstacles to transition in that end-user commitment to, or endorsement of, the individual research effort funded by a third-party sponsor does not always result in effective transition when it comes time for the end-user to take over operational funding or complete efforts for final transition to operations. Methods for the end user to be represented at the NUOPC ESG level will therefore need to be explored to endorse needed directions and desired capabilities. It is recommended that these methods include an annually reviewed prioritized list of operational needs, sponsorship of conferences, workshops and other forums, managing a Fellowship Program, and so on. The challenge here is in ensuring that prioritized goals are realistic, reasonable as to timelines and implementation, resourced appropriately at both the basic, applied, and transition work levels, and tied back in with the researchers so that new capabilities, innovations, and breakthroughs are incorporated into the short and long range operational visions.

This collaboration could include focusing research on topics of particular interest to the various operational centers, testing and evaluation of the numerical weather prediction (NWP) systems in

operations, and pre-implementation testing of proposed changes to the operational system. It could also include evaluation of next generation NWP components for possible elevation to operations. To attract collaborators from the research community it is recommended that there be incentives in place. These incentives should include:

- A Fellowship Program directed towards long-term R&D goals
- A Visiting Scientist Program to perform technology transfer functions at the production centers
- Leveraging researcher facilities such as dedicated or shared computer resources
- Access to data archives, special observations, and remote sensing resources
- Access to NUOPC numerical code libraries including the operational codes and assistance in running and modifying code libraries
- Participation in NUOPC through integrated product teams and other similar processes
- Invitational travel grants for participation in NUOPC or community forums

3.4 Long Term National Agenda and Short Term Management Plan.

R&D requirements and the desired capabilities need to be clearly articulated by end users. It is recommended that overarching five and ten year plans be developed and reviewed at least every three years and a management plan be updated every two years through a panel consisting of agency production center representatives, operational users, and subject matter experts from the science community. Each agency may further maintain a subordinate list of agency-unique research needs. A process will be established wherein the NUOPC staff will query the member agencies in order to develop a consolidated list of common operational needs, which will then be prioritized by agency representatives and reviewed for scientific and implementation feasibility, implementation schedule (i.e., long term vs. short term efforts), and adequate resourcing. This research plan will then be approved by the ESG and be the basis for a call for proposals through Broad Agency Announcements (BAA), or other programs (e.g., the NRC Fellowship Program), and the short term management plan. It is recommended that a standing NUOPC TTP Committee be formed to manage this process, produce a prioritized listing of research/operational requirements, the five and ten year plans, and the management plan for NUOPC TTP activities.

Resources need to be focused on the required R&D through an established iterative process between operations (provide what is critical), production (provide what is feasible to implement), and researchers (provide advice on what is possible). It is recommended that the NUOPC staff (i.e., the standing TTP Committee) develop, through joint technical meetings with the primary development community, a list of research and development priorities (a National agenda) based on the approved list of operational needs.

Review and oversight at appropriate intervals is needed for operations to retain flexibility in the face of changing priorities and to keep R&D from proceeding off track to an end point that may no longer be desired. It is recommended that the NUOPC staff convene annual management and

technology reviews. During these reviews, GNWP research and development efforts will be reviewed to identify those that are addressing established NUOPC goals.

The short term management plan should be based on available NUOPC funding for the research agenda items, clearly articulated with performers, milestones, expected outcomes, delivery mechanisms for results, and training or documentation that will be required. The plan should also specify who signs off as approving authority and who is responsible between the researchers and production site programmers for testing, delivery dates, test reports, and final deliverables with respect to code, publications, or other products.

NUOPC management staff should provide recommendations for accelerated research and development to the Tri-Agency via the ESG. It is recommended that NUOPC management staff direct available focused research and development funds to address the most pressing common requirements or the most promising common operational capabilities.

3.5 Potential Risks and Roadblocks.

The major challenges and risks to this effort involve ensuring an adequate process is established to create and regularly update a detailed and robust set of coordinated research priorities, sufficient agency financial, personnel and management support to fund and execute a credible Fellowship Program, and a process to ensure the agencies are adequately engaged and represented, and the results are effectively transitioned.

There are some risks involved in engaging with the broader community with respect to the effectiveness of this effort and the reputation of the involved agencies. Ensuring that adequate funding is secured is a primary risk. Additional risks include establishing regular interactions so that focused research efforts are competently managed and that the operational centers are also resourced to implement the resulting capabilities, and that the developed priority list of research topics is realistic and reasonable. Each operational center should select a competent liaison to the NUOPC TTP research group that also has the authority to make commitments. Also, access to high-performance computing and other computer and shared national resources, as well as federal laboratories and production centers needs to be streamlined and improved.

Additionally, as there is no new money confidently projected, a NUOPC research program would be in direct competition with current efforts in single-model ensembles, deterministic models, mesoscale forecast systems, aerosol modeling, remote sensing and other current investments at the operational prediction centers.

Furthermore, entrainment of foreign nationals with expertise in multi-model ensembles and GNWP efforts from the United Kingdom Meteorological Office (UKMO), ECMWF, Australia's Bureau of Meteorology (BOM), the Japan Meteorological Agency, the People's Republic of China Meteorological Service, and National Taiwan University's National Meteorological Center, could be challenging.

3.6 Costs.
It is recommended that a Fellowship Program be established to work on NUOPC's research agenda. Fellowships can be easily directed towards fulfilling long-term R&D goals. Ideally, each production center will host up to three fellows on two year cycles. A Program Manager is needed to manage the fellows, for a total of 10 full time equivalent (FTE) positions. Estimating $200K per year for one FTE yields a total of $2M per year for 10 FTEs, beginning in 2013. This dollar figure is scalable based upon the amount of research the Tri-Agency wants performed. In addition to working on the research agenda, fellows would also be expected to visit the production centers for one to two months per year to interact and exchange ideas with operators and visiting scientists to enable effective collaboration. Travel costs by the fellows are included in the $2M per year recommendation.

3.7 Recommendations.
(1) NUOPC management, through a Research Needs Panel on a standing TTP Committee, produce a coordinated, prioritized list of operational needs and research requirements every two years. Overarching five and ten year plans will also be developed to guide NUOPC research activities.

Recommended process: NUOPC staff should query Tri-Agency members to develop a consolidated list of common operational needs, which will be prioritized by agency reps and reviewed for scientific and implementation feasibility, scheduling, and adequate resourcing. End users at the ESG level will endorse needed directions and desired capabilities in the plan. Five-year and ten-year overarching plans should be developed and reviewed at least every three years. The management plan/currently funded efforts list should be reviewed and prioritized every two years through a panel consisting of agency production center representatives, operational users, and subject matter experts from the scientific community. The list should be released through the standing TTP Committee.

(2) Establish a Fellowship Program, through the Research Needs Panel on the standing TTP Committee, for laboratory and academia performers to propose against. Assess agency willingness to contribute resources. The recommended structure of the Fellowship Program is three fellows each per center (Navy, NOAA, and Air Force), on two year cycles, with a Program Manager at the NUOPC management level administering the collective Fellowship Program. The two year National research agenda and Fellowship Program cycles will be offset by one year (e.g., the research agenda is generated in odd years, and new fellowships start in even years).

Recommended process: Effect collaborative management through a joint proposal review process and individual agency management and contracting of their share of the joint contribution. Use the research priorities list as the basis for the call for proposals and the short term management plan. Ensure that goals are resourced appropriately at both the basic, applied, and transition work levels. Resources should be focused through an iterative process between operations, production, and research. The NUOPC staff will develop, through joint technical

meetings with the development community, a list of R&D priorities based on the approved list of operational needs. NUOPC management staff will direct focused R&D funds. Regular management and technology reviews will be held.

(3) Sponsor dedicated workshops or special sessions at large community conferences such as AGU or AMS. Sponsorship and organization can be done on a rotating basis among the agencies.

(4) Support research facilities that provide computer resources, access to NUOPC software, data archives, and observations.

(5) Allow outside researchers to participate in NUOPC through integrated product teams, travel grants for participation in community forums, and other similar processes.

4. RESEARCH TRANSITION FACILITY (RTF).

4.1 Purpose.

A method for transitioning research advances into NUOPC is required to accelerate operational improvements. Therefore, it is recommended that a Research Transition Facility (RTF) be established at each Tri-Agency production center to address technology transfer objectives. However, before discussing the proposed NUOPC RTFs, and to assist in an understanding of how new technology and advancements are currently incorporated, a short review of existing Tri-Agency technology transfer processes is necessary.

4.2 Existing Tri-Agency Technology Transfer Processes.

4.2.1 NOAA.

The National Weather Service's National Centers for Environmental Prediction (NCEP) are comprised of nine individual centers. Two of these centers are of primary interest for NUOPC—NCEP Central Operations (NCO), which executes the operational suite of numerical analyses and forecast models and prepares the products for dissemination, and the Environmental Modeling Center (EMC), which develops and improves the numerical weather prediction systems through broad partnership with the research community. EMC identifies new or improved software algorithms for testing, integrates that software with the operational codes, and performs extensive retrospective and real-time testing and evaluation prior to transferring the software to NCO for implementation. Thus, EMC is NCEP's research and technology transfer agent, and it is expected that they would continue to perform this role as host for an RTF for NOAA under our recommended plan for NUOPC.

4.2.2 U.S. Navy.

Fleet Numerical Meteorology and Oceanography Center (FNMOC) has global modeling responsibility for all of DoD and as the Navy's corporate laboratory, the Naval Research Laboratory (NRL) provides the DoD's only full-spectrum research program in NWP.

Furthermore, NRL functions as a national and international scientific collaborator for the Navy, as well as the primary developer and transition agent for global NWP and global ensemble prediction systems for operational application at FNMOC. The collocation of the research and operational centers has proven to be a very effective model for technology transfer and under the NUOPC partnership, it is expected that NRL would continue to function in this role. The laboratory is well poised to successfully compete for funding to perform in this manner, and has a proven and reliable track record of successful collaboration and transition following the formal process governed by the Navy's Administrative Model Oversight Panel.

In addition, Information Assurance (IA) issues require a security clearance to work directly on FNMOC computers, where most development is done and, in the future, remote access may even be prohibited, as that has already become the case for some systems. Therefore, the collocation of NRL with FNMOC, with NRL acting as a host for an RTF, will be critical for effectively engaging the outside research community to work on aspects of the global NWP problem important to DoD operational interests. NRL employees will be a critical link, serving as intermediaries and advisors, as well as collaborators in the development and evaluation of systems that will run effectively and efficiently on FNMOC computer systems.

4.2.3 U.S. Air Force.

The Air Force Weather Agency (AFWA) does not run a global atmospheric model, but uses global products from NCEP and FNMOC to provide information to various regional models, specialized models, and other meteorological applications. AFWA also post-processes global model data from NCEP and redistributes it to DoD users. AFWA makes use of community-based modeling systems, where available, and technology transfer is done on-site by AFWA staff and various AFWA contractors. It is currently envisioned that AFWA's role in NUOPC will be to perform post-processing of the global ensemble members from NOAA and the Navy, to create products for direct customer support and for use in downstream applications. An RTF focusing on ensemble post-processing would be hosted at AFWA for scientists interested in research in this exciting new field.

4.3 RTF Functions.

The NUOPC RTFs proposed by the interim TTP Committee would have two overarching technology transfer objectives, designated as "outward" and "inward" facing. *Outward* facing objectives refer to goals of improved collaboration and communication with the broader research community by the production centers. *Inward* facing objectives pertain to those actions taken to assist NUOPC member agencies in improving their operational capabilities. This section describes the typical functions and processes that would likely need to be accomplished by NUOPC RTFs.

4.3.1 Outward Facing Objectives.

In order to incorporate research advances into operations, NUOPC must collaborate with and encourage involvement from the research community. The most effective manner in which to

accomplish this is through a Visiting Scientist Program at the RTFs, collocated with the Tri-Agency production centers.

To support outward facing goals, it is recommended that the RTFs serve as software and data repositories. These repositories should include source codes for all releasable GNWP and ensemble system components:

- Data pre-processing
- Data quality control
- Data assimilation systems (e.g., land surface, sea surface temperature, etc.)
- Radiative transfer models
- Land surface models
- Model initialization
- Global forecast models
- Input/Output routines
- Ensemble methods
    - Initial condition perturbations
    - Boundary condition perturbations
    - Physics perturbations
- Ensemble post-processing

The repositories would also need to contain other software tools depending on their functions. Examples are:

- Common configuration management system (i.e., Subversion)
- Executables for all source codes for ease of testing individual components
- Scripts for running the entire system (or portions of the system)
- User-friendly graphical user interfaces (GUI) to enable easy system set-up and tailored output
- Earth System Modeling Framework (ESMF) infrastructure
- Web-based observation monitoring/data impact tools
- Post-processing and product building
- Data management tools
- Copies of operational libraries, where possible, or suitable alternatives

To support the NUOPC metrics program, which will be discussed in section 5, test case data should also be included in the software and data repositories:

- Standard test cases
- Global observational database

    - Rawinsonde observations (RAOB)
    - Data from different atmospheric satellite sensors
    - Feature tracked winds
    - Land surface stations
    - Aircraft, including Unmanned Aerial Vehicles (UAV)
    - Ships
    - Fixed and drifting buoys
    - Coastal-Marine Automated Network (C-MAN) stations
    - Sea surface temperature (derived from various platforms)
    - Special Sensor Microwave Imager (SSMI) ice concentration
- Model history restart files
- A year's worth of global tropical cyclone cases (in 2008, there were 102 storms)
- The ability to grab and save data for interesting case studies

In addition to serving as a software and data repository, RTFs would provide other support services including:

- Documentation of all supported software
    - In-line documentation of all code, down to and including each subroutine
    - Software design documents that provide the "big picture"
    - Users guides that explain how to execute software
- On-line information on software coding standards of the operational centers
- Metrics to guide software evaluation
- Help desk functions for remote users
- "Hands-on" support for visiting scientists
- Periodic training classes/material

Lastly, in order to function effectively, RTFs will require the following resources:

- Operational-like computational resources to support research, development, test and evaluation (RDT&E)
- Large data storage capacity
- Up-front investment to develop user-friendly tools for GNWP/ensemble execution
- New funding for Visiting Scientist Program
- Staff and resources to support visitors from the research community

4.3.2 Inward Facing Objectives.

The preceding section addressed how to facilitate coordination and collaboration with the research community so NUOPC could leverage their research advances. This section will

outline how those research advances could be incorporated and applied towards operational improvements.

After performing the necessary outward facing functions to engage the broader research community, individual agencies will then turn toward inward facing functions to improve operational capabilities. It is recommended that a cost benefit analysis be performed first before pursuing any new course of action. The next step in the process is a Demonstration and Validation (DemVal) phase to assess the value of model code changes. It is envisioned that the following functions would be performed by NUOPC RTFs during DemVal:

- Preliminary transition review
    - Ensure system addresses anticipated operational need
    - Ensure sufficient operational resources are available
    - Ensure system will meet downstream model requirements
    - Ensure scientific/technical documentation is sufficient to move forward
- Source code
    - Configure for operational environment
    - Include in-line documentation adequate for cursory review
    - Include capability to produce all output required for validation testing
    - Ensure code meets all operational coding and architecture standards
    - Ensure code is compatible with operational configuration management
    - Ensure code meets all Information Technology (IT) and Information Assurance (IA) certification requirements
    - Correct code errors/problems as encountered during DemVal
- Documentation
    - Deliver scientific/technical documentation as needed to justify transition
    - Deliver approved transition plan
    - Deliver approved validation test plan
    - Deliver preliminary software design document
    - Deliver preliminary software users guides
- Demonstration (Pseudo Operational)
    - Execute system within operational environment constraints
        - Execute within operationally available time windows
        - Execute with operationally available libraries and data
        - Provide output required to drive all needed downstream systems
        - Provide input/output (I/O) adequate to meet all operational product requirements
    - Run system stably for sufficient cases to address validation test needs
        - Include both archived test cases and pseudo real-time runs
        - Provide output products required to address validation test plan
        - Rerun all tests as required following code changes and corrections

- Validation
    - Address all tests required by a validation test plan
    - Ensure system meets validation standards relative to current operational system
    - Include external review of validation test results
    - Deliver validation test report with recommendation for transition or additional development

After completion of the DemVal phase, the next step in the "research into ops" process is the Operational Implementation phase. During this phase, operations testing and transition to operations is completed and coordinated between each production center's RTF and production facility. The following steps would generally occur at the NUOPC RTFs during the Operational Implementation phase:

- Operational transition review
    - Ensure system addresses clearly defined operational need
    - Ensure sufficient operational resources are available
    - Ensure system addresses all downstream model requirements
    - Ensure system documentation is sufficient for operational use
    - Ensure life-cycle support requirements are addressed
- Source code
    - Deliver code fully configured for operational environment
    - Freeze code for duration of operational testing to ensure there are no undesired changes
    - Provide for full operational control with no developer intervention
- Documentation
    - Deliver approved operational test plan
    - Deliver final software design document
    - Deliver final software users guides
- OPSCHECK
    - Ensure code runs stably within full operational environment
    - Ensure execution meets operational time windows
    - Ensure code produces all output needed to drive downstream systems
    - Ensure code conforms to operational configuration management
- OPSTEST
    - Run system in full operational mode (i.e., parallel operations)
    - Run system stably for sufficient time to meet OPSTEST plan
    - Assess system ability to meet operational skill requirements
    - Deliver OPSTEST report with recommendation for transition or further development

Lastly, the typical final steps in the process would include:

- Deliver final source for operational configuration management
- Deliver life-cycle support and maintenance plan
- Deliver upgrade plan
- Deliver final documentation

Figure 1 depicts the actions and coordination elements of the RTF model, and serves as a graphical summary of the preceding sections.

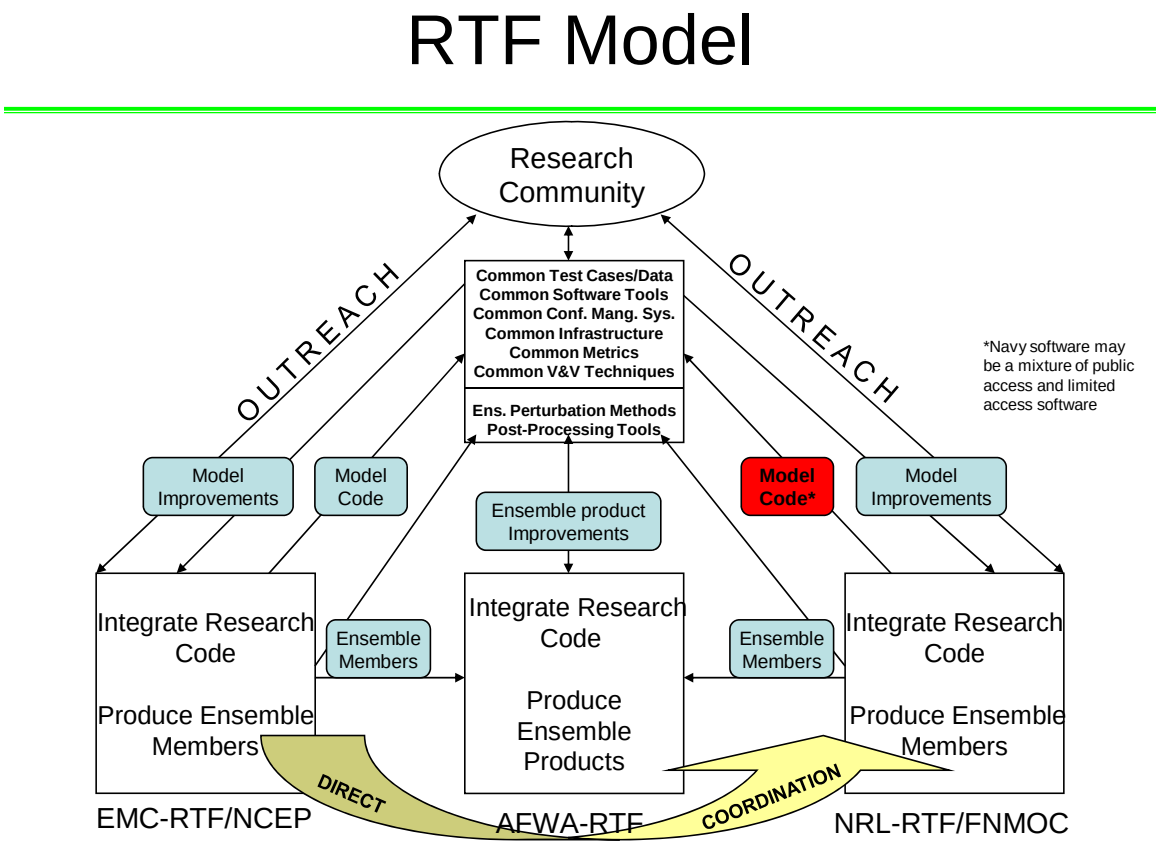


Figure 1. RTF Model.

4.4 Distributed Testbed Center (DTC).

During the development of this report, interim TTP Committee members conferred several times and deliberated over the best way to perform technology transfer functions in the NUOPC era. Two primary courses of action emerged as a result of these discussions: an RTF at each of the production centers (as described in this chapter), or a single, consolidated technology transfer

function external to the production centers, dubbed a Distributed Testbed Center, or DTC (see Appendix 3 for details on how a NUOPC DTC might be designed).

Several committee members had reservations about a NUOPC DTC. First, they were concerned it would be too much like the Weather Research and Forecasting (WRF) DTC, which is the result of a grassroots effort that focused on enabling the research community. NUOPC is not a grassroots effort, it is being downward directed from the top, and its focus is on the operational community versus the research community.

Second, the current WRF DTC construct manages a fine-scale model. There are increased logistical obstacles to overcome with global ensemble systems—the current focus of NUOPC. Due to the sheer size of the global database, and the output produced by global ensembles, long-haul communications issues and associated bandwidth requirements will be demanding. Consequently, it was thought a centralized DTC facility might be less than ideal.

Third, IA obstacles pose a considerable risk to a consolidated, external technology transfer function, such as a DTC. DoD limitations include access restrictions to facilities, computer networks, and software codes. Many NWP researchers are foreign nationals. The DoD is unlikely to openly release all NUOPC codes to the public. Furthermore, there are IA limitations between USAF and USN. Just because certification and accreditation (C&A) processes are satisfied by one agency, doesn't automatically mean it will pass another agency's C&A processes. A single DTC would face many difficulties working under these conditions.

Fourth, each agency has different missions and decision-making processes. Joint decision making and the full alignment of NUOPC processes among the Tri-Agency would be a daunting task, and it would impede the rapid transition of new technologies or improvements into operations. Thus, Tri-Agency members should retain their autonomy, something that would be difficult to do under the single DTC construct.

Fifth, a DTC would likely cost twice or three times as much as the RTF solution. In order for a NUOPC DTC to function effectively, it would have to emulate each of the Tri-Agency production centers, which would necessitate the purchase of additional resources and computing power.

Finally, using the European meteorology centers as an example, technology transfer functions work best when collocated with production centers. ECMWF and UKMO are successful with technology transfer at their centers because they have technology transfer functions onsite, they procure sufficient computing resources for R&D, they align researchers with operators and ensure researchers understand the operational requirements, and they provide resources to staff

and support visiting scientists. Furthermore, DemVal at ECMWF and UKMO is performed at the operation centers, not offsite, as an external DTC would require.

For these reasons, the majority of the interim TTP Committee members agreed that the RTF structure described in this document was the most likely arrangement to meet NUOPC goals. It could be argued that a single DTC would be more effective at outreach through its ability to engage the broader research community with fewer access restrictions (for example) and that the RTF solution does not promote the national NUOPC objectives too much beyond "business as usual." RTFs could lead to a NUOPC DTC in the future; however, based on current monetary, computer processing, and security restraints, and all the other reasons cited above, the majority of interim TTP Committee members agreed the RTF solution was the best course of action. Furthermore, the proposed RTFs differ from "business as usual" in the following manner:

- RTFs will maintain all of the releasable, shared software under ESMF
- RTFs will have common test cases with common datasets for testing
- RTFs will have common user-friendly software tools to allow them to easily design and set-up test cases and analyze results
- RTFs will work with common metrics and software for analyzing multi-model ensembles
- RTFs will have coordinated and common sets of techniques for perturbing ensembles
- RTFs will have the common ability to produce multi-model ensembles, if possible, or to work together to produce multi-model ensembles, when needed, for testing
- RTFs will assist in the establishment of a Visiting Scientist Program to enable NUOPC outreach efforts
- RTFs will have personnel dedicated to assisting researchers in running software in the operational environment
- RTFs and operational centers will use the same configuration management software and will periodically synchronize their codes

Additionally, the interim TTP Committee considered pros and cons of both the RTF and DTC options to ensure all aspects of each were considered. The following tables summarize the advantages and disadvantages of each alternative.

Table 1 outlines the pros and cons of a DTC:

Advantages	**Disadvantages**
Strong national collaboration among the Tri-Agency	Logistics of downloading and running global model data from an off-site location are cumbersome and costly
Strong centralized management	Information assurance concerns associated with the sharing of data will cost a lot of time and money to overcome, if at all
Strong outreach program with easier research community accessibility	Performing research transition from a centralized off-site facility is more difficult and has proven to be less successful than performing the function on-site
	Joint decision making and the full alignment of NUOPC processes across the Tri-Agency will unduly hamper the rapid transition of technology into operations
	Higher cost, particularly for purchasing computer resources to emulate each agency's production center
	More duplication of effort
	Requires competitive selection (and periodic recompete) if the DTC is located at a non-government site
	More obstacles to overcome in getting funding and buy-in from the individual NUOPC agencies to support a centralized off-site facility

Table 1. DTC Advantages and Disadvantages.

Table 2 outlines the pros and cons of the RTF proposal:

Advantages	**Disadvantages**
The logistics of managing global model data are less challenging	Collaboration among Tri-Agency members will not be as unified
Information assurance problems are less prohibitive since each RTF will be directly linked to its production center, adhering to its center's certification and decision-making processes	Less centralized Tri-Agency management
On-site research transition, such as the model followed in Europe, has a proven track record of being more effective	Foreign nationals may find it more difficult to perform NUOPC research in the more restrictive RTF construct
One-third to half the cost of the DTC	
Each NUOPC Tri-Agency member retains autonomy with respect to code changes, timetables, scheduling, etc.	
More effective at translating research into operations	
Less duplication of effort	
Utilizes government facilities, avoiding contract competitions	
Less obstacles to overcome in getting funding and buy-in from the individual NUOPC agencies to support on-site technology transfer functions	

Table 2. RTF Advantages and Disadvantages.

During the course of committee deliberations, a third option was also discussed: a DTC/RTF hybrid. This hybrid consisted of a centralized DTC with three satellite RTFs, one at each production center. However, it was determined that the sharing of code and data would still present many challenges under this scenario. Additionally, a DTC was seen as adding a superfluous layer to the RTF model. Furthermore, the majority of the interim TTP Committee members were concerned with the cost of maintaining a DTC (whether separately or under the hybrid version), and the duplication of effort and resources required to make the DTC function effectively. Thus, the hybrid DTC/RTF was not considered a viable alternative.

Thus, based on the above analysis, the majority recommendation of the interim TTP Committee is that RTFs, structured as described in this section, be adopted as the model for NUOPC technology transition activities. Interim TTP Committee members felt the RTF option had the best chance for success given funding, security, and logistical constraints, and the goals of such a technology transition facility.

4.5 Potential Risks and Roadblocks.

Interaction with the research community poses considerable risk since there are many foreign nationals working on NWP R&D projects at laboratories and universities around the globe. Many of the codes that will be part of NUOPC will be developed by DoD funding and therefore fall under International Traffic In Arms Regulations (ITAR) and raise IA concerns. However, the RTF construct should minimize many of these issues since individual agencies would have greater control over hiring decisions and code release.

It is likely that some NUOPC participants may not want to put parts of, or even any of their codes in the public domain due to proprietary or intellectual property interests. This could be a potential obstacle to overcome.

Managing the impact of changes on any individual model in the multi-model ensemble is another potential issue. For example, if one agency adopts the data assimilation, parameterization, dynamic core, etc. from another system, it may improve their bulk performance metrics, but degrade ensemble performance. In this case, which impact of the change is more important?

Sufficient computer resources, transition support resources, scheduling, milestones, configuration management, release dates, etc., are also important matters that must be properly considered.

Lastly, the USAF or USN could, in the future, decide to run model code from one of the European centers. This code would probably be licensed and the license agreement would most likely not allow unrestricted distribution of the code, which could pose several challenges.

4.6 Costs.

Ideally, each production center will host up to three visiting scientists for RTF work on-site, beginning in 2013. Calls for visitors would likely be made via Broad Agency Announcements,

or through other programs, such as the UCAR VSP. Additionally, a PM would be needed to manage the visitors, resulting in a total of 10 FTE positions. (However, the PM for the Fellowship Program may be able to manage the visiting scientists as well, depending on workload, reducing the RTF FTE recommendation by one FTE.) Estimating $200K per year for one FTE yields a total of $2M per year for 10 FTEs. This dollar figure is scalable based upon the amount of funding and operational improvement desired. This estimate doesn't include the cost of computers and people on-site to assist the visiting scientists.

In order to assist the visitors and to ensure that visiting scientists are able to make positive contributions to the NUOPC TTP process as quickly as possible, some script writing will likely be required to have standard NWP tools immediately available to visiting personnel. A number of programmers would be necessary at each RTF to write these scripts as well as develop user-friendly GUI programs to reduce visitor spin-up time (this is very similar to what is done at the European centers). This process is successful in Europe because visitors in the European centers can begin making contributions and perform modeling research in weeks versus months with easy to use scripts and GUIs. It is recommended that one $200K FTE position (for one year only) is needed at each RTF to write scripts and standardize GUIs across the RTFs, yielding a onetime cost of $600K, beginning in 2013. This recommendation is an expansion of existing functions at the individual agencies. Maintenance of the GUIs will be absorbed into the operational functions and RTFs of the production centers after the initial one-year period.

A common configuration management (CM) process is also needed across the Tri-Agency. The amount of code that will be handled in the NUOPC era is monolithic: multiple global models, land surface models, ocean/ice/wave data assimilation (DA) systems and (potentially) models, data processing and quality control systems, model initialization schemes, model and DA system adjoints, multiple methods for producing global ensembles, software for producing multi-model ensembles, the ESMF infrastructure, data monitoring tools, graphics, post-processing, and product building software. It is recommended that the RTFs use the software tool Subversion to standardize NUOPC CM software across the Tri-Agency. NUOPC codes will have to be periodically synchronized between centers as well as between the operations and research functions. CM personnel at the RTFs would also need to maintain the large global databases and standardized test cases used to support the RTFs. It is the recommendation of this committee that 1.5 FTEs would be needed at each of the Navy and NOAA RTFs for CM activities: 0.5 FTEs for data assimilation (DA) work, 0.5 FTEs for global modeling, and 0.5 FTEs for ensembles. It is further recommended that the Air Force RTF would require one FTE: 0.5 FTEs for post-processing and 0.5 FTEs (potentially) for DA, global modeling, and ensembles (to maintain Navy and NOAA baselines, and back-up capabilities). Thus, a total of four FTEs are recommended for RTF CM activities by 2013. Assuming a cost of approximately $200K per year for one FTE, a total of four FTEs would equal $800K per year. The CM process at each RTF would be an extension of current CM processes and would act to synchronize CM requirements at the operational centers.

Thus, the total projected cost of the RTFs is approximately $2.8M per year. However, this dollar figure is scalable based upon the amount of operational improvement and Tri-Agency collaboration desired. However, scalability is primarily tied to the number of RTF visitors, which is estimated to cost approximately $2M per year; the funding for CM processes (estimated at $800K per year) has less scalability since CM is required regardless of how many visitors work in each RTF. However, total costs should generally decrease (increase) as the number of visitors decrease (increase) below (above) the three visitors this committee recommends per RTF.

Lastly, significant computational resources will be needed to meet model development and technology transition requirements. These resources may or may not be available in the processing centers today. Therefore, a thorough review of RTF computational requirements and a cost assessment is required and highly recommended to ensure computational requirements of the RTFs will be met by current and future computer systems at each center. In the event that either current or planned computer systems are inadequate to provide resources for RTF operations, additional NUOPC funds would be required to obtain necessary resources.

4.7 Recommendations.

(1) Establish an RTF at each of the Tri-Agency production centers (to include AFWA).

RTFs will require computer processing power and storage that mimics the operational structure of their respective production centers. RTFs will maintain all of the releasable, shared NUOPC software. In addition, they will store common test cases and datasets for testing. RTFs will also have personnel dedicated to assisting researchers in running software in the development and operational environments.

(2) Establish coordinated managerial oversight of the three RTFs.

Coordinated collaboration among the three RTFs will be administered through the standing TTP Committee, or other appropriately designated body.

(3) Establish a common CM system among the RTFs.

The RTFs and operational centers will use the same CM software, Subversion, and will periodically synchronize their codes. Dedicated personnel are needed at the RTFs to perform this function.

(4) Establish a Visiting Scientist Program.

The great wealth of talent available from the broad research community will be leveraged through a strong outreach program. The recommended structure of the visitor program is three visitors each per agency (Navy, NOAA, and Air Force) with a Program Manager at the NUOPC management level administering the collective Visiting Scientist Program. The Program

Management of the visitor program could be accomplished by the same person managing the recommended Fellowship Program (see chapter 3).

(5) Develop user-friendly GUIs to assist the visiting scientists.

RTFs will have common easy-to-use software tools to allow visitors to easily design and set up test cases, as well as analyze results.

5. METRICS PROCESS.

5.1 Purpose.

Within the interim TTP committee, the metrics sub-committee was tasked to define the process by which the Tri-Agency develops, maintains, and upgrades joint (or National) metrics for the multi-model global ensemble that forms NUOPC. In this document, the term "metrics" will take on several meaning. "Metrics" will be used to refer to tools for measuring the overall forecast skill of NUOPC numerical output and, to fully assess the impact of NUOPC on the operational forecast centers, the discussion of metrics will be extended to include derivative product guidance that each agency provides to its respective customer base. For example, the ability to quantify uncertainty in a forecast may be as valuable, or more valuable, than the overall skill of a particular forecast. Or the impact of a change on ocean model products may be equally important as looking at atmospheric metrics alone. Other system performance goals, such as computational efficiency, computer resource requirements, code portability, and other factors that affect an agency's ability to utilize and execute the global ensemble will be important to consider when defining metrics. Finally, metrics for strategic goals other than superior prediction will also be required. We will need to define methods for determining whether or not NUOPC is indeed leading to accelerated transition of new technology, eliminating duplication of effort, and providing the partner agencies with an acceptable level of return on their investment.

5.2 Current Status.

At present, NOAA and the U.S. Navy each have their own standardized test cases and metrics process for their global prediction systems, the Global Forecast System (GFS) and the Navy Operational Global Atmospheric Prediction System (NOGAPS), respectively, and for the global ensemble prediction systems based on these models. While metrics for global deterministic forecasts are historically well-defined, with each agency having its own mission-related priorities, appropriate metrics for use with ensembles is a less mature science and will be an area for continued research. Therefore, it is important that the NUOPC metrics process not only consider how metrics for the multi-model ensemble will be defined initially, but also how they will be modified and improved as research brings forward new ideas and capabilities for measuring performance and impact.

Both NOAA and the Navy perform testing on their dependent modeling systems and downstream applications, as deemed necessary, while publishing/communicating information

regarding projected hardware or software changes to other user agencies (e.g., Air Force, Federal Aviation Administration, Environmental Protection Agency). These other agencies must then determine what impact/change the announced modification will have on their own modeling systems and applications and ultimately to the support they provide to their customers. Often times, fully tested changes in common software at one agency are re-tested at the other (user) agency(s) to determine impacts to user agency customer support (accuracy, timeliness, etc.). One issue for consideration is whether a more comprehensive common set of metrics, i.e., joint test cases and metrics, could reduce the amount of re-testing that is required by multiple agencies in cases where software applications are shared in whole or in part.

5.3 Metrics.

5.3.1 Performance Metrics.

One of the primary NUOPC goals is to "achieve a superior National global prediction capability." Thus, before establishing a process to define metrics that will assist in determining whether NUOPC is providing a superior prediction capability, we must first establish a baseline that better describes what the new capabilities provided by NUOPC should be measured against. Since both the Navy and NOAA are currently running operational global ensemble prediction systems, we propose that the overarching standard for NUOPC is that the multi-model global ensemble must result in a combined or merged global ensemble that is "better" than either agency's existing ensemble system evaluated individually. This issue of what comprises a better ensemble (e.g., more skillful, more valuable, etc.) will be the centerpiece of any discussion that proposes to establish metrics for NUOPC. Most importantly, that judgment does not have to be founded solely in terms of performance metrics. There may be cases where accepting a merged ensemble whose performance skill is "as good as" the individual ensemble skill would be acceptable if other NUOPC metrics of success were being met.

5.3.1.1 Deterministic Forecast Metrics.

Standard performance metrics for deterministic GNWP skill include widely accepted statistical metrics such as anomaly correlations (AC) and root-mean-square (RMS) errors (see Appendix 4 for one such example). These metrics are typically computed at multiple levels for various regions around the globe, using standardized test cases that require running complete data assimilation cycles from the operational global dataset for several months at a time, from at least two different seasons. Other more incident-driven metrics may be included in the testing process to evaluate the impact of changes to the model or its data assimilation system. For example, tests may be performed to evaluate how the proposed model modifications affect tropical cyclone track error, or the forecast position, central pressure, and accumulated precipitation from a significant winter cyclone event.

Depending upon the type of modification being made to the modeling system and how extensive the changes are, an agency may also find it necessary to test the impact of the changes on

downstream modeling systems or other applications. For example, changes to global model physics that significantly affected tropical cyclone tracks would clearly call for testing of the regional hurricane/cyclone models. Modifications to the atmospheric boundary layer physics that significantly affected surface fluxes would require investigation of the impact of the new fluxes on aerosol, ocean, wave, and ice model behavior and on decision aids such as an automated ship-routing application. Similarly, the impact of model modifications on such post-processed parameters as visibility, icing, turbulence, etc., may also have to be investigated, along with the overall impacts of such changes to decision aids dependent on the accuracy of these forecasted parameters. The choices made for downstream testing are not standardized at this time, but are determined on a situation-specific basis. As a result, downstream testing of system modifications may vary within an agency as well as between agencies. Therefore, while we recommend that downstream impacts be communicated to other partners, we recommend limiting common metrics to those that assess NWP skill.

While deterministic forecasts do not fall under the purview of NUOPC, the skill of the deterministic model clearly affects the skill of the global ensemble run by each agency. Therefore, it is this Committee's recommendation that the partner agencies agree on an inclusive set of common metrics for assessing GNWP skill, as well as a common set of test cases that will be routinely executed by each agency as part of their validation and verification efforts. Furthermore, it is highly recommended that the metrics used include some measure of statistical significance of the result. This commonality will improve communication and interpretation of test results between agencies, as well as allow for a more straightforward comparison of the net effect of comparable changes on each modeling system if software subsets are exchanged and tested. This effort should be coordinated with the Joint Center for Satellite Data Assimilation, a NOAA/NASA/Navy/Air Force partnership, which is also interested in defining joint metrics for measuring improvement in GNWP performance resulting from the addition of new satellite data sources to the operational data assimilation systems of the NUOPC partners.

5.3.1.2 Ensemble Forecast Metrics.

The metrics that are suitable for evaluating performance of a deterministic model forecast may not be the same metrics that are desirable or suitable for evaluating performance of an ensemble forecast. At a minimum, there are additional factors that must be considered when evaluating the overall quality of a probabilistic forecast, including the techniques used to produce the ensemble itself. In addition, there are new parameters and, obviously, new types of information that can be extracted from an ensemble forecast, which requires additional metrics to be defined to capture the quality of the data available. One consideration is the ensemble mean, which is the average, or weighted average, of the individual ensemble members. While it may be, on average, more skillful than any individual ensemble member, the averaging process also filters out some of the detail that will be observed in the individual members. Another parameter of interest is the ensemble spread and how well it represents the true variability of the atmosphere, as measured by the observational data. The spread of the ensemble allows us to assess the reliability of the

predictions, as the distribution of possible outcomes represented by the different ensemble members provides information about the uncertainty or relative likelihood of various forecast scenarios. Ensemble predictions can be used to provide quantitative, as well as qualitative, estimates of probabilities.

Since ensemble predictions are relatively new compared to GNWP, the science of ensemble verification and probabilistic skill metrics is still the subject of much ongoing research. There is no general agreement on the best skill score and the choice may very well be dependent upon the particular application being considered. Probabilistic metrics such as the Briar Skill Score (BSS), Ranked Probability Skill Score (RPSS), Relative Measure of Predictability (RMOP), and Relative Operating Characteristic (ROC) curves are being applied to measure skill of ensemble forecasts (see Appendix 5). Additionally, Talagrand diagrams are routinely used to verify whether the observed probability distribution is well represented by the ensemble spread. For more information on these metrics and others, reference the following website for the World Weather Research Programme/Working Group on Numerical Experimentation (WWRP/WGNE) Joint Working Group on Verification:
http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html.

Since the basic tenet of NUOPC is a multi-model ensemble to improve the National prediction capability, it is critically important to develop a set of all-inclusive joint metrics that can be used to track the skill of the NUOPC ensemble and the statistical significance of the improvement. It is also important to compare that skill against the skill of the individual agency ensemble systems to assess how the National capability compares to each agency's ensemble forecast skill. For reference points, both NOAA and the Navy have already defined metrics that are applied to their individual global ensemble systems, and the Navy and the Air Force have defined metrics for their joint multi-model ensemble system, Joint Ensemble Forecast System (JEFS) (see Appendix 5). Another important example that can be looked at is NAEFS, http://www.emc.ncep.noaa.gov/gmb/ens/NAEFS.html, an existing and ongoing multi-model ensemble project involving the US, Canada and Mexico, which is also tackling the issue of defining probabilistic metrics for global multi-model ensembles. In addition, information about ensemble metrics used at other operational centers such as ECMWF, UKMO, and BOM are readily available, and there are many open literature references that can be consulted.

5.3.1.3 Application of Metrics.

The type of testing that needs to be performed for new modeling system development can be very time consuming, and it demands significant computer resources be made available to the developers and transition agents, both for computation and for data/model output storage and its subsequent analysis. Furthermore, such testing requires a fair amount of sophistication and experience on the part of the evaluation team, who need to be sufficiently knowledgeable of the overall system and its performance characteristics to identify likely sources of error and to diagnose the myriad problems that can arise in these scenarios. Once the developmental

evaluation team has determined that their new modifications are sufficiently robust and meet the standards for operations, according to an agreed upon set of metrics, the final stages of validation and verification can begin. This final evaluation is always performed on the operational machines using the operational database, running continuously in parallel to the operational system, and typically requires the assistance and expertise of the operational center's transition team.

Given the large number of performance metrics that are normally computed, it is common for a system modification to have mixed results (i.e., some metrics show an improvement in skill while others show a degradation in skill). Thus, the decision to move forward into operations is not always straightforward and may ultimately come down to a judgment call by the operational center or its oversight/decision board. The use of a scorecard (Appendix 4 provides one such example), which is a weighted sum of the collective metrics used to measure forecast skill, along with a definition of what is an acceptable vs. unacceptable score, is one way to make the decision-making process more objective. Ultimately, however, most operational centers operate under a "do no harm" rule. If the impact of the modification is judged to be non-negative (i.e., the collection of metrics is sufficiently positive that the expectation is that product users will not discern any noticeable degradation of forecast skill), and the impending upgrade provides a technology or new scientific capability that lays the groundwork for additional improvements to come, the change will go forward into operations. In such cases, the longer range payoff is the driving force behind the decision-making, not the immediate gain to be made.

This current culture of using a combination of objective metrics and more subjective, goal-oriented decision-making will be a challenge for the standing TTP Committee. If the current decision-making processes at the individual agencies were to be replaced by a purely objective process, there would be an even greater burden to define a metrics process that will not unduly hamper or inhibit the individual agency's needs to meet its mission goals. However, many difficulties can be envisioned from a decision-making process that imposes strict objectivity and thus joint concurrence for transition decisions. Those difficulties, coupled with the NUOPC goal of accelerating technology transition leads us to recommend the following: 1) a common set of metrics for use by all partners should be agreed upon, and 2) the transition decisions based upon those metrics should remain independent and under the sole authority of the agency responsible for the planned software modification. That is, each agency will measure the impact of model modifications with a standard suite of tools and against a standard measurement of skill, but each agency will maintain the authority to implement changes in its solely-owned component of the NUOPC capability with or without the concurrence of partner agencies. The one decision that would require consensus would be a decision to alter the way the joint ensemble was formed from the individual ensemble members. Such a decision would necessarily require approval by the ESG or other designated representative of the ESG. Finally, the standing TTP Committee will coordinate with the standing UEO Committee Panels and the operations community to ensure the NUOPC Metrics program focuses on metrics of interest to operations.

5.3.1.4 Selection and Maintenance of Metrics.

Since the definition and selection of specific metrics is a highly technical area, we recommend that a NUOPC Scientific Metrics Committee (SMC) be formed and its efforts coordinated by the standing TTP Committee to research, formulate, and recommend an appropriate set of joint metrics for use in the evaluation and decision-making process regarding modifications to NUOPC and partner agency global models and model ensembles. The SMC need not be a separate committee from the standing TTP Committee. In fact, it is recommended that the standing TTP Committee perform the functions of the SMC either by creating a sub-committee or by performing SMC functions in their entirety as a standard part of the standing TTP Committee charter. The Scientific Metrics Committee should be comprised of one subject matter expert from each operational center— Fleet Numerical Meteorology and Oceanography Center (FNMOC), NCEP, and AFWA-- and one subject matter expert from each global ensemble transition agent – NRL Code 7500 and EMC, nominated by their respective civilian technical directors, or equivalent. Two additional subject matter experts may be nominated by the other five members from the research community or operational community at large. All committee members must be approved by the standing TTP Committee, which will also select the chair of the committee. The Scientific Metrics Committee will be tasked with reviewing the state-of-the-art in GNWP and probabilistic metrics applications, consulting with downstream customers and end users of ensemble products, and evaluating and consolidating disparate agency needs, to define a single set of joint metrics for evaluating GNWP skill and another set for evaluating ensemble prediction skill and *impact*.

The latter is not to be overlooked. It is critically important that this committee take into account how the NUOPC ensemble products will be used by the various agency customers when defining the metrics that will be utilized. Therefore, in addition to measuring the overall skill of the ensemble, we suggest that the committee also look at the more difficult issue of defining metrics that measure impact on downstream applications as well as value to the customer. For example, many products may be delivered to customers using Web-based technology. In that case, tracking the number of "hits" received by each type of product might be one way to define how useful customers found a particular product, especially if the numbers changed significantly one way or the other over time. However, even if a product is used by only a very limited number of customers, it may be providing an essential piece of information. Quantifying the ***value*** of individual products, and their improvements, to customers will be more challenging. Some of the options that might be considered include annual customer surveys, collections of case studies noting impacts of particular products on specific forecasts, etc. Any ideas that encourage--or mandate--obtaining regular feedback from customers would be useful.

Measuring impact on downstream applications is another extremely challenging area. Each agency will obviously have different applications that will be dependent on the NUOPC multi-model ensemble. While we do not expect this committee to identify those products, if the agencies identify the application as a routine part of their evaluation package, it would be the

role of this committee to define the metrics for that downstream application. As for "the how" the different agencies' downstream applications would be considered in the overall decision-making process for changes to the operational NUOPC ensemble, it is viewed that this is outside the purview of the standing TTP Committee and part of the standing UEO Committee Panels' responsibilities. However, caution should be taken in defining a process that would be so time-consuming to execute that it would effectively negate the primary NUOPC goal of accelerating technology transfer. The amount of downstream application and end user testing currently performed by the agencies prior to implementing new technology can provide some guidelines as to the relative importance placed on these matters. That said, downstream impact is expected to become more important than it is now as the use of ensemble products in downstream applications increases. However, emphasis should be placed on communication and collaboration between partners rather than blind joint decision making.

It is recommended that the Scientific Metrics Committee meet annually to review the current metrics and make recommendations for changes based on feedback over the past year, as well as changes in the state-of-the-science. In addition, this committee should play the primary role in recommending topical areas for additional metrics-related research, based on observed gaps in knowledge and shortfalls in our ability to quantify skill and impact in ways that best meet the needs of the NUOPC members. All recommendations of the Scientific Metrics Committee, including initial metrics and changes to those metrics, would be presented, in writing, to the ESG or its designated representative for approval. Suggested topical areas for future research would be communicated in writing to the various government funding agencies for national meteorological research and to NUOPC leadership.

5.3.2 System Metrics.

Before any system upgrade can be seriously considered for operations, its impact on overall operational system execution must be determined. The operational product suite at each operational center is carefully scheduled, based on a number of factors. These factors include the optimum window of data availability, the length of time it takes to run a system, the number of systems that will be executing simultaneously, the computer resources that are available, and perhaps most importantly--the time the end products have to be in the hands of the customer to serve their useful purpose. In addition, each operational center typically has several operational computers, which can be updated or changed every few years, which places additional demands on operational software. Ideally, the software developer is aware of these additional demands and is taking into account the fact their software not only has to provide improved skill or capability but also must meet operational requirements for robustness, portability, and speed of execution. Furthermore, the developed software must fit into the available memory and bandwidth allocated for the system.

So, in addition to scientific metrics for measuring skill and metrics for assessing value to the customer, a set of metrics for determining if the specific software under consideration will meet

other operational requirements is needed. First, the software must be portable--able to be successfully compiled and executed on different hardware with different operating systems--and reproducible, achieving the same results repeatedly when given the same data. The choice of specific operating systems and computers will depend upon where the software is intended to be used. If it is only targeted for one of the NUOPC partner agencies, then the transition team for that agency's operational center can determine what constitutes a sufficient demonstration of portability and reproducibility. If the software is intended for use by multiple partners, then broader testing will be required to satisfy the transition teams at each of the potential operational centers. Second, the software must be robust. That requirement is typically demonstrated by executing the upgrade as part of the overall system, in a continuous mode with real data, for a period of several weeks. Finally, the software must execute within the system parameters set aside for operational computation, including constraints placed on wall time, memory, and bandwidth.

In each case, the developer must work with the agency's operational center and transition team to determine if the new software can meet these system requirements. Since the system requirements will change with time, there is no reason to make any specific reference as to what they are. Each agency already has some type of transition team that can define what these metrics will be at the time they are required, and a process for evaluating them. It is recommended that as long as the software is targeted for operations at only one operational center, then that agency's decision-making process regarding these system metrics remains an independent decision. If the software is intended to be shared by multiple agencies, then each agency's transition team may choose to independently work with the developer to evaluate the software. However, one agency's rejection of the software for reasons of failure to meet system metrics would not necessarily prevent another agency from using the software if it met their purposes.

5.3.3 Strategic Metrics.

The most difficult, but perhaps most important, set of metrics to define will be the metrics that help determine whether or not NUOPC is indeed meeting its strategic goals. While it may be difficult to define quantitative metrics to measure the overall success of our overall goals, we recommend that such metrics at least be part of the evaluation process as we track the progress of this joint agency effort. Since NUOPC goals are identified in terms of three factors, the discussion in this section is framed around those three factors. Also, in each section, cost vs. benefit is evaluated, for ultimately, every agency is interested in identifying cost savings, if possible, or identifying the increased costs associated with implementing and sustaining this improved capability. Additionally, since the interim TTP Committee's goal is to propose a process to identify the NUOPC ensemble metrics rather than come up with the metrics themselves, only a few ideas are suggested as to what some of the metrics might be. As for the process, it is recommended that the NUOPC Program Manager task the standing TTP Committee

to prepare a more in-depth analysis and recommendation of strategic metrics for approval by the ESG or its designated representative.

5.3.3.1 Metrics to Determine Accelerated Technology Transfer.

To determine if some process has been accelerated, it is necessary to somehow compare recent accomplishments with those of the past. There are a couple of sets of records that are fairly well documented, and some we may need to start documenting. There are several common GNWP skill scores that have been tracked for years, such as 500 millibar anomaly correlation and tropical cyclone track errors. If technology transfer is accelerated, the trend line of changes in these scores may be accelerated compared to the historical trends; however, since NUOPC's focus is on global ensembles, we need to quickly start tracking agreed-upon metrics for the agency's individual ensembles, so we will have a baseline to measure future ensemble skill against, as this is one way of seeing indirect evidence of accelerated technology transfer. Another metric that could be tracked for this purpose would be the time elapsed from "proof of concept" to the operational implementation of the new technology, marking the important decision points along the way, and watching for evidence that the speed to capability of new technology is improving. It might be possible to piece together historical information about time to transition from historical project files and other documentation. The other set of records that is fairly complete is the record of upgrades made to the global modeling systems and the global ensembles. That set of records can be watched for indications that upgrades into the systems are more frequent after full implementation of NUOPC.

The other aspect of technology transfer that can be measured relates to the engagement of the research community at large. Records could be kept of how many different researchers or research agencies contribute new software or system upgrades that actually transition to operations. Similarly, records could be kept of how many researchers outside of the operational agencies are working with the operational systems, particularly the joint NUOPC ensemble system. Ideally, if NUOPC is successful, we would see those numbers increasing each year, although we have to be careful not to interpret use of a system as synonymous with contributing to improving a system's skill.

Finally, to look at a cost vs. benefit analysis for this area, one area of primary interest would be to investigate and track how the investment in basic and applied research in global modeling and global ensemble prediction, in particular, changes from the pre- to post-NUOPC years (normalized, of course, to some base year's dollars). And to that, we would have to add any cost of procuring resources and hiring staff to assist outside researchers in participating in the development, testing, and implementation of their software, as well as using the operational system in their own research.

5.3.3.2 Metrics to Demonstrate Reduced Duplication of Effort.

To demonstrate reduced duplication of effort, we would look for evidence that the operational agencies and their transition teams: 1) are no longer doing something they once did, 2) are using something from a partner agency rather than developing it themselves, 3) have increased collaborations and are thereby sharing "brainpower" more effectively, and/or 4) are sharing hard resources, among other possibilities. Obvious examples of being able to use one another's software rather than developing it from scratch removes a duplication of effort and *may* save money, as long as the effort to re-tool that software and get it running effectively on the other agency's computer systems doesn't take longer than writing the software from scratch. The more the agencies standardize their computer hardware, operating systems, compilers, libraries, data processing, software standards, etc., the more straightforward sharing of software would become, but until that time, it is not always a given that it is more efficient to borrow software than build it, so this is one of those areas where a cost analysis can be beneficial.

It is, however, always beneficial to share ideas and lessons learned with one another. Even if one agency can't utilize another agency's software directly, they can certainly save time when building their own software if they liberally draw upon the knowledge gained by the other agency as they develop, debug, and test their own software. Many of the problems encountered are likely to be the same, and knowing how someone else solved the problem can be an invaluable time saver. Forming work groups that would teleconference or e-meet on a regular basis to discuss topics of common interest would be beneficial. A related area where much time and resources could be saved is when agencies *are* using common software, they could relax their own requirements for testing and accept the fact that the other agency has already tested the software sufficiently to determine that it is operationally acceptable. This agreement could greatly shorten the time needed to implement software obtained from another agency to conducting only those tests that would be needed locally because of system differences or other system dependencies.

5.3.3.3 Metrics to Assess a Superior National GNWP Capability.

If we limit the definition of superior to mean "more skillful," then we can use the metrics that would be defined by the Scientific Metrics Committee as the primary tool for assessing whether or not NUOPC is achieving this goal. And while this type of assessment would be an important part of determining success on this third goal, defining a superior National capability is not that simple. At the very least, it is recommended that a cost component should be added so that there is a comparison between the percent improvement in skill for dollars invested both pre- and post-NUOPC. Finally, the ultimate test is whether or not NUOPC leadership is satisfied that NUOPC has enabled them to better fulfill their agencies' missions.

5.4 Potential Risks and Roadblocks.

The bottom line is that NUOPC will not be a success unless each of the partner agencies ultimately feels that it is better off post-NUOPC than it was pre-NUOPC. If the decision-making

processes defined by NUOPC are too cumbersome, or the costs for completely implementing and sustaining the effort are too high, or the management structure takes too much time away from too many valuable people – and the payoff is not large enough to justify those costs – the effort will fail. A point in GNWP has been reached where the gains in skill are small incremental improvements that over time average out to be about the equivalent of a gain in skill of one forecast day every 10 years. The great new discoveries that lead to large leaps ahead are very few to non-existent. And while there are differences in models, and some models are more skillful than others overall, when significance testing is applied to the various skill scores, the differences are not as significant as one might think.

Now, we are entering into a new era of forecasting, one based on probabilistic prediction. Will the trends for ensemble forecast systems be different? Can we find ways to achieve those large leaps forward in skill by the way we formulate the ensemble members…since we can't expect to do it through improvement to the global models themselves? The additional challenge for NUOPC success will be to show that *multi-model* global ensembles, in particular, will be significantly better than single-model global ensembles for the purpose of meeting the various agencies' customers' needs. From a scientific viewpoint and from a practical viewpoint, that is a challenge. Expectations of senior management should be realistic on this point.

It will be critically important to carefully define the overall management and decision-making processes for NUOPC. An over-emphasis on joint decision making or joint testing for impact on downstream applications can quickly paralyze the upgrade process and immediately defeat the goal of accelerating technology transfer. Too many layers of management, with too many committees and working groups, result in everyone spending their time attending meetings and writing documents, rather than improving the scientific products. We need to move quickly through the former so as to focus on the latter.

For metrics processes, the single largest potential roadblock is likely to be agreeing on common metrics that will meet the different requirements and applications of the various agencies. Using the concept of "weighted" scorecards should assist in this process, where larger weights could be assigned to those parameters of interest to all parties, as one example. Another consideration for ensemble metrics might be to include the possibility of a suite of scorecards used to "tune" ensemble members to manage the spread of the final product, with each agency focusing on a subset of scorecards that best meet its own requirements.

5.5 Costs.

We don't expect the costs for formulating the metrics to be exceptionally high, as these costs will be mostly limited to labor and temporary duty (TDY) costs for the committee members involved. We estimate that the Scientific Metrics Committee will need funding during the first year (as early as 2010/2011) for four weeks salary and two weeks travel/temporary duty costs to research and initially formulate the metrics. With seven committee members on the Scientific Metrics

Committee (assuming two will be at the meeting site), we would need roughly $120K for salaries and $25K for travel for that committee. Additionally, we would ask for $20K to be set aside for any visits members might want to make to talk to customers regarding their use of probabilistic products, for a onetime cost of $165K for the committee's total effort. This recommendation is an expansion of existing functions at the individual agencies. After the development of the initial products and processes, it is recommended that the NUOPC metrics program be folded into the routine functions of the production centers and RTFs.

For the interim TTP Committee, we estimate we would need two week's salary and one week travel/TDY costs to work on strategic metrics, plus an additional week's salary for the Metric sub-committee to finalize the documentation. We believe costs for this group's efforts have already been funded. There are, of course, the more indirect costs related to lost productivity on other work efforts, in that time spent on NUOPC committee work is time not spent on other projects that consumed peoples' time pre-NUOPC, but we have not attempted to place a dollar value on those costs.

5.6 Recommendations.

We make the following recommendations regarding metrics associated with NUOPC's technology transition processes:

- Joint metrics are needed in addition to individual agency metrics.
    - The partner agencies should agree on an inclusive set of common metrics for assessing GNWP skill; the metrics used should include some measure of statistical significance of the result.
    - The partner agencies should develop a set of all-inclusive joint ensemble metrics that can be used to track the skill of the individual ensembles, the NUOPC ensemble, and the statistical significance of any improvements.
    - The partner agencies should agree on a common set of test cases that will be routinely executed by each agency as part of their validation and verification efforts for GNWP and global ensemble prediction.
    - Transition decisions based on these metrics should remain independent and under the sole authority of the agency responsible for the planned software modification.
- NUOPC needs to address Performance Metrics, System Metrics, and Strategic Metrics.
    - Performance Metrics should include metrics for both deterministic and probabilistic global forecast systems.

- A Scientific Metrics Committee should be formed to research, formulate, and recommend an appropriate set of joint metrics for use in evaluating and decision-making regarding changes and modifications to the NUOPC and agency global models and model ensembles.

- Performance metrics would include measures of forecast skill, statistical significance of the changes, assessments of ensemble behavior, impacts on downstream applications and derivative products, and value to the customer/end user.

- All recommendations of the Scientific Metrics Committee, including initial metrics and changes to those metrics, would be presented, in writing, to the ESG or its designated representative for approval.

- Suggested topical areas for future metrics research would be communicated in writing to the various government funding agencies for national meteorological research and to NUOPC leadership.

- System Metrics should include metrics for portability, reproducibility, wall time, memory and bandwidth requirements.

  - The developer would work with the targeted agency's operational center and transition team to determine if the new software can meet their system requirements.

  - One agency's rejection of the software for reasons of failure to meet system metrics would not necessarily prevent another agency from using the software if it met their purposes.

- Strategic Metrics should include metrics for determining if the NUOPC goals of accelerated technology transfer, reduced duplication of effort, and superior National global prediction capability are being met.

  - The NUOPC Program Manager should task the standing TTP Committee to prepare a more in-depth analysis and recommendation of strategic metrics for approval by the ESG or its designated representative.

  - Metrics should include cost vs. benefit analyses where possible.

The Scientific Metrics Committee (SMC) will be nominated and approved by the standing TTP Committee in FY10. The SMC, in turn, will have its recommendations ready to submit to the ESG six months after appointment.

Having well-defined metrics for NUOPC will provide a means for making informed decisions not only about the value of a specific piece of model software or new technology for creating better ensembles, but also for evaluating the overall value of the program to the individual agencies. Defining joint metrics and test cases for common use by each agency will improve collaboration and communication, and make test results more meaningful and useful to one another, saving each time and resources. In addition, these metrics will assist the broader research community in participating in the technology transfer process by defining in advance the standards that must be met.

6. SUMMARY.

In order to accelerate the transition to new technology, eliminate unnecessary duplication, and achieve a superior National global prediction capability, a joint process for technology transition must be formulated and applied by the NUOPC Tri-Agency. A standing TTP Committee is needed in order to manage key technology transfer functions. These functions include the establishment and management of outreach programs, Research Transition Facilities, and a strong metrics program.

The standing TTP Committee will encourage and stimulate research community involvement in NUOPC to facilitate the rapid transition of new improvements or technologies into operations. Outreach will require providing stimuli to the research community in the form of visitor programs, funding, and access to computing resources. A Fellowship Program is recommended to elicit researcher interest and assistance in tackling the prioritized Tri-Agency research requirements list. A robust program will cost approximately $2M per year for 10 FTEs, beginning in 2013. However, this dollar figure and the number of FTEs are scalable based on the amount of research the Tri-Agency wants performed.

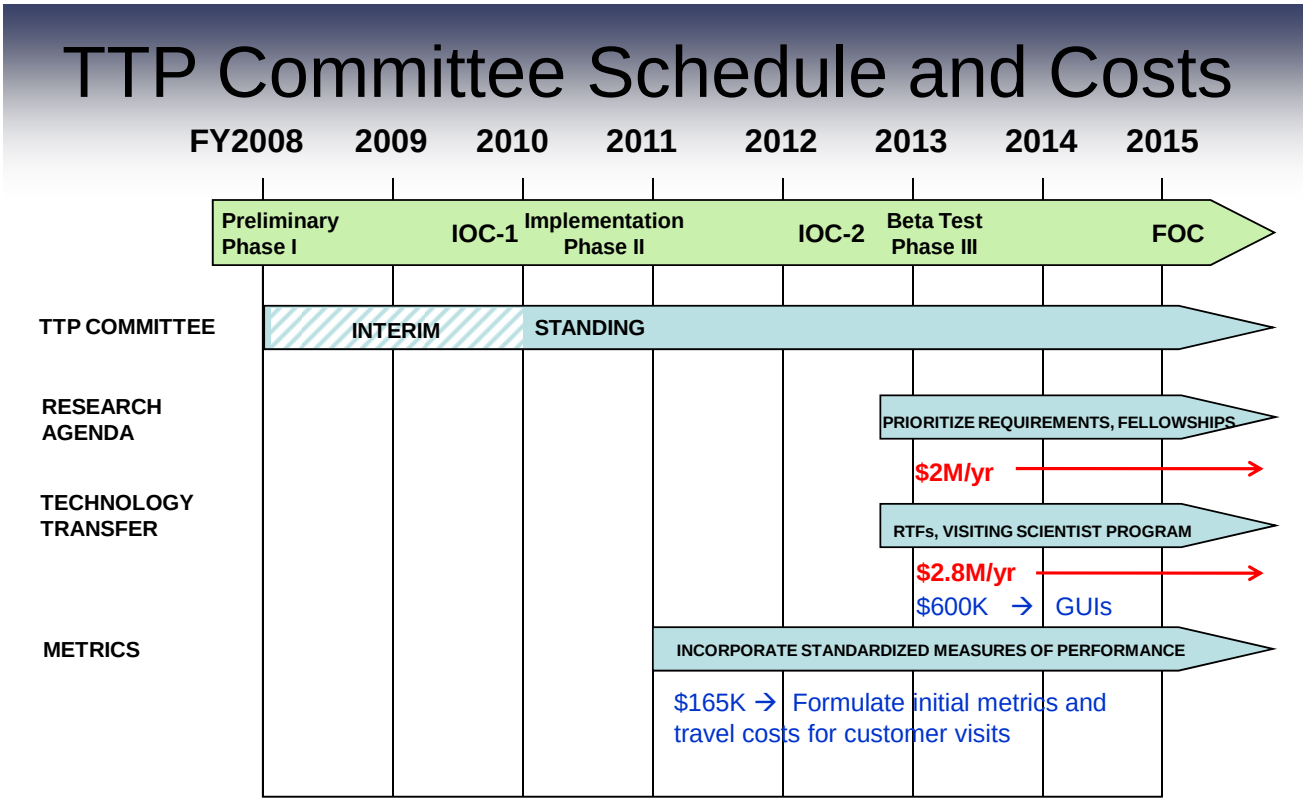
The establishment of Research Transition Facilities (RTFs) at each Tri-Agency production center (including AFWA) will enhance collaboration and communication with the research community, and assist Tri-Agency production centers in improving their operational capabilities. A Visiting Scientist Program is required to perform on-site RTF work. A robust Visiting Scientist Program will cost approximately $2M per year for 10 FTEs, beginning in 2013. This dollar figure is scalable based on the amount of operational improvement desired. In addition, the RTFs require a common CM system to synchronize efforts among the Tri-Agency. Starting in 2013, four FTEs are needed at the RTFs for CM requirements for a cost of approximately $800K per year. This recommendation is an expansion of existing functions at the individual agencies. Lastly, a onetime cost of $600K, which is also an expansion of existing functions, is needed to write initial scripts and standardize GUIs at the RTFs. Maintenance of the scripts and GUIs will be

incorporated into the routine functions of the production centers and RTFs after this initial stand-up effort.

The proposed NUOPC metrics program will incorporate standardized measures of technical capability. This program requires an initial onetime cost of approximately $165K to formulate initial metrics and cover travel costs for customer visits. This recommendation is an expansion of existing functions at the individual agencies. After the development of the initial products and processes, the NUOPC metrics program will be folded into the routine functions of the production centers and RTFs.

The total cost of the proposed NUOPC TTP functions is approximately $4.8M per year, beginning in 2013. However, this cost is mostly scalable based upon the amount of research and operational improvements/collaboration desired by the ESG.

Figure 2 below summarizes the TTP Committee's schedule and costs.



Total cost: $4.8M/yr, beginning in FY13 (*mostly scalable, based on the amount of research and operational improvements desired*)

Figure 2. TTP Committee Schedule and Costs.

APPENDIX 1 – List of Acronyms and Definitions.

AC	Anomaly Correlation
AFWA	Air Force Weather Agency
AGU	American Geophysical Union
AMS	American Meteorological Society
APL	Applied Physics Laboratory
BAA	Broad Agency Announcement
BOM	Bureau of Meteorology (Australia)
BSS	Briar Skill Score
C&A	Certification and Accreditation
CM	Configuration Management
C-MAN	Coastal-Marine Automated Network
DA	Data Assimilation
DemVal	Demonstration and Validation
DoD	Department of Defense
DTC	Distributed (or Developmental) Testbed Center
ECMWF	European Centre for Medium-range Weather Forecasts
EMC	Environmental Modeling Center (NCEP)
ESG	Executive Steering Group
ESMF	Earth System Modeling Framework
FNMOC	Fleet Numerical Meteorology and Oceanography Center
FTE	Full-Time Equivalent
GFS	Global Forecast System
GNWP	Global Numerical Weather Prediction
GUI	Graphical User Interface
IA	Information Assurance
I/O	Input/Output
IT	Information Technology
ITAR	International Traffic in Arms Regulations
JEFS	Joint Ensemble Forecast System
NAEFS	North American Ensemble Forecast System
NASA	National Aeronautics and Space Administration
NCEP	National Centers for Environmental Prediction
NCO	NCEP Central Operations
NOAA	National Oceanic and Atmospheric Administration
NOGAPS	Navy Operational Global Atmospheric Prediction System
NRC	National Research Council
NRL	Naval Research Laboratory
NSF	National Science Foundation
NUOPC	National Unified Operational Prediction Capability
NWP	Numerical Weather Prediction
Ops	Operations

PDF	Probability Density Function
PM	Program Manager
RAOB	Rawinsonde Observation
RDT&E	Research, Development, Test & Evaluation
RFI	Request For Information
RMOP	Relative Measure Of Predictability
RMS	Root-Mean-Square
ROC	Relative Operating Characteristic
RPSS	Ranked Probability Skill Score
RTF	Research Transition Facility
R&D	Research and Development
SMC	Scientific Metrics Committee
SSMI	Special Sensor Microwave Imager
S&T	Science and Technology
TDY	Temporary Duty
THORPEX	The Observing System Research and Predictability Experiment
TIGGE	THORPEX Interactive Grand Global Ensemble
TTP	Technology Transition Processes
T&E	Test and Evaluation
UARC	University Affiliated Research Center
UAV	Unmanned Aerial Vehicle
UCAR	University Corporation for Atmospheric Research
UEO	Unified Ensemble Operations
UKMO	United Kingdom Meteorological Office
USAF	United States Air Force
USN	Unites States Navy
VSP	Visiting Scientist Program
VTC	Virtual Testbed Center
V&V	Validation and Verification
WRF	Weather Research and Forecasting
WWRP/WGNE	World Weather Research Programme/Working Group on Numerical Experimentation

APPENDIX 2 – Glossary.

Configuration Management (CM)	Establishing and maintaining the consistency of a system's or product's performance and its functional and physical attributes with its requirements, design, and operational information throughout its lifecycle.
Deterministic forecasts	A single forecast; ignores forecast uncertainty.
Distributed (or Developmental) Testbed Center	A centralized, consolidated technology transfer function located external to the operational centers.
Earth System Modeling Framework (ESMF)	The software for building and coupling multi-component climate, numerical weather prediction, data assimilation, and other Earth science software applications.
Ensemble forecasts	Ensemble forecasts are designed to capture the probabilities for weather events and the range of uncertainty inherent in each forecast situation. Ensemble forecast systems provide information on uncertainty and the most likely forecast outcome(s).
Ensemble members	Different ensemble forecasts (or members) are generated by changing the initial conditions, the boundary conditions, or the model characteristics (e.g., dynamical formulation, numerical method, physical parameterization, and horizontal and vertical resolution).
Fellowship Program	A program that partners with academia to work together in the pursuit of knowledge or improved capabilities.
Information Assurance (IA)	Measures that protect and defend information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities.
International Traffic in Arms Regulations (ITAR)	Restrictions the United States (US) places on the release of critical technologies to non-US persons.
Operational/production centers	AFWA, FNMOC, and NCEP
Outreach	An effort to connect one group's ideas or practices to the efforts of other groups or organizations.
National Unified Operational Prediction Capability (NUOPC)	An agreement to coordinate the activities of the Tri-Agency partners in order to accelerate the transition of new technology, eliminate unnecessary duplication, and achieve a superior National global prediction capability.
Research agenda	A prioritized, coordinated research requirements list.
Research Transition Facility	A proposed technology transfer function collocated with each Tri-Agency operational center to facilitate the rapid transition of new technologies into operations.

Visiting Scientist Program	A program that provides opportunities for visiting scientists in the academic community and other government laboratories to interact with colleagues at the operational centers in an attempt to improve or enhance existing capabilities.

APPENDIX 3 – DTC Proposal.

In order for the NUOPC program to be ultimately successful it needs to entrain the overall community, including university researchers, and researchers at the mission agencies. Experience gained during the development of the Weather and Research Forecasting (WRF) program is useful to reference. The WRF program was a grassroots level effort to try to link the research and operational mesoscale communities together in a way that is better than what has existed in the past in order to more efficiently transfer new mesoscale NWP techniques and systems from research into operations. The program includes a well managed code system that contains a full end-to-end NWP system with many options including two substantially different cores, various physics options, pre- and post-processing systems, data assimilation, and verification systems. In addition, there is a strong user support service provided, which includes documentation development, user support desks and tutorials. This code system is used worldwide with over 6000 registered users and portions of the WRF system are used in two operational centers in the US and in a number of foreign countries, particularly in Asia. A major part of the WRF program is a Developmental Testbed Center (DTC) that provides much of the user support and code system maintenance mentioned above. It also provides a testing and evaluation process that has been useful both in operations and in research. While there are many successes with the WRF program, it can be regarded as enjoying only moderate overall success to date since it has only garnered modest buy-in within the agencies that would be necessary for full success. Regardless, it has been successful in linking the research and operational communities and lessons learned from WRF can be used in creating the NUOPC outreach program.

1. Purpose of the Outreach Program. The motivation for the NUOPC program reaching out to the broad community beyond the operational community is the same as it is for WRF program. Put simply, the problem of making improvements in operational NWP forecasts systems at a rate that the customers appear to be demanding is a very large one, one that probably can't be accomplished by the personnel at the operational centers alone. If a way can be found to leverage the extensive expertise in the research community and focus it on improving the operational models, then the rate of improvement of the operational models can be accelerated. The processes outlined below are an attempt to achieve such a leveraging.

2. Work Performed by the Committee. The bulk of this section will need to be written after the subcommittee has completed its work. At this time the plan is for the Committee Chair to prepare a draft of the report then circulate it to the committee for comments. Those comments will be incorporated into the draft by the Chair and re-circulated to the committee for further comment and modifications before being submitted to committee members for final comment and modifications. Many of the initial ideas presented in this first draft are from experience the Chair has had with the WRF program.

3. Reaching Out to the Research Community. The research community has a reputation of only wanting to work on problems they find interesting ("leave me alone, give me plenty of money and I will do wonderful things") and that they generally don't want to work with operations. Indeed this has led to the concept of a "valley of death," the void between research and operations. While this may be true for a few researchers most really are anxious to see the results of their work be used practically. In numerical weather prediction research this translates into wanting to see their efforts improve the operational models at some point. So in principle it should not be difficult to establish collaboration between the research community and operational centers.

This collaboration could include focusing research on topics of particular interest to operations, testing and evaluation of the NWP systems in operations, and pre-implementation testing of proposed changes to the operational system. It could include evaluation of next generation NWP components for possible elevation to operations. In order to attract collaborators from the research community there needs to be incentives in place. These incentives could include:

Financial resources

Grants program for targeted research

Visitor program (with the Virtual Testbed Center (VTC)/DTC or one of the operational centers)

Facilities

Computer resources

Access to data archives

Access to NUOPC codes including the operational codes

A facility where researchers can participate in testing and evaluation (i.e., DTC)

Membership in the decision making process for NUOPC

Advisory committees

With sufficient financial incentive, the grants and visitor programs probably would not be necessary. However, those resources in the form of grants and visitor positions are not likely to be more than modest so incentives in the form of facilities available to work on NUOPC-related problems and a form of ownership in the program will be necessary. In the sections that follow, the focus is primarily on facilities and financial components.

Membership of the research community in the management of NUOPC is left to other committees. However, there should be members from the research community on committees that influence resource sponsors and the choice of NUOPC development efforts that are selected

for priority emphasis. By participating in NUOPC management the research community will get a sense of the transition possibilities and be encouraged to participate in that process. Also, as we noted above, most researchers would like to see the results of their research used for practical applications and a sense that they are truly influencing the improvement of operational forecasting will be an incentive to participate in the transition process.

4. DTC/VTC. A DTC is a Developmental Testbed Center patterned after the WRF DTC. A VTC is a Virtual Testbed Center. A VTC is intended to mean that the center is not at one particular location; rather, it is distributed among a number of locations (for example, at or near the three operational centers).

Unfortunately, VTC conjures up images of holograms and video games which certainly is not what is meant. The term "Developmental" when used in DTC (as with the WRF DTC) conjures up images of a place where NWP codes are developed. For the WRF DTC, this is definitely not the case. In fact, the WRF DTC carefully avoids any sense that it develops NWP codes since one of its main objectives is testing and evaluation that could pit competing codes developed by different developers against each other. By avoiding being a code developer, the WRF DTC can be viewed as an unbiased assessor of any comparison. The WRF DTC is also a distributed facility with components currently at NCAR and NOAA/Earth System Research Laboratory (ESRL). Other components have been discussed but for the moment these are the only two that are operating effectively.

It has been proposed in the past that the "D" in the WRF DTC refers to "Distributed," which better describes the role of the WRF DTC. Therefore the "D" in the NUOPC DTC will stand for "Distributed." Lastly, when referring to the existing DTC in the WRF program, it will be referred to the WRF DTC, and when referring to the NUOPC implementation, it will be referred to as the NUOPC DTC.

4.1. Concept. The Distributed Testbed Center (DTC) of NUOPC (referred to below as either DTC or NUOPC DTC) would contain a number of components from the various NUOPC partners. For example, there might be a component closely associated with each of the operational centers. In fact, each of the centers currently has a facility in which a DTC component can be constructed: Naval Research Laboratory (NRL) for the U.S. Navy, National Centers for Environmental Prediction (NCEP)/Environmental Modeling Center (EMC) for NOAA, and the National Center for Atmospheric Research (NCAR) for the Air Force Weather Agency (AFWA). However, in setting up this distributed facility, care should be taken to insure there is a clear separation between the development parts of these organizations and the testbed center for the reason noted above. Currently, with the WRF DTC, the NCAR component is in a separate part of the organization (in the Research Applications Laboratory (RAL)) from the development part (Earth & Sun Systems Laboratory (ESSL)/Mesoscale and Microscale Division (MMM)). In ESRL, the DTC is located within the Global Systems Division (GSD), but is a separate entity from the parts of GSD that perform model development. In the NUOPC DTC, a

similar separation will exist, for example, EMC and NRL. Presumably at NCAR, a separation will be similar to the current one at the WRF DTC. By a separation, it is meant that there is an entity within EMC or NRL that is clearly identified as part of the NUOPC DTC and that it and the people that work there are not part of model development. Each of these organizations will have to assure the community that when testing occurs within their component of the DTC that there is no bias toward codes that originate within their host group and that they are neutral to the outcome of the tests. There has been a lot of experience in the WRF DTC in how to maintain this neutral reputation and this can be shared with NUOPC.

4.2. Maintenance of a Code Repository. Each operational center already maintains a repository of their operational codes, but these are usually behind firewalls for which there is limited access by users outside the firewall. In addition, the adding of new experimental codes to the operational repositories is extremely limited in order to maintain the security of the operational codes. This situation will not be adequate for close collaboration between the research and operational communities where any potential user should have easy access to the NUOPC codes (as is the case with WRF). Thus, it is advisable to maintain a research-level code outside the operational repositories that would contain as a minimum the operational codes, but could also contain any number of experimental codes. A responsibility of a NUOPC DTC would be to maintain this repository, allowing check-ins that would then be carefully tested to assure the integrity of the repository after the check-ins. The DTC could also be responsible for assuring that the repository it maintains is kept reasonably in sync with the current operational repositories. Note: various components of this research repository could be held at several of the distributed centers.

4.3. User Support. Once a repository is made available to the community there needs to be support provided to the users. This support would be in the form of providing the codes to the users (perhaps through Web-based downloads), supporting documentation to describe the codes and how to use them, help desks to answer questions, and tutorials to introduce the NUOPC codes to new users. In addition, the NUOPC DTC will need to maintain test cases to exercise systems and for testing and evaluation, metrics to evaluate performance relative to other system configurations, and tools to store and access model output for future reference. All of these are functions of the WRF DTC and experience from it can be used in designing these services for NUOPC. These activities in particular tend not to be of great interest to either model developers in the research community nor to operational centers and are a reason why a separate DTC system should be considered for NUOPC.

Note that user support services are services that operational centers generally do not wish to provide since that is not their primary, or even secondary mission. Their mission is to provide operational forecasts and so they tend to see user support as a diversion of resources away from operations.

4.4. Visitor and Grants Program. A major service of the WRF DTC is to provide opportunities for members of the research community to visit with components of the WRF DTC or the

operational centers. This is in the form of a visitor program where the DTC sets aside on the order of $100K-$200K per year to provide visitors with one month salary plus travel and per diem for a month visit with the DTC or one of the operational centers. The DTC releases an announcement of opportunity annually seeking proposals for visitor projects. These proposals are evaluated by its advisory board which selects the proposals to be funded. A similar program can be set up for the NUOPC DTC.

In addition, if NUOPC were to set up a grants program the DTC could manage a competition process for this funding including releasing an announcement of opportunity and using its advisory board for making a priority list of submitted proposals. Individual funding agencies would then make a final determination as to the distribution of funds that they have provided to the grants program. Note that the visitor program funds could also be handled in a similar way.

A visitor program is relatively low cost and so represents a huge amount of leveraging for collaboration with the community in technology transfer and also for meeting the research requirements that are set out by the NUOPC program. This may argue for a visitor program over a grants program since it would likely be relatively straight forward to incorporate visitors into a program of rapid transition to operations where the visitors work directly with an operational center in making the transition. These visitors could even be jointly funded between the major NUOPC finding agencies and an operational center with costs still running at about $20K/visitor/year.

This however, does not imply that a grants program would not have value as well. The costs per PI are higher than for a visitor but the grants program can be easily directed towards fulfilling the longer term R&D, identified by the operational centers, that are necessary to fully meet the longer term NUOPC goals.

4.5. Management. Each of the distributed components of the NUOPC Distributed Testbed Center would have a local lead who would report up through the normal management structure of the host organization (e.g., NCAR, NRL, and NCEP). Each of these local leaders could be considered a Deputy Director of the NUOPC DTC and collectively they would form the management committee of the DTC. One of these deputies would be named the overall director and each of the deputies would report to the Director for overall leadership of the DTC. Duties of the Management committee working with the Director would be to set an annual program for the overall DTC, deciding what parts would be accomplished by each of the distributed components. These plans in turn would be approved by each of the funding agencies.

The Director would report to an executive committee that has overall responsibility of the management of the DTC. This committee would be made up members of the various funding agencies. It will appoint the Director and Deputy Directors, approve annual plans and receive annual reports.

A key committee for a group that links the operational and research communities will be an advisory board. This committee will be made up of members of the broad community: universities, mission agencies, operational centers and perhaps the private sector. It would provide the DTC with advice on its annual program, its visitor program, and a possible grants program. It would suggest changes in the DTC program that might extend out to five years. It could also be used to review accomplishments in the previous year.

If the outreach program of NUOPC is to succeed it must meet unmet needs of the operational community. There are a number of ways this can happen. One objective of the DTC will be to provide the operational centers with guidance on the value of new NWP technology through its testing and evaluation program, aid through its visitor program and it can participate in pre-implementation testing programs. If the operational centers can develop a prioritized list of research topics that they would like to see addressed, these topics could be addressed through announcements of opportunity for the visitor or grants programs.

Another objective of the DTC will be to provide to the research community access to the operational models. This will allow evaluation of operational model performance particularly when there are forecast busts which in turn should lead to operational forecast improvements.

Finally, there will need to be people stationed at the operational centers that are dedicated to receiving new technology from the DTC, informing their particular center of DTC testing results and participating in pre-implementation testing of new technology that has been determined will be introduced into operations. This link, which will be referred to as the Operational Testbed Center (OTC), need not be extensive and could be limited to one or two people per center. They must be dedicated to maintaining the link between the operational center and the DTC and they could be considered part of the overall DTC funding.

5. Potential Risks and Roadblocks. We identify five potential risks in implementing the program outlined above. These are the lessons learned from the WRF program:

5.1. Promising More than Can Be Delivered. This was a lesson learned in both the U.S. Weather Research Program (USWRP) and in WRF. If promises are made to the research community to entice them to participate, NUOPC must follow through on those promises very early in the program. It does not suffice to keep putting off implementation to later dates with promises of good things to come.

For example, if there is going to be a grants program, it has to be adequately funded before it is announced. See below for projected costs, but a lesson learned from USWRP is that the level of funding needs to be high enough that the success rate of proposals is reasonable. A success rate under 10% will quickly discourage the community and they will become cynical about it. One way to deal with the success rate issue is to be up front with what the average grant will amount to and how many can be expected to be funded.

The same can be said for the visitor program. WRF DTC experience with its visitor program is that the number of proposal submissions seems to adjust to the number of expected awards, so the success rate stays at about 30%, which is reasonable.

If NUOPC is going to have a DTC it needs to be adequately funded from the beginning and be able to quickly spin up to a level of support to the community similar to what is available for WRF.

5.2. Not Addressing Operational Center Needs. This is a criticism that we have heard frequently from NCEP regarding the WRF program. To some extent this is a control issue, but the outreach program does need to ensure that the operational centers are getting a benefit from the program. This benefit of course has to be manifested as improving operational forecast accuracy in a way that does not force unreasonable additional burdens on the operational centers. This can be achieved through help with pre-implementation testing by the DTC or research community, it can be achieved through focusing on developing technology specifically identified as necessary by the operational centers, and it can be achieved with focused research efforts. Focused research efforts will require that the operational centers develop a priority list of research topics.

A way to ensure that the DTC program does not force additional burdens on the operational centers is to place a person at each center to act as a go-between with the DTC and the OTC. This individual can work with the DTC to link the operational codes with the repository at the DTC and can help with implementation testing at the center prior to making new codes operational.

5.3. Foreign National Issues. This could be a show stopper for interaction with the research community since there are many foreign nationals working on research and development projects at universities and at NCAR. There are even foreign nationals working with NOAA labs and centers through contractors or closely associated research centers. Many of the codes that will be part of NUOPC will be developed by DoD funding and therefore can come under International Traffic In Arms Regulations (ITAR). There are separate regulations related to providing support to nationals from a small group of embargoed nations. These issues have been encountered by the WRF program and solutions have been studied extensively. NUOPC is encouraged to use the experiences of the WRF program in addressing this issue.

One strategy has been to put all the codes that the WRF program maintains in the public domain. Once in the public domain anyone can have access to the codes anywhere in the world and the ITAR regulations don’t come into play in accessing the codes. This also applies to foreign nationals working with the codes in the US. The only issue left has to do with what can be considered as user support to foreign nationals from embargoed countries such as answering help desk questions, helping them get the WRF systems running on their computers and so forth. This is avoided by requiring all users to note their national affiliation when coming to the WRF

DTC for help. We must respectively decline to provide any help to people from embargoed nations.

It is likely that some of the NUOPC participants may not want to put parts of or even any of their codes in the public domain. Then the issue of foreign nationals working with those parts of the NUOPC codes may be problematic. It is beyond the scope of the outreach committee to solve this problem, but we recommend that attention be given to it early in the development NUOPC.

5.4. Information Assurance. This will be of particular importance to all the operational centers when they begin to deal with codes that come from repositories outside their firewall. It is likely that a "research repository" will be maintained outside the operational center firewalls, perhaps by the DTC, so information assurance will need to be practiced carefully. The WRF DTC is addressing this issue and AFWA is familiar with ways to confirm that codes are "clean." This experience can be tapped by NUOPC.

5.5. Estimated Costs. The costs estimates are shown below. Depending on the commitment of the participating agencies, this could be a show stopper.

5.5.1. Visitors Program. The current WRF DTC visitor program can be used as guidance. This program offers visitors one month salary and travel and per diem for one month at the DTC or an operational center. Costs vary with visitor but typically it is about $20K per visitor per year if overhead costs can be minimized. The DTC visitor program has an annual announcement of opportunity and it likes to fund approximately 10 visitors with each announcement.

Annual cost--$200K

5.5.2. Grants Program. A typical National Science Foundation (NSF) grant to a university researcher is approximately $100K/year including overhead of the PI's institution. These typically cover about two months of a PI's salary and one or two student assistants plus travel and overhead. This number is of course going to vary with the PIs and the number of students supported by the grant but should run about $100K/researcher/year. The annual costs shown below assume that all the funding will come from new money, but it is highly likely that at least some of the funding can come from existing research grants programs within the participating agencies.

Annual cost--$1M for 10 grants.

5.5.3. DTC/VTC.

For the current WRF DTC the cost breakdown is approximately as follows

	FTE	Budget
Code maintenance	3.5	$750K
User Support	3.0	$700K
T and E	3.5	$750K
Management	0.5	$100K
Visitor Program		$200K
Infrastructure (travel, computers, Adv. Panel, workshops)		$200K
Total	10.5	$2.7M

It is estimated that the total staffing is about half of what it needs to be to accomplish all the demands that are being placed on it. Removing the visitor program from the above numbers (since it is shown separately above), a reasonable estimate of a fully functional NUOPC DTC would be approximately $4.8M. These costs would be allocated among the various components of the distributed DTC. Some of the costs could be shared with other programs (for example, with the WRF program unless a separate DTC for NUOPC is desired) and could use existing personnel at the various DTC components.

Annual costs--$4.8M

Total annual costs for the outreach program--$6.0M

6. Recommendations. We recommend that the following actions be started immediately:

- Begin developing a management plan for DTC

- Begin developing an implementation plan for DTC

- Assess agency willingness to contribute resources to a grants program

- Assess agency willingness to support a visitor program

- Assess agency willingness to support a DTC

7. Benefits and Measures of Success. The benefits of the program have been noted above. Whether these benefits are being met can be measured by:

Accelerated milestones for operational improvements in forecast accuracy

Objectives of the operational centers that are being met through DTC and research community testing and evaluation and through targeted research

Interest by researchers in meeting operational needs, measured by:

- Visitor requests

- Grant proposals

- Downloads of NUOPC codes

Note: After the DTC proposal was written, committee members expressed a desire to communicate one final item of consideration for the ESG, and that was the location of the NUOPC DTC. A non-government agency is not recommended as the DTC host (i.e., centralized facility) as it would result in the need for competitive selection and periodic recompete processes. Thus, to avoid the competition (and re-competition) issue, it is recommended that a government agency serve as the DTC host. Government agencies also offer many advantages over non-government agencies, including less conflicts of interest, and fewer contractual and legal entanglements.

APPENDIX 4 – NOGAPS Scorecard.

Type	Level	Area	Parameter	Error-Type	Fcst Time	Weight
Field	Surface	Tropics	Tropical cyclone	Track error	96 hrs	4
Field	500 mb	N Hem	Height	AC	96 hrs	4
Field	1000 mb	N Hem	Height	AC	96 hrs	1
Field	500 mb	S Hem	Height	AC	96 hrs	1
Field	1000 mb	S Hem	Height	AC	96 hrs	1
Field	850 mb	Tropics	Wind	RMS	72 hrs	2
Field	250 mb	Tropics	Wind	RMS	72 hrs	1
Field	850 mb	N Hem	Wind	RMS	72 hrs	1
Field	250 mb	N Hem	Wind	RMS	72 hrs	1
Buoy Data	Surface	Global	Wind	Speed Error	72 hrs	2
Raob Data	850 mb	Global	Wind	RMS	72 hrs	1
Raob Data	250 mb	Global	Wind	RMS	72 hrs	1
Raob Data	850 mb	Global	Temperature	RMS	72 hrs	1
Raob Data	250 mb	Global	Temperature	RMS	72 hrs	1
Raob Data	500 mb	Global	Height	RMS	72 hrs	1
Raob Data	50 mb	Global	Height	RMS	72 hrs	1

AC = anomaly correlation

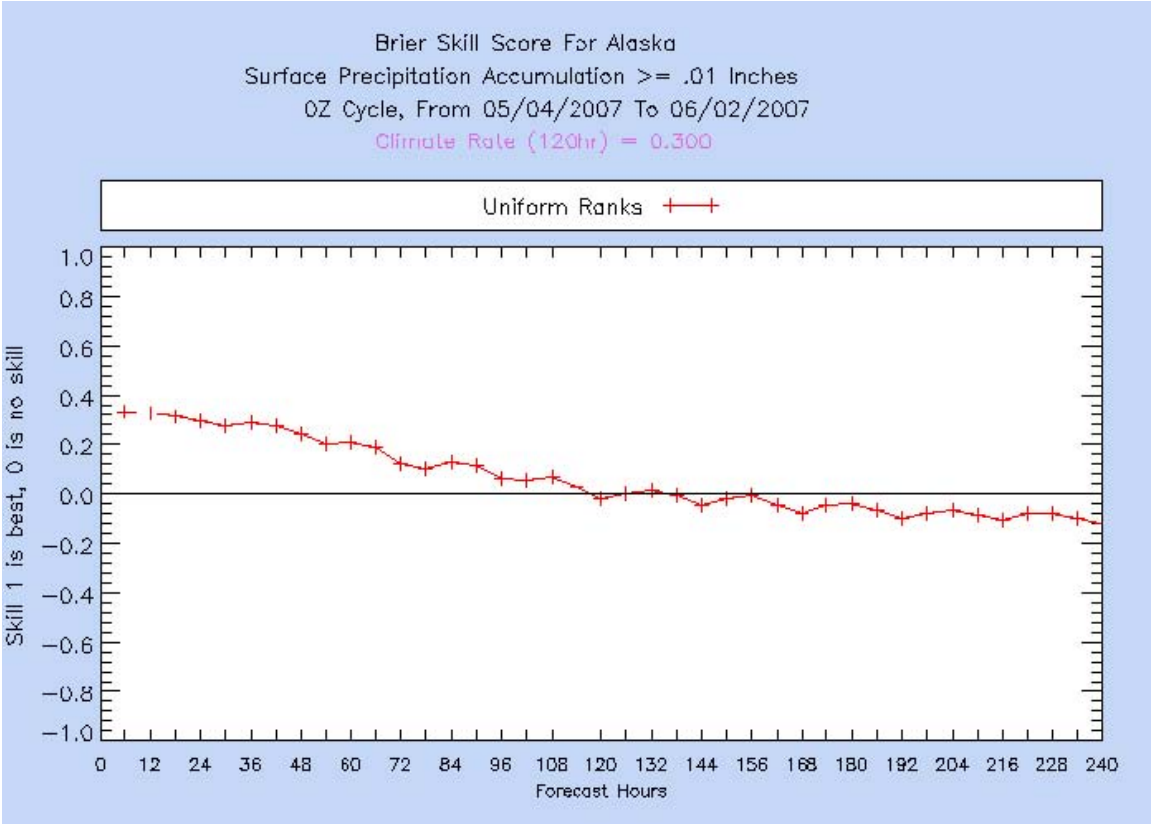
RMS = root mean square error

The scorecard is used to rate NOGAPS data assimilation/medium range test results versus a control (usually operational) NOGAPS. For each category there is assigned the positive weight, zero, or the negative weight, depending on whether the results were better than, neutral, or worse than the control. For TC track error, RMS, and speed error the criterion for a positive score is that the error must be at least 5 percent less than the control with a statistical significance level of

95 percent. For anomaly correlations, the scores must be statically higher with a confidence level of 95 percent. If the opposite is true, i.e. the control is significantly better, a negative score is assigned. The total score is the sum of all categories. A perfect score of one test over another is 24. A poor test would be any that received a negative total score.

APPENDIX 5 – JEFS Description of Ensemble Verification Charts.

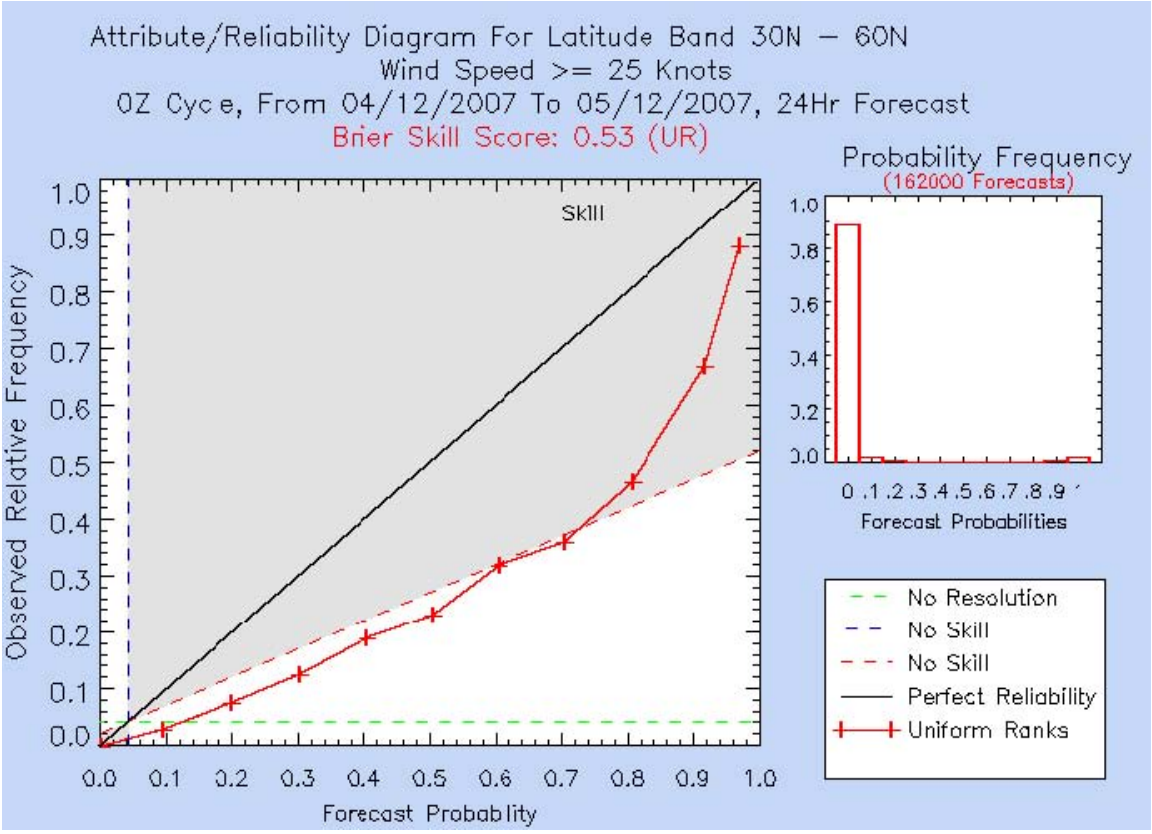
1) Brier Skill Score



The Brier Score Skill (BSS) is the comparison of the Brier Score (BS) of the ensemble forecasts compared to the BS of the reference forecast – in this case climatology. The BSS measures the skill of probability forecast for a specific weather event. It uses the Brier Score which is similar to the Root Mean Square Error (RMSE) commonly used in deterministic weather forecasting, except that is uses probability forecasts of an event compared the frequency of occurrence of an event instead of model forecast values compared to observed values. A perfect probability forecast will have a BSS of 1 and probability forecast which is no better than climatology will have a BSS of 0.

The climate rate is referenced to show how uncertain the event is overall, useful to ensemble model developers. For more information on the Brier Skill Score, see “Forecast Verification” by Ian T. Jolliffe and David B. Stephenson.

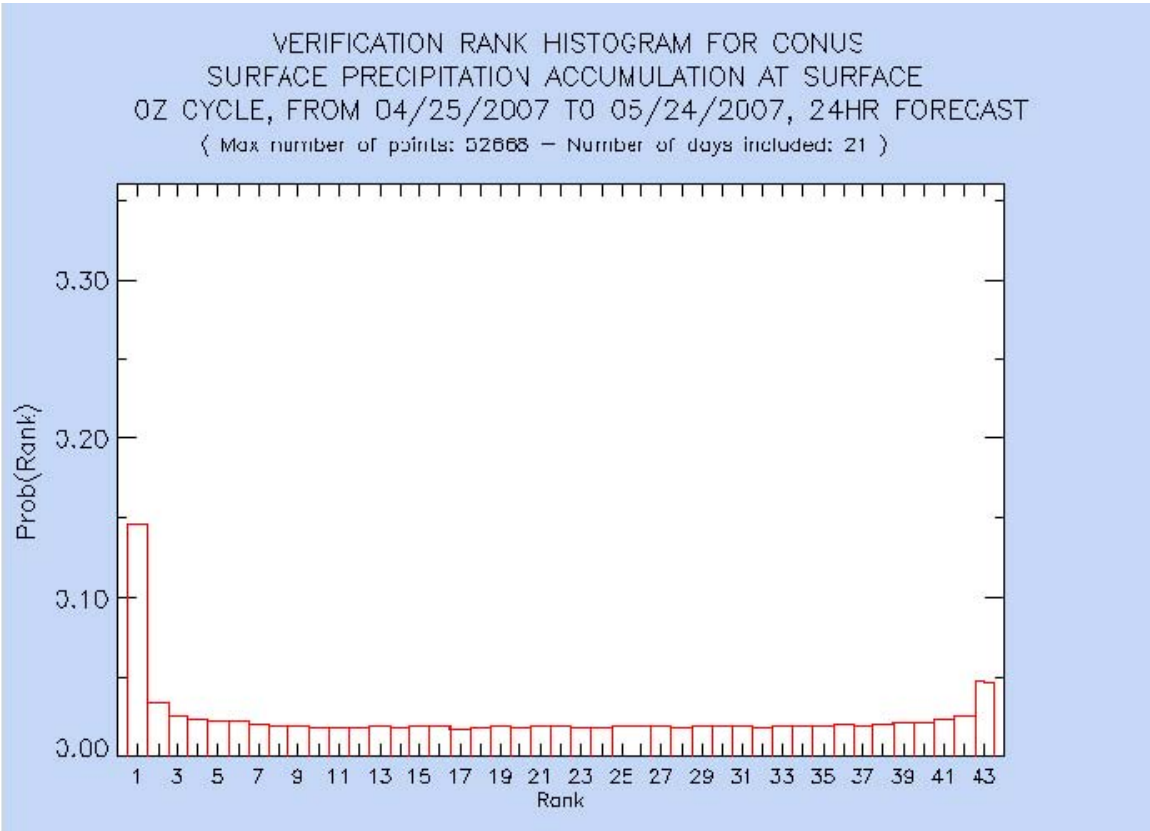
2) Reliability/Attribute Diagrams



The Attribute/Reliability diagram is a plot of observed frequency of an event to its forecast probability. The forecast probabilities have been sub-divided into ten bins. For a perfect forecast, the forecast probability will be the same as the observed frequency and the plotted line will lie on the diagonal. Forecasts with skill will have the plotted forecast probabilities falling in the shaded region on the chart. The Brier Skill Score is included on the chart to show the overall skill of the forecast.
The Probability Frequency diagram shows the frequency a particular range of probabilities are forecast by the ensemble system. The number of forecasts included in the diagram is indicated above the diagram. (162000 forecasts points in this case)

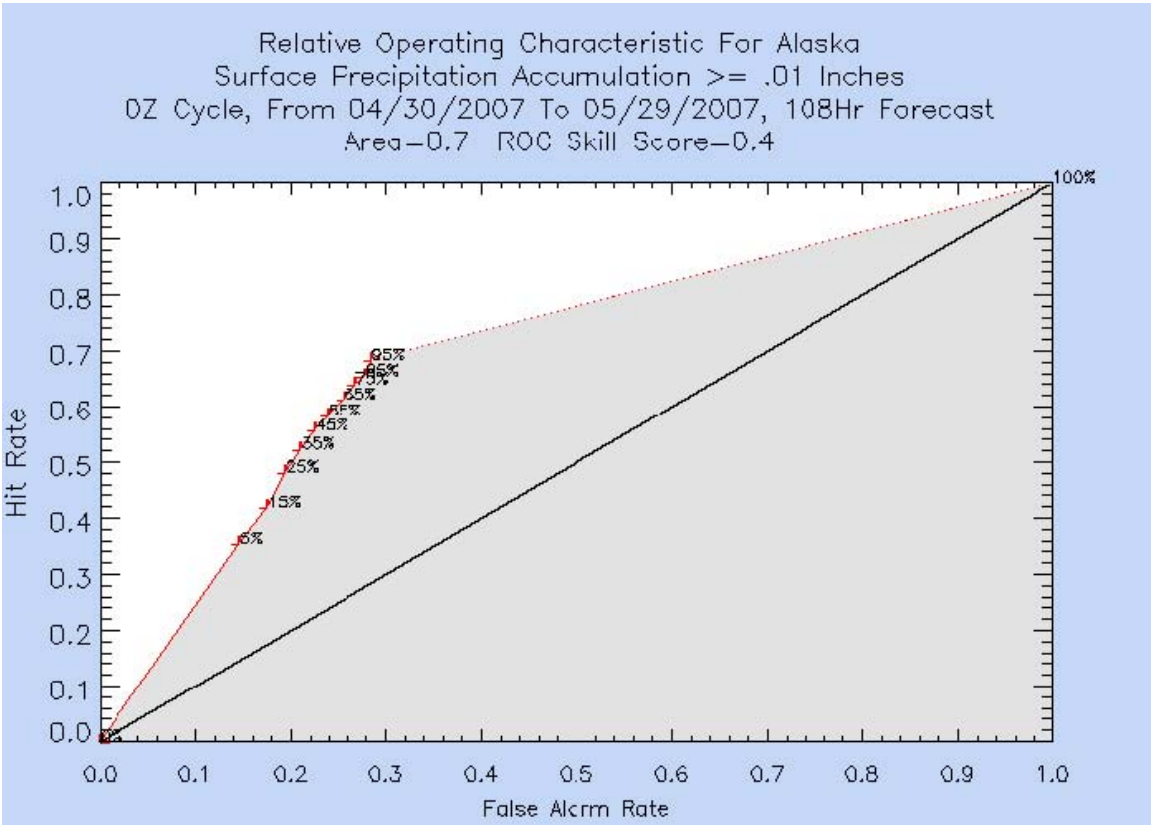
For ensemble model developers we also include the no skill and no resolution lines. The red dashed line (labeled no skill) marks the point where the forecast probabilities contribute to the overall skill of the forecast. Technically, this represents forecasts where the resolution term in the Brier skill score is larger than the reliability term (See "Statistical Methods in the Atmospheric Sciences" – Daniel Wilks) for a thorough explanation. The dashed green line (labeled no resolution) is the climatological frequency of the event. Forecasts probabilities on this line are no better than climatology. The dashed blue line is a reflection of this line on the x axis. Ensemble forecast probabilities which contribute skill to the ensemble will be plotted above the red dashed line and to the right of the dashed blue line (in the shaded area).

3) Verification Rank Histograms



The verification rank histogram indicates the number of times truth fell within the ordered ensemble members. An ideal ensemble with the correct amount of dispersion of the ensemble members will have truth falling randomly and evenly across the ranks - the bars of equal height. Ground truth for AFWA's ensemble is the global model from United Kingdom Meteorological Office (UKMO). The number of points being plotted is shown under the title along with the actual number of days present in the period being displayed.

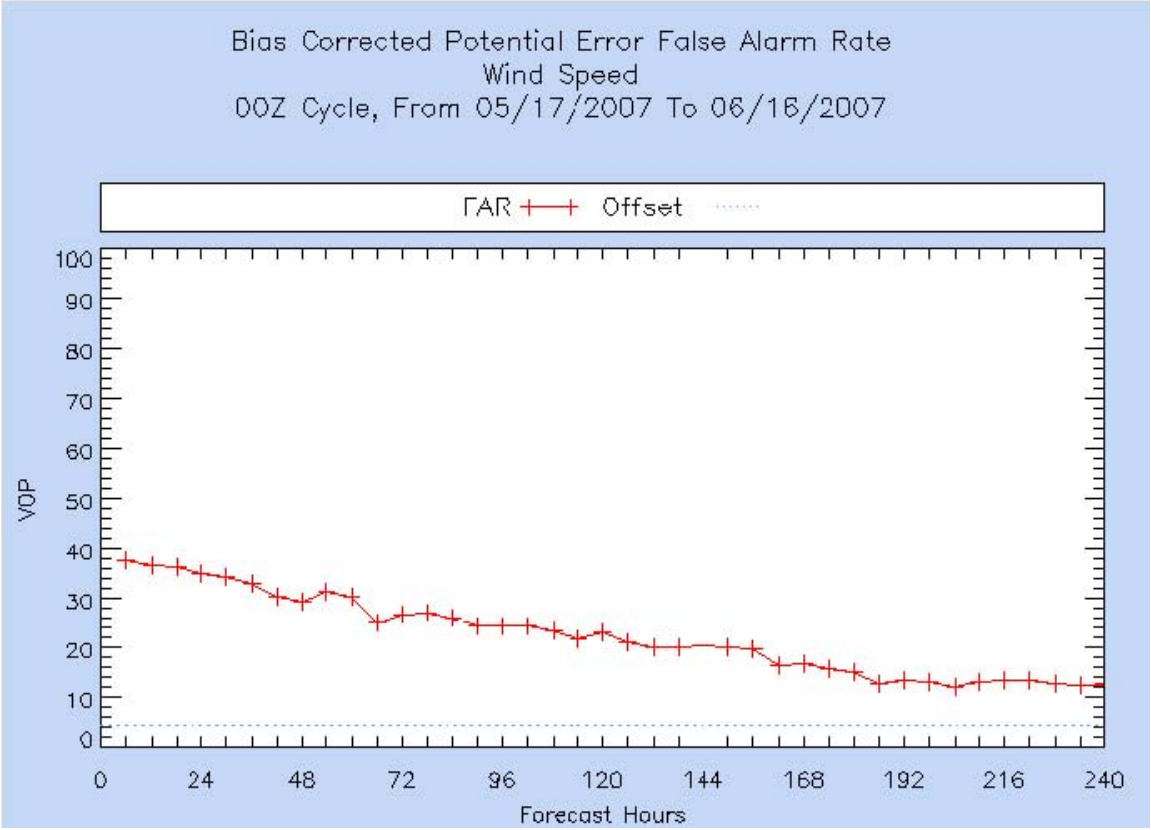
4) Relative Operating Characteristic (ROC)



The ROC curves is a plot of hit rate versus false alarm rate plotted at various decision thresholds for an event. The ROC measures the extent to which observations of the event correspond to relatively high probability forecasts and observations of the non-event correspond to relatively low probability forecasts. The gray area above the diagonal is a measure of the amount of skill of the forecast system (ROC Skill Score – 0.4 in this case). A forecast system with no discrimination ability will have the plotted curve (red line) lie on the diagonal. Since the gray area above the diagonal is zero this forecast system would have no skill. A forecast system with perfect discrimination ability will have the plotted curve on the left and top axis. This will lead to the gray area filling the chart above the diagonal and skill will be perfect.

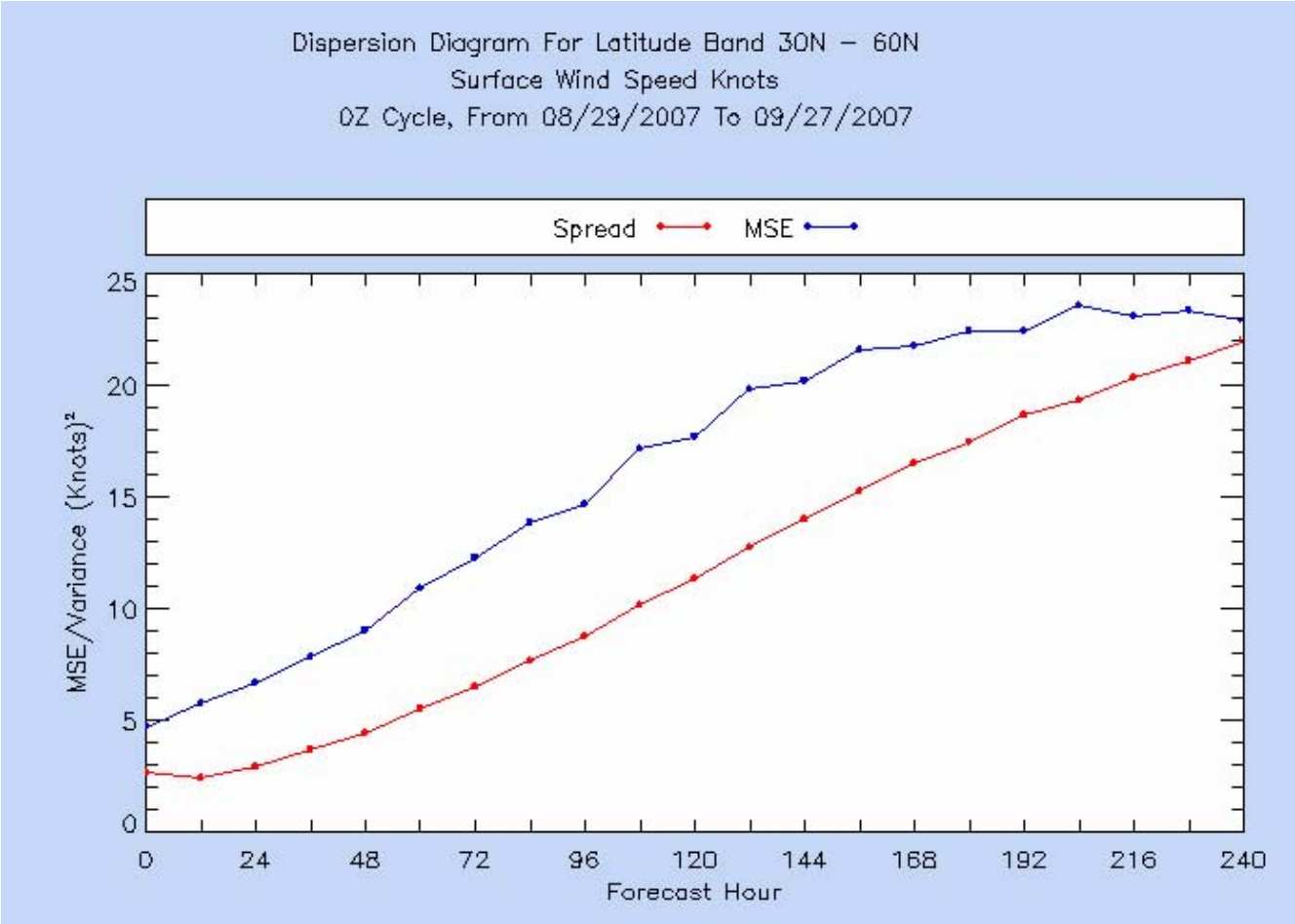
The numbers on the curve show the various thresholds used to calculate the ROC curve.

5) Potential Error False Alarm Rate



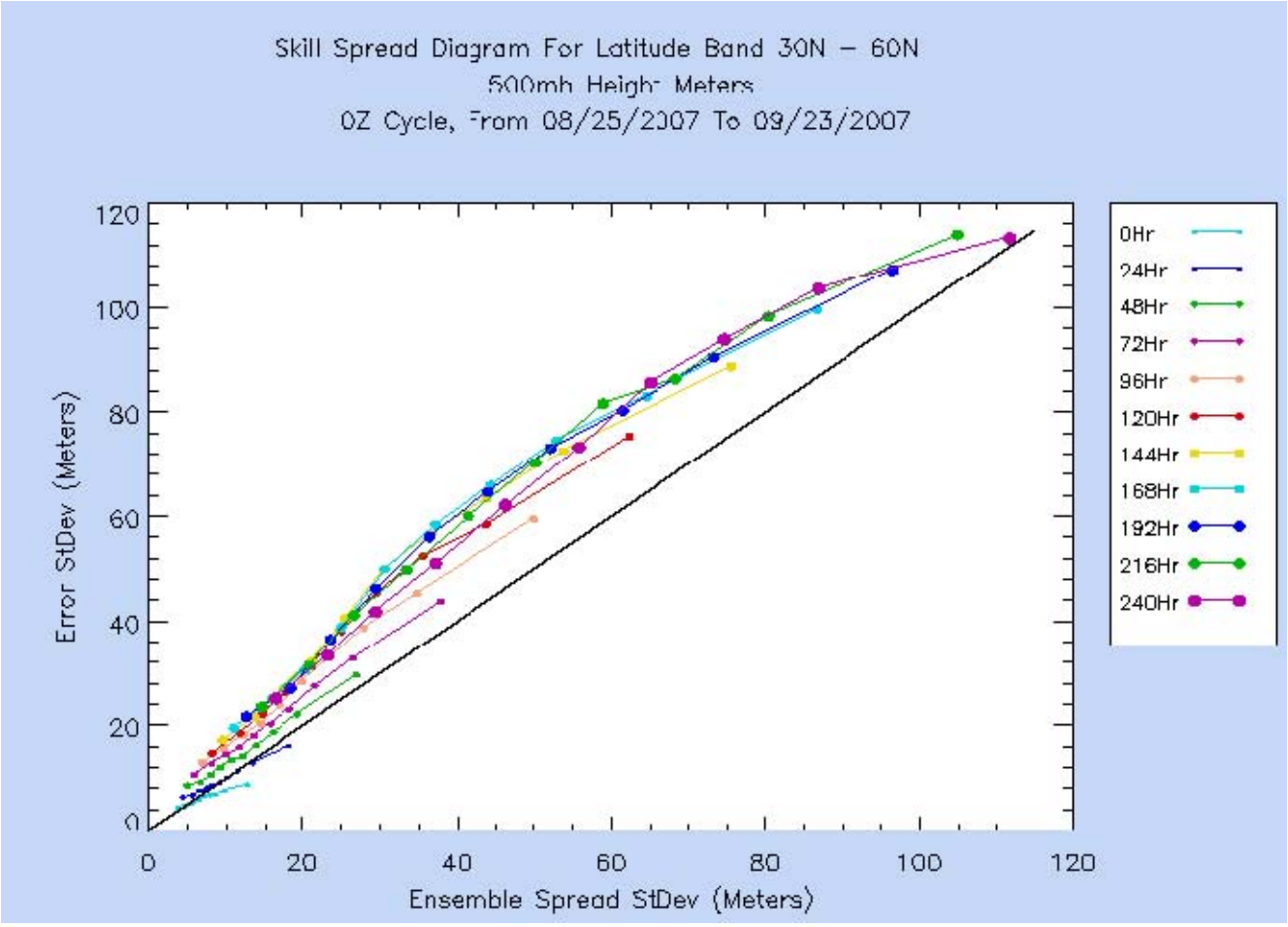
The potential error false alarm rate (PEFAR) is a measure of the percentage of ground truth observations that fall outside of the potential error bounds. These bounds are 2 standard deviations above and below the mean, where both mean and standard deviation are calculated from all the ensemble members at any single grid point. The goal is for the FAR percentage to be approximately 4.56%. This number is derived from the area of the standard normal distribution ("bell-shaped") curve that is more than 2 standard deviations from the mean. It is counterintuitive that errors in the forecasts are acceptable. The reason for the nonzero rate is a compromise between error minimization and forecast sharpness. For example, a forecast temperature of -40 to 120 degrees F is almost guaranteed to never be wrong at the mid-latitudes, but such a wide range provides no useful information about the most likely temperatures. The nonzero rate allows for narrower forecast ranges which ideally encompass truth over 95% of the time. The graphics that appear on the JEFS page illustrate the ensemble system's performance against the target FAR percentage, which appears as a horizontal line at the 4.56% error level. This error rate is plotted as a function of forecast lead time from 6 to 240 hours. All grid points within the domain or latitude band are grouped together on these plots, and the PEFAR values are cumulative over a period of 30 days.

6) Dispersion Diagram



This chart compares the mean squared error (MSE) of the ensemble mean compared to ground truth with the ensemble variance (Spread) of individual members around the ensemble mean. For a perfectly reliable ensemble forecast, the spread in the ensemble forecasts about the ensemble mean should equal the error in the ensemble mean compared to ground truth– the blue and red lines should be on top of each other. In this case the ensemble spread is smaller than the mean error – ensemble spread has been under- estimated.

7) Skill Spread Diagram



This chart shows the degree of correlation between the ensemble mean error standard deviation (difference of ensemble mean from ground truth) and the standard deviation of the ensemble members from around the ensemble mean. The standard deviations have been subdivided into 10 equally populated categories (the dots are the categories). An ensemble with the ideal amount of dispersion around the ensemble mean will have all lines on the diagonal.