Task 3.5.2 Visualization & Extraction via NEIS Final Report

I. Overview

Task 3.5.2 successfully expanded the prototype NEIS system adding capabilities to help others visualize, interact, and integrate HIWPP data with other geolocated environmental data. Through the end of the project, the following improvements were made:

- Developed and implemented architecture for high performance real-time research data ingest system for new high resolution gridded model data, point data. This system provided existing global operational models, global satellite observations, as well as available experimental HIWPP data to users, including FIM, GFS, and NAVGEM. All data is made available for at least the previous 24 hours, often up to 48 hours.
- Created high performance service layer allowing fast access, visualization, and integration capabilities to real-time research data.
- Developed, Side by Side or Multi sphere display allowing users to visually compare forecasts in real-time. The **Figure 1** below provides an example of the side by side capability with forecasted simulated IR imagery on left, with actual observed global composite IR imagery on right.
- Created quick access comparison features to allow easy access to comparisons of HIWPP FIM, GFS, and NAVGEM data against observations over the last 7 days.
- Implemented tiling capability to allow progressive disclosure of very high resolution gridded data including non-hydrostatic GFDL FV3 and NCAR MPAS data. **Figure 2** shows a comparison of 1 hour precipitation from these models for the Hurricane Sandy case.
- Created initial high performance processing capability to provide analytic capabilities allowing users to run analytics or algorithms remotely only sending relevant data for display
- Integrated Statistical Post Processing data from task 3.5.1.
- Updated visualization capability (TerraViz) to provide high performance tools for model evaluation and comparison, including contours, wind barbs, vectors, and particle trails, as well as analysis tools to probe underlying data.
- Significant bug fixes, refactoring, and testing suites to improve overall stability and performance of client
- Delivered updated visualization client for MacOS and Windows users several times a year.

Technology developed as part of this tasking has been successfully transitioned and used in other applications outside of the HIWPP project including the Science on a Sphere Explorer (SOSx) program, and for visualization of VIIRS data from the NESDIS JPSS program. Additionally TerraViz was successfully used as a main presentation medium within the NOAA Exhibit booth at both the 2015 American Geophysical Union Fall Meeting and the 2016 American Meteorological Society Annual Meeting.

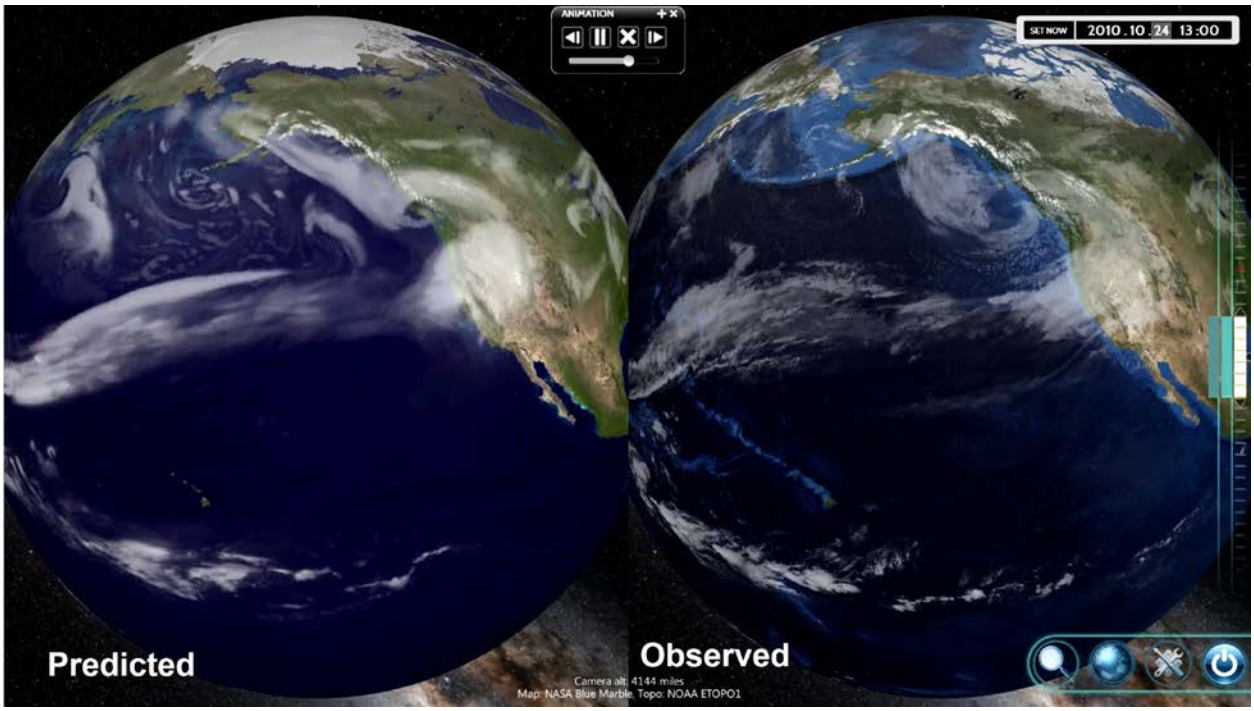

Figure 1. Side by side comparison of forecast from the FIM model with Global Infrared Observation Composite

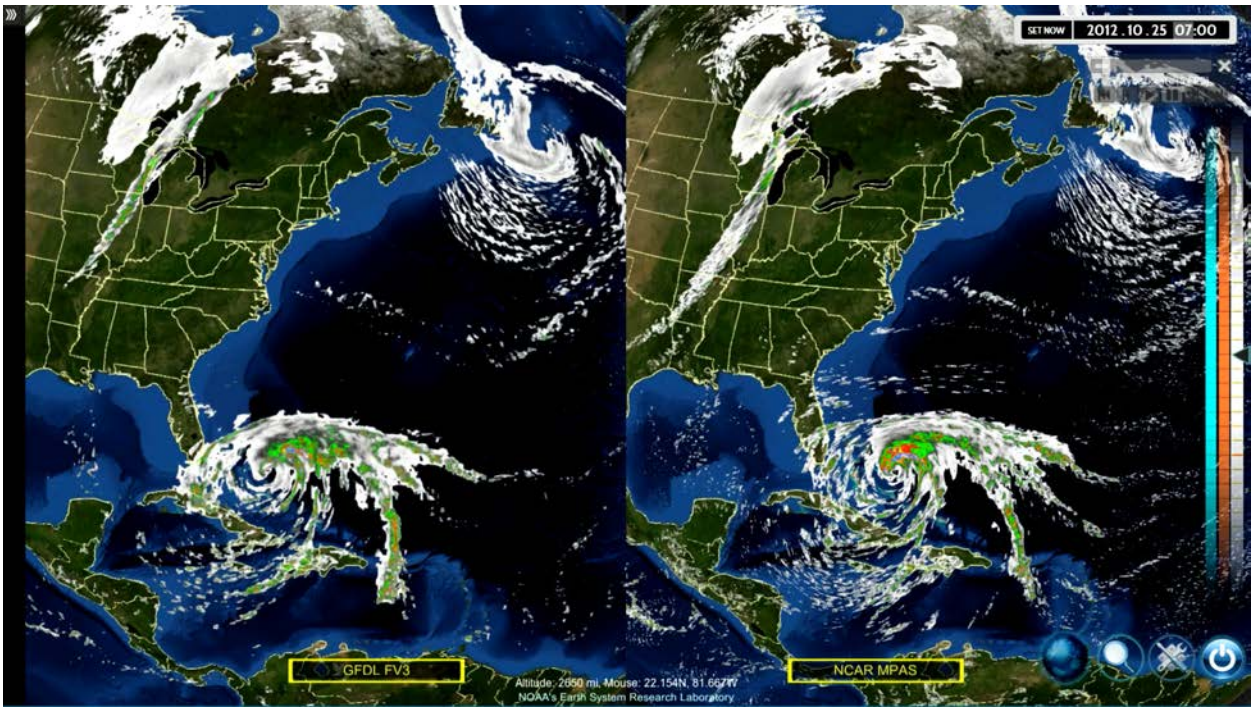

Figure 2. Comparing Non Hydrostatic FV3 and MPAS 1 Hour Precipitation for the Hurricane Sandy Case

II. Background

Through HIWPP, significant advances in science were made to improve global prediction systems. Along with improvements to models themselves, advances in visualization and extraction capabilities were necessary to effectively test and evaluate the resulting large

volume (i.e. “Big Data”) prediction systems. A key part of the evaluation process is the ability to visualize how these experimental models are performing by providing capabilities to integrate other time- and space-matched earth information data in real-time, such as existing global predictions systems and global observations. Through the Test Program Visualization and Extraction task, ESRL leveraged the NOAA Earth Information System (NEIS) and developed an advanced access, visualization and analytics framework to provide on-demand access and integration of experimental global-scale models, and other earth system data. This capability will complement those now present at the operational center (NCEP) by allowing for more extensive testing of new technologies that may be a large departure from the current operational system. It will also allow for more involvement of a diverse community of users in the development, testing and evaluation process,. as well as prepare for future systems where data volumes greatly exceed existing bandwidth available to end user systems.

III. Task 3.5.2 Team

Team Member	Responsibilities
Jebb Stewart	Program Management, Technical Guidance, Backend Services
Jeff Smith	Software Engineer, Backend Services, Visualization and User Interface
Jonathan Joyce	Software Engineer, Visualization and User Interface
Chris Golden	Software Engineer, Visualization and User Interface
Eric Hackathorn	Software Engineer, Visualization and User Interface
Julien Lynge	Software Engineer, Visualization and User Interface
Tracy Hansen	User Interface, Capability, and Service Consultant
Don Murray	Consultant
MarySue Schultz	Software Engineer, Backend Services
Chris MacDermaid	GSD/ITS Project Manager
Randy Pierce	Software Engineer, Backend Services

Table 1. The various team members and their roles over the lifetime of the project. Note: not all members were available for the entire project

IV. Task Activities and Development

A. Hardware Procurement

To support near real-time high performance visualization of forecast data produced as part of the HIWPP project, new hardware was needed. The team evaluated existing options and decided on a hardware configuration that included large network based storage purchased in conjunction with the Task 3.5.4 Real-Time IT Operations, High Performance NVidia GPU's for server based visualization and data processing, as well as local Solid State Drives (SSD's) to store temporary or pre-processed data for fast retrieval.

B. Data Transfer Evaluation

To meet the requirements allowing users to have high performance access to remote high resolution data, the data format for transferring data is important. The evaluated was performed based on several criteria:

- The speed of data read and decompression on the client.
- The resulting compressed file size which impacts data transfer speeds.
- The speed of compression on server.

The NEIS team evaluated several formats including JPEG, zipped raw data, wavelet compressed data. Ultimately, the best tradeoff was a compressed Direct Draw Surface (DDS) file format. This format can be directly read by the GPU on the client side and the resulting file size was comparable to JPEG imagery.

C. Service Side Development

For this tasking, server side development focused on speed and event driven nature. As soon as data was available, it should be processed and sent to end user. The development focused on asynchronous frameworks to improve overall utilization of hardware and event driven nature to process data on demand. For the data ingest software, the developed system was based on Apache Storm, an "open source distributed real-time computation system. Storm makes it easy to reliably process unbounded streams of data". Apache Storm is used by Twitter along with many other organizations to process data in real-time. Through Apache Storm, data flows through the system as a series of events and individual processing functions act on data that contained matching metadata tags. Processors were built to extract metadata, reprocess, combine data into new fields, or generate preprocessed data to improve the speed on the client side. Ultimately, the data ingest systems processed and delivered multiple terabytes of Gridded, Imagery, and Point data daily.

Another significant open source project used is the Vert.x Java Framework. Vert.x provided an asynchronous messaging framework and a polyglot nature and handled all interactions with the remote client through service API's. Vert.x was chosen because of it ability to provide a polyglot environment. The Polyglot nature allows users to write code in numerous different languages and use the same underlying api to access data. This was important for the processing framework as it was desired to allow users to develop algorithms in languages they were familiar with to allow processing data server side before it is sent to the user. Additionally the Vert.x framework provided a high performance platform for the HIWPP service API's

handling many thousands to millions of requests per second important for scaling the application.

While the main focus of HIWPP development was on high performance research grade services. The development team worked to insure the system developed would work and could be supported in an operational environment. The Both the Vert.x Framework and Apache Storm software were chosen because they provided fault-tolerant, distributed based platforms. Various monitoring and alerting systems were developed to insure operators had knowledge of the state of the system at any time and low-level problems automatically resolved. These were viewed as important steps in development to ease the transition to an operational environment from research.

For Task 3.5.2, the team investigated using GPU's on the server side to improve the overall performance of the system. Software was developed to use GPU's in two different ways. For the remote processing of data, some complex algorithms benefited from using the many core's available in GPU's and significant speedup was observed. The remote processing capability has hooks for those developing algorithms to take advantage of the GPU's if needed. GPU's were also used for wavelet data compression of data. Ultimately another data format was chosen but it's believed this compression will be useful when more volumetric capabilities are developed.

D. Visualization Development

Before the HIWPP project, NEIS was using a visualization software called TerraViz developed on the Unity Game Engine platform. This project leveraged this technology to allow new high performance visualization of NOAA data through both space and time. The main focus at the time was on interoperability and discoverability. NEIS could pull data in from a number different sources and provided an improved interface to filter out the best products for users. This common operating platform removed the need for knowledge of different file formats or server locations. With HIWPP the focus turned to high performance visualization of large dynamic datasets and development of analysis tools to allow users to better compare, probe, and understand the underlying data.

Through these development efforts the following enhancements were made:

- Side by Side Spheres for Comparison up to 4.
- Greatly improved local data capability.
- Tiled data support allowing progressive disclosure for very high resolution data (Non-Hydrostatic) both in sphere mode or in flat map mode.
- Analysis tools including real-time data probing and transects (**Figure 3**).
- Annotation capabilities for improved presentation support.
- Several performance improvements to support multi-globe and high resolution data including caching optimization and multi-threaded data loading.

- New visualization capabilities including contours, wind barbs, vectors, particle trails.
- Support for visualizing aircraft flight tracks, including graphing/probing of dropsondes.
- Custom color palettes.
- Dynamic color palettes (optimal palette is generated automatically as user zooms in).
- Point data integration.
- Custom algorithms capability allowing users to access server side processing of remote data.
- Improved performance lower end systems.
- Integrated feedback mechanism to populate google feedback form which is directly received by developers.
- "Tour" capability to script how and when the visualization loads data, animates, as well as integrate ancillary content such as images, websites, or powerpoint slides.
- Significant bug fixes, refactoring, and testing capabilities to improve client stability and performance.
- Improved data discovery and ability to easily add new data sets.
- Quick access comparisons for HIWPP models (500 mb, Derived IR, Surface precipitation, Mean Sea Level Pressure).

Software releases were made available to open initiative teams several times per year for both MacOS and Windows 64 users. These releases were published on the CoG HIWPP Open Data Initiative Website.

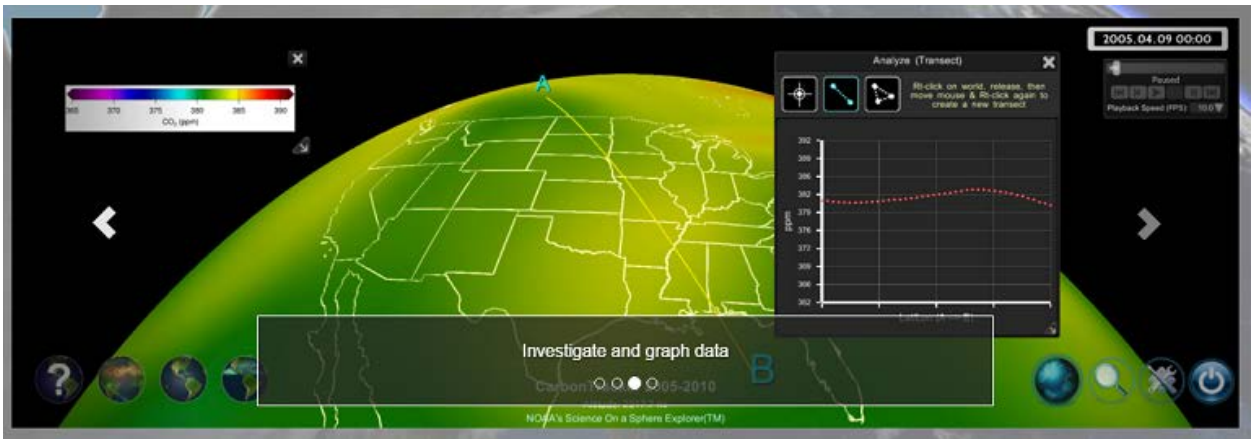

Figure 3. Using TerraViz analytical tools to probe data

V. Challenges and Lessons Learned

Throughout the project, task 3.5.2 encountered several challenges that impacted the schedule and deliverables. The largest problem that faced the team for task 3.5.2 was the change of the target audience for the software. During initial development and hardware procurement, it

was expected the resulting software would be delivered to the "Trusted User" program with a limited number of users on specific hardware and with high speed connections to the NEIS servers. Following legal review during the first year of this task, the "Trusted User" program became the "Open Data Initiative" where any interested party could request access to the client. In terms of development, this meant working with lessor client hardware and unknown internet connectivity. This change of scope meant design decisions made prior to the change required modifications to support these potential new users. Ultimately the changes were numerous and invasive to the existing system and the decision was made to stick to the original scope. NEIS remained a restricted part of the Open Data Initiative where users with reasonable hardware and internet connectivity were allowed to participate.

Tasks 3.5.2 operated on a strict schedule of deliverables and limited room for problems. This situation is not advisable for a research project. Future tasking for research projects should allow more flexibility of schedule and deliverables as more knowledge is learned.

Towards the end of the first year, the NEIS team lost a key software engineer. This position was filled promptly within 1.5 months. Time was lost both because of lack of coverage during the interview and hiring process, and the on boarding required to bring the new software engineer up to speed.

Additionally, delays from other dependent HIWPP tasks resulted in challenges in having proper data and hardware available for testing and implementation of NEIS concepts. NEIS team was able to move forward with test with example data and prototype hardware, but final testing and implementation was delayed until hardware was in place and data was reliably being transferred.

Throughout the HIWPP project, the NEIS team had various outside users and testers. Only a handful of reliable users provided regular feedback for improvements and feature requests. Future projects should attempt to find a dedicated user pool to provide more reliable source of feedback and future improvements.

VI. Research to Applications (R2A)

Through this research, technology was successfully developed and is currently being used in applications beyond HIWPP. Technology will continue to live beyond HIWPP through the following projects:

- The visualization technology is used as part of the NOAA Science on a Sphere Explorer (SOSx) to create an interactive Earth for flat screen displays including those projected on walls, computers, and large displays, providing teachers, students, and the public their

own personal SOS. In addition, tools included in the application allow users to zoom into, probe, and graph the data, as well as add supplementary material including websites, videos, pictures, and placemarks. Lesson plans and pre-programmed tours through standards-relevant topics are available for teachers through the Science on a Sphere Explorer website.

- The NEIS processing infrastructure along with the visualization technology was the key technology used within the NOAA Exhibit booth at the American Geophysical Union Fall Meeting in both 2014, and 2015. This software was also successfully used by many scientist to present their research at the 2016 American Meteorological Society Annual Meeting alongside the NOAA Science on a Sphere.
- The NESDIS JPSS team is using NEIS to generate high impact visualizations of the VIIRS data for scientific outreach and presentations.

VII. Presentations

Stewart, J. 2014: NOAA Earth Information System (NEIS), NOAA Environmental Data Management Workshop, Silver Spring, MD, September 9th – 11th 2014.

Stewart, J. 2014: Challenges in High Performance Data Access and Visualization, NOAA Environmental Data Management Workshop, Silver Spring, MD, September 9th – 11th 2014.

Stewart, J. 2014: Data Formats and NEIS, NOAA Environmental Data Management Workshop, Silver Spring, MD, September 9th – 11th 2014.

Stewart, J., E. Hackathorn, J. Joyce, and J. S. Smith, 2014 Real-time Data Interaction and Visualization using the NOAA Earth Information System (NEIS), Meteorological Technology World Expo, Brussels, Belgium, October 21st - 23rd, 2014.

Stewart, J. Q., E. Hackathorn, J. Joyce, C. MacDermaid, R. Pierce, and J. Smith, 2014: High Performance Real-Time Visualization of Voluminous Scientific Data Through the NOAA Earth Information System (NEIS). Visualization Technologies Enhancing Earth and Space Science Data Usability (IN040), San Francisco, CA, Amer. Geophys. Union, Dec. 15 – 19 2014.

Stewart, J., E. Hackathorn, J. Joyce, and J. S. Smith, 2015 High Performance Real-Time Visualization of Voluminous Scientific Data Through the NOAA Earth Information System (NEIS). 31st Conference on Environmental Information Processing Technologies (EIPT), Phoenix, AZ, Amer. Meteor. Soc. CD-ROM, 15B.2, January 4th – 8th 2015.

Stewart, J. Q., E. Hackathorn, J. Joyce, and J. Smith, 2015: Breaking it down, using modular services to improve the NOAA Earth Information System (NEIS). Scalable and Adaptable Architecture for Earth Science Cyberinfrastructure (IN31C), San Francisco, CA, Amer. Geophys. Union, Dec. 14 – 18 2015.

Stewart, J., E. Hackathorn, J. Joyce, and J. S. Smith, 2015 Breaking it down, using modular services to improve the NOAA Earth Information System (NEIS).. 32nd Conference on

Environmental Information Processing Technologies (EIPT), New Orleans, LA, Amer. Meteor. Soc., January 11h – 14th 2016.