NOAA Global Systems Laboratory

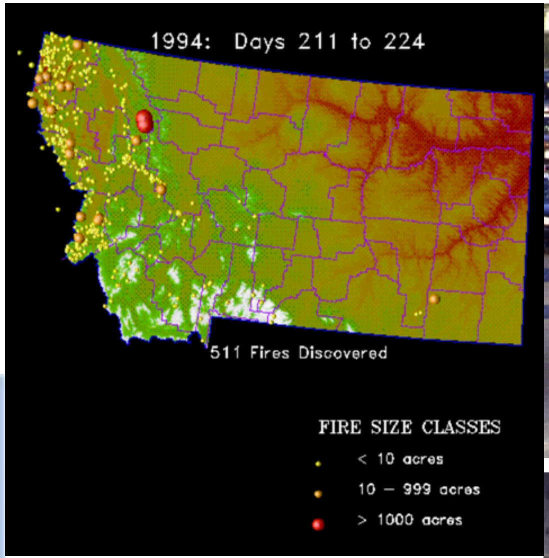
Verification and Evaluation

Matthew Wandishin
Branch Chief
Forecast Impact and Quality Assessment Services Branch

GSL Verification

Model/ assimilation system	Horizontal resolution	Number of vertical levels	Assimilation frequency	Implemented at NCEP
RUC1	60 km	25	3 h	Sep 1994
RUC2	40 km	40	1 h	Apr 1998
RUC20	20 km	50	1 h	Apr 2002

Forecast verification at GSL goes back roughly three decades, back when we were known as Forecast Systems Laboratory



Why Verification?

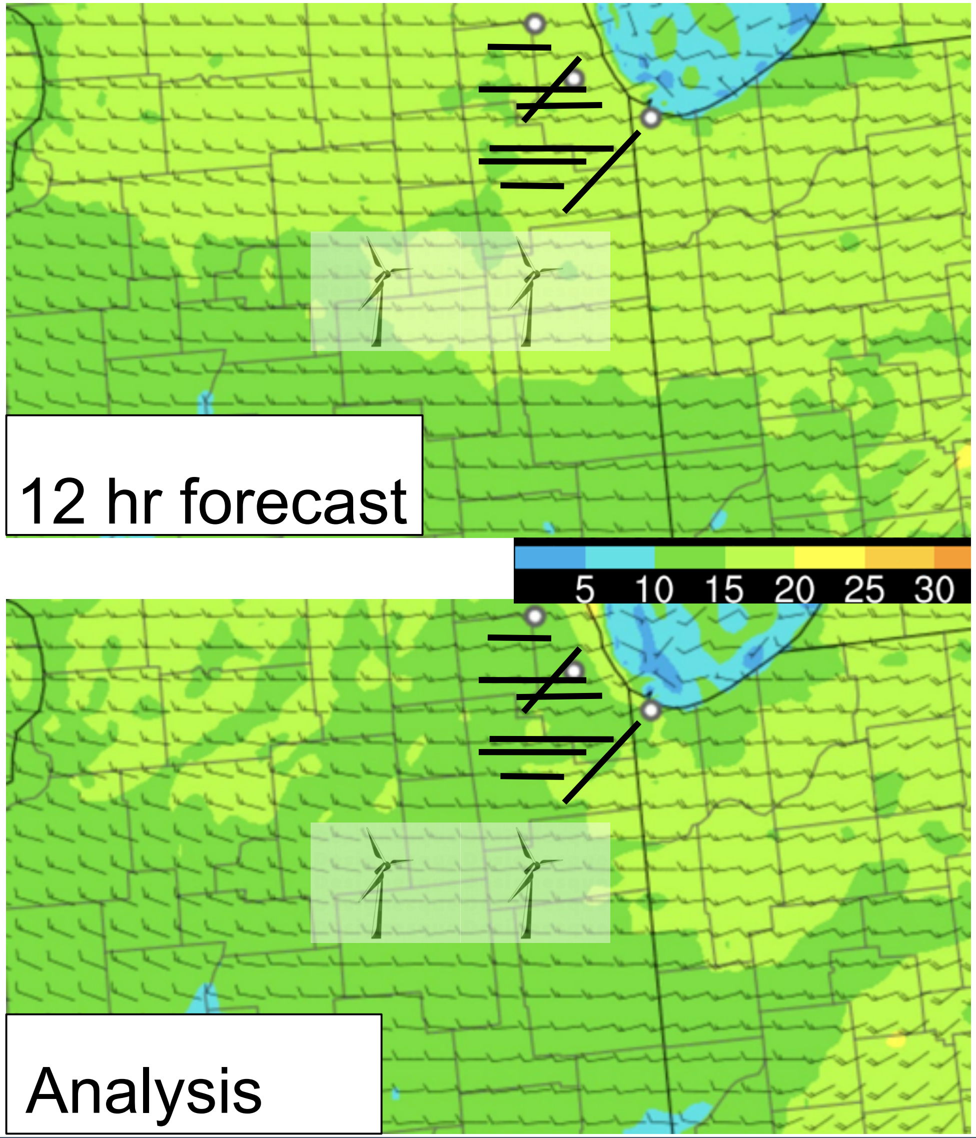
Users

Model Development

Managers

Verification

What is Verification?



Q: Was this a good forecast?

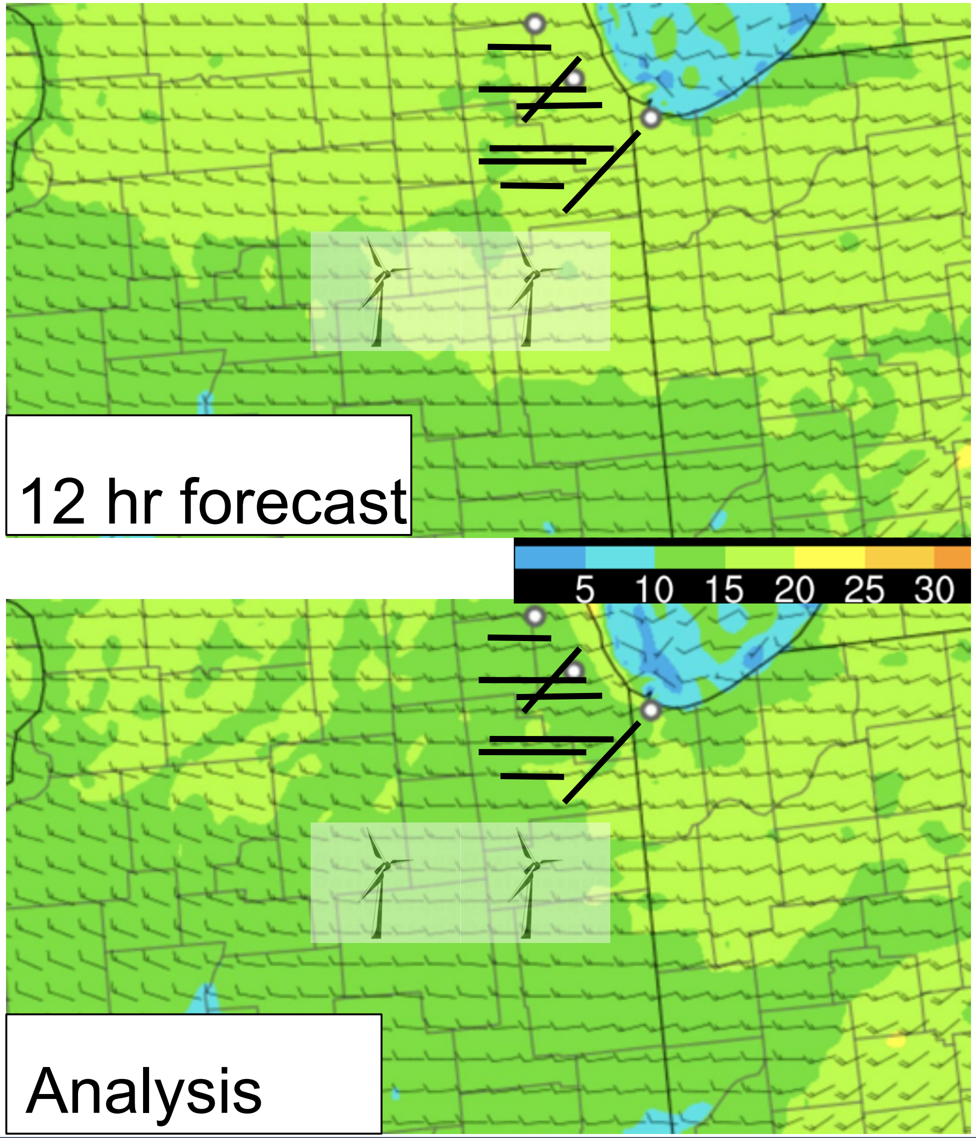
A1: Wind directions are good but speed is ~5 kts too fast around Chicago and to the southwest.

A2: Check land-surface representation, low-level clouds, fluxes.

A3: Forecast correctly identified west-to-east operations on main runways. So an effective forecast.

A4: Wind speeds were too high and in a steep part of the power curve, resulting in an expectation of nearly twice the level of electricity production as was actually possible. Could lead to expensive purchases on the spot market.

Why Verification?



Verification provides feedback about forecasts

- To forecasters--what do we need to improve; how might we go about it? (Jeff Hamilton)
- To users--how does the forecast perform near our decision points; how can we best make use of this product? (Tanya Peevey)
- On economic benefit--can we quantify the benefit to society of forecast improvements, the return on investment? (Dave Turner)

From the Last Review

2015 Recommendation		Presenter
B6.7	Coordinate a community effort on model validation and verification involving GSD with NCEP, DTC, MDL and others	Jeff Hamilton
D4.2	Begin incorporating social science perspectives and knowledge into decision support activities to help realize Grand Challenge goals	Tanya Peevey
B6.10	Identify, track and embrace broader metrics of GSD's success even if those metrics are outside of GSD's direct or sole influence, with particular focus on measures of key stakeholder outcomes	Dave Turner

NOAA Global Systems Laboratory

Verification for Model Development

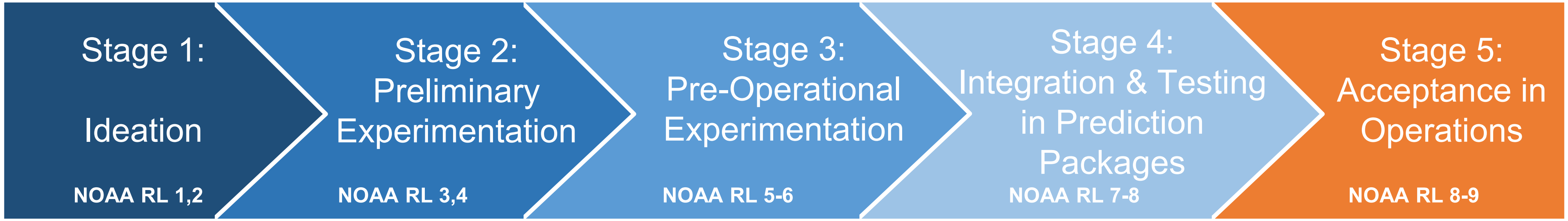
Jeff Hamilton
Associate Scientist, Verification and Assessment Branch

Global Systems Laboratory

Verification for Model Development

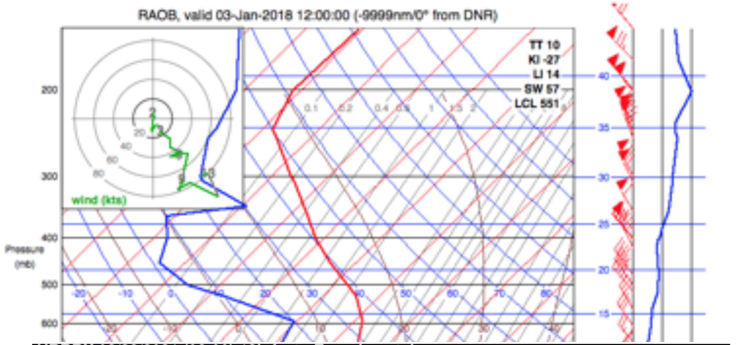
- Requires both subjective and objective verification
- Stages in the research-to-operations (R2O) process
- Typical model development cycle includes
  - "Sandbox" tests; e.g., evaluating change over particular cases
  - Retrospective runs; e.g., multi-week runs to ensure change has no ill effects
  - Experimental runs; e.g., realtime runs to evaluate impacts
  - Operational implementation at National Centers for Environmental Prediction
- Developers need objective verification at each stage

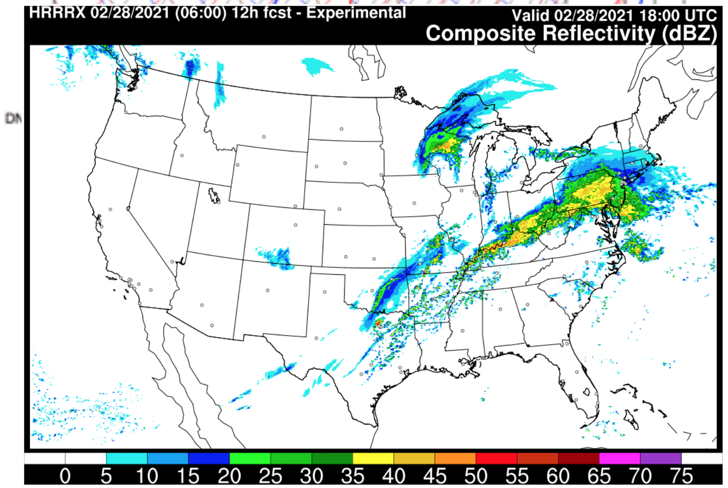
NOAA's R2O Pathway

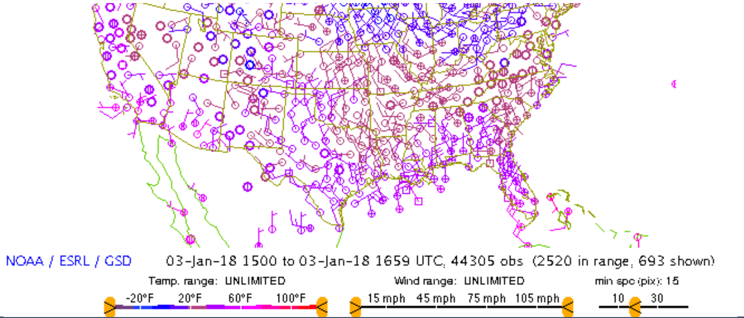


Verification System – Requirements

- Hypothesis driven
- Requirements on the GSL model verification tools
  - Quickly compare experimental model and a reference (operational) model in a user-driven web interface
  - Display statistics with a diverse set of plotting options
  - Evaluate against a wide variety of observations
    - Radar reflectivity, precip, surface stations, radiosondes, aircraft, more
    - Point and gridded data; both continuous and discrete (categorial) data

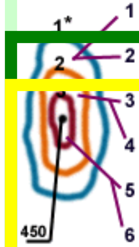






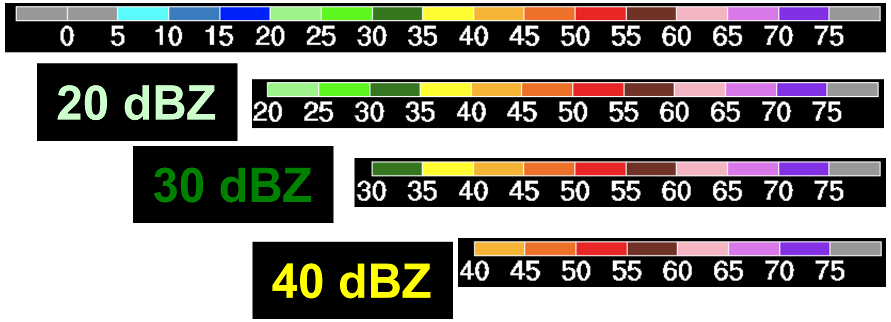
Verification System – Reflectivity Example

Intensity Threshold

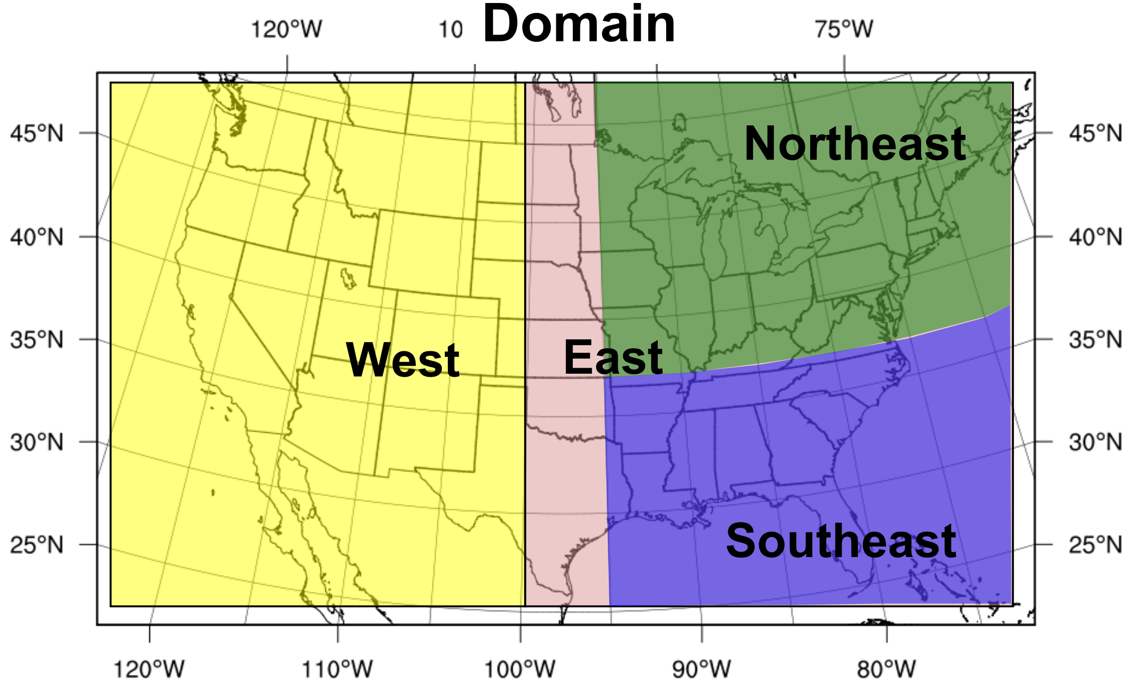


VIP** Level	Echo Intensity	Precipitation Intensity	Rainfall Rate (in/hr) Stratiform	Rainfall Rate (in/hr) Convective	Reflectivity in (dBZ)
1	Weak	Light	Less Than 0.1	Less Than 0.2	Min. Signal - 30
2	Moderate	Moderate	0.1 - 0.5	0.2 - 1.1	31 - 40
3	Strong	Heavy	0.5 - 1.0	1.1 - 2.2	41 - 45
4	Very Strong	Very Heavy	1.0 - 2.0	2.2 - 4.5	46 - 50
5	Intense	Intense	2.0 - 5.0	4.5 - 7.1	51 - 57
6	Extreme	Extreme	More Than 5.0	More Than 7.1	> 57

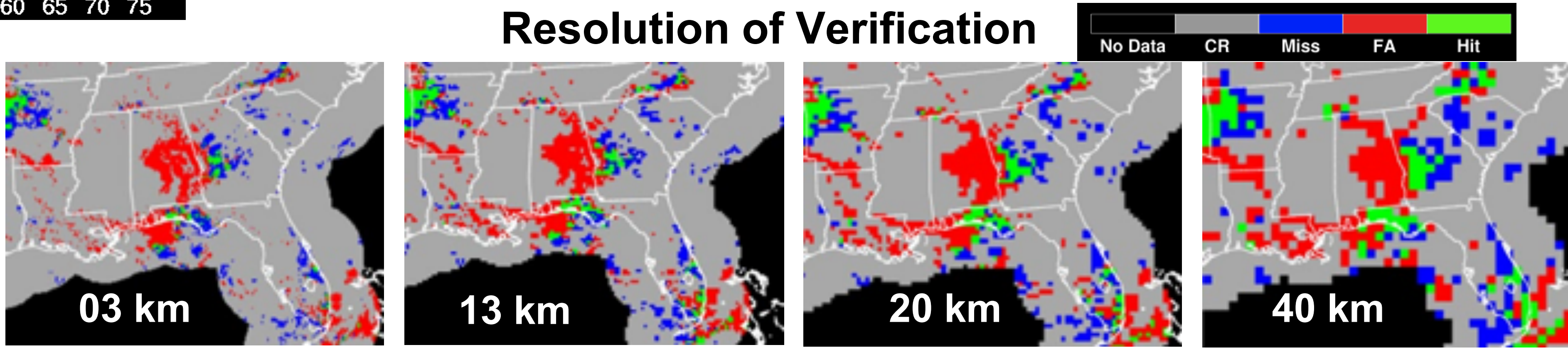
Highest precipitation top in area in hundreds of feet MSL (45,000 feet MSL).



Domain



Resolution of Verification

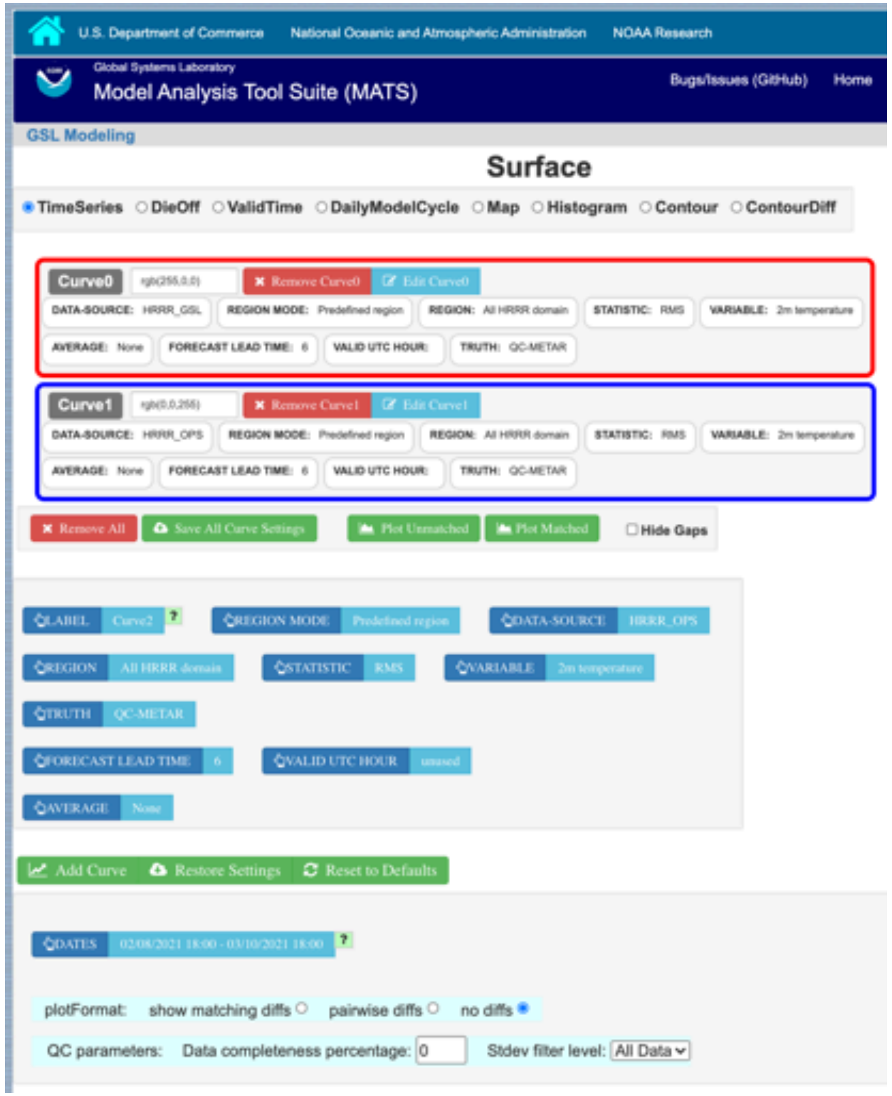


Verification System Details – 1

- Database-driven architecture
- Metadata is the main driver
- Automated workflow extracts NWP output and observations, matches them, computes aggregated sums of the matched pairs, and stores in database
- User interface provides access to the DB and creates plots desired by the developer

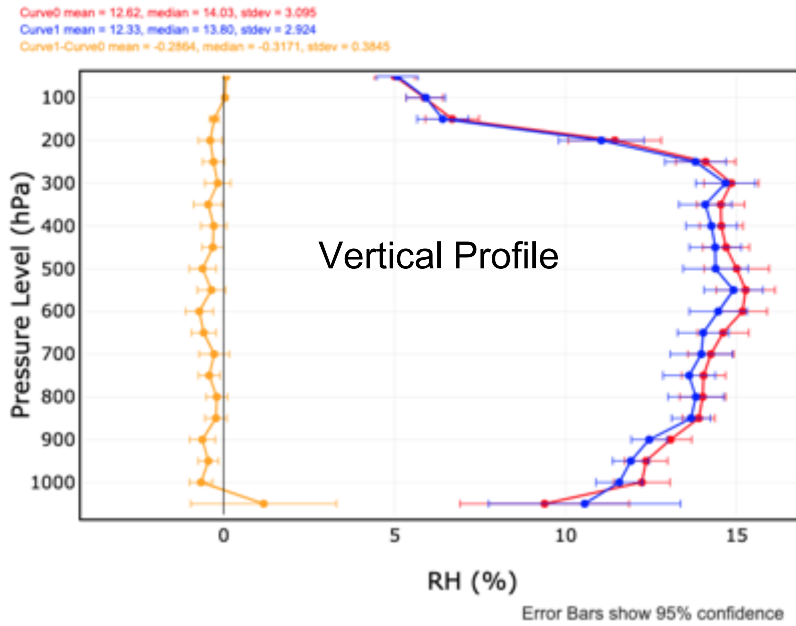
Verification System Details – 2

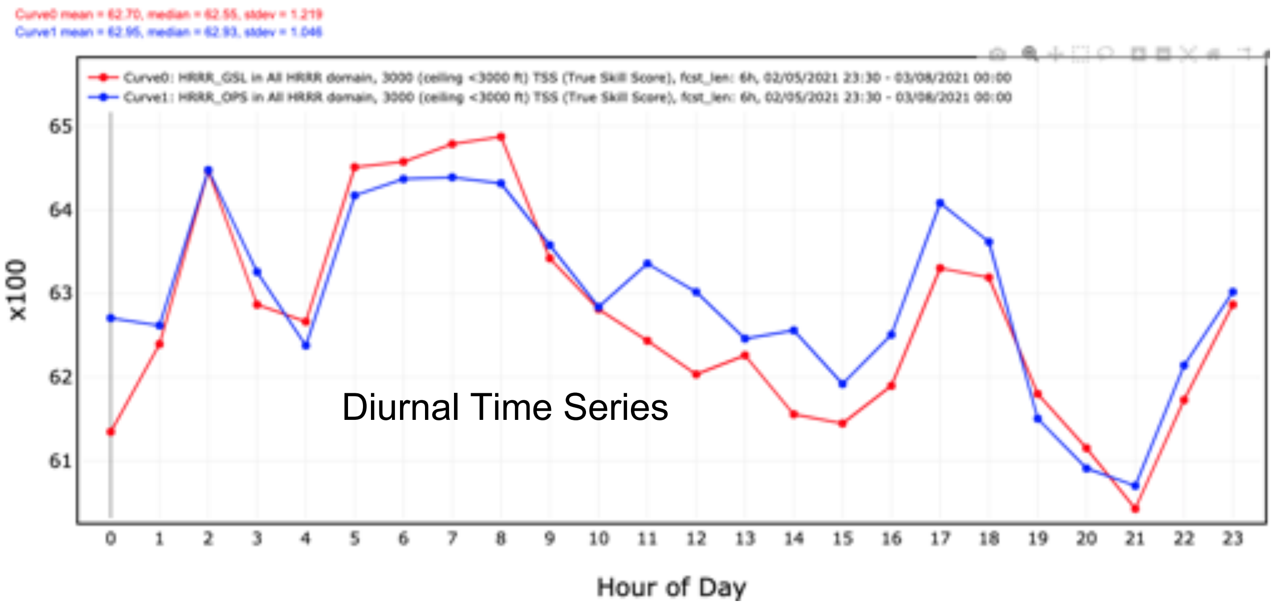
- Curve options dynamically generated
- Multitude of plot types (e.g., time-series, dieoff, etc)
- Multiple data sources per plot
- Numerous options for the user (geophysical variable, statistic, forecast time, date range, time averaging, etc.)
- Match the cases or treat independently
- Trivial to change to other plot types
- For comprehensive overview of GSL system, see Turner et al. 2020
- Collaboration with DTC's METplus team lead to METexpress (2015 Recommendation B6.7)

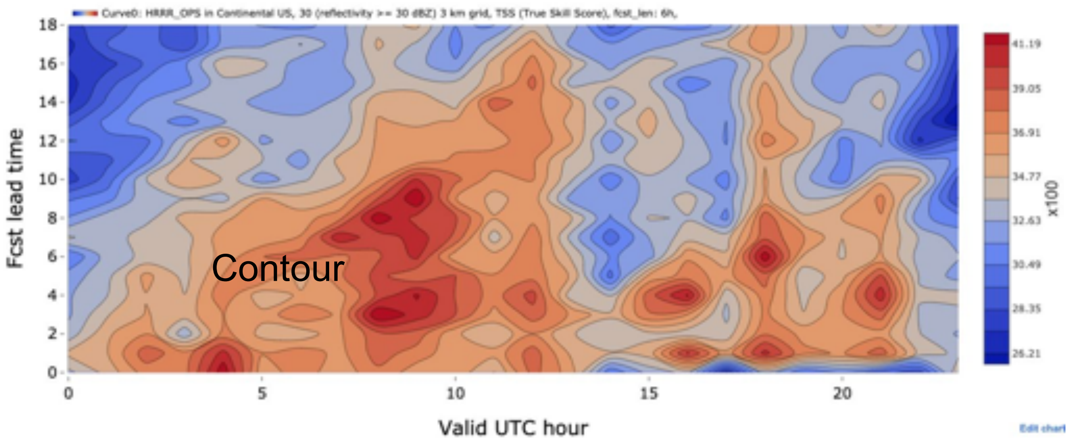


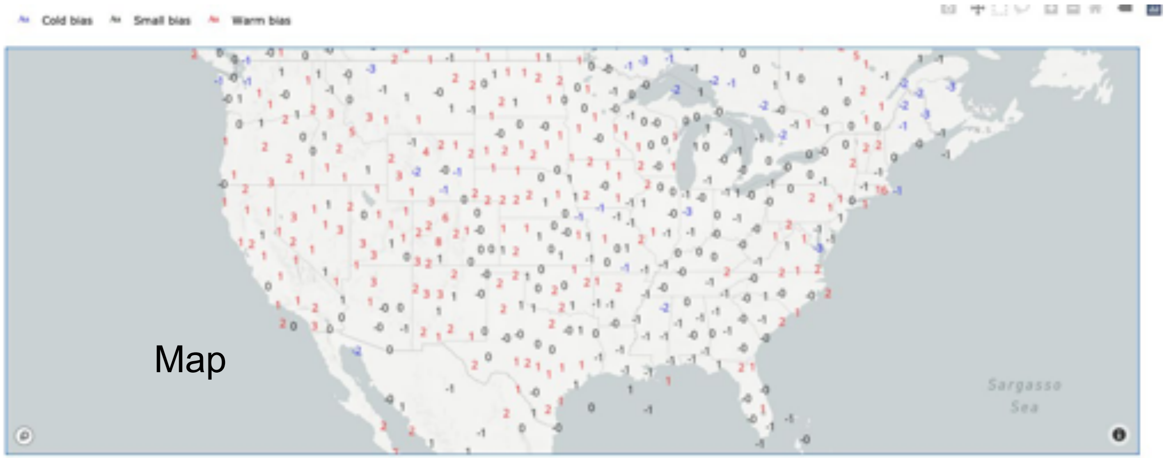
Verification System – GUI Example

Verification System – Plot Types examples



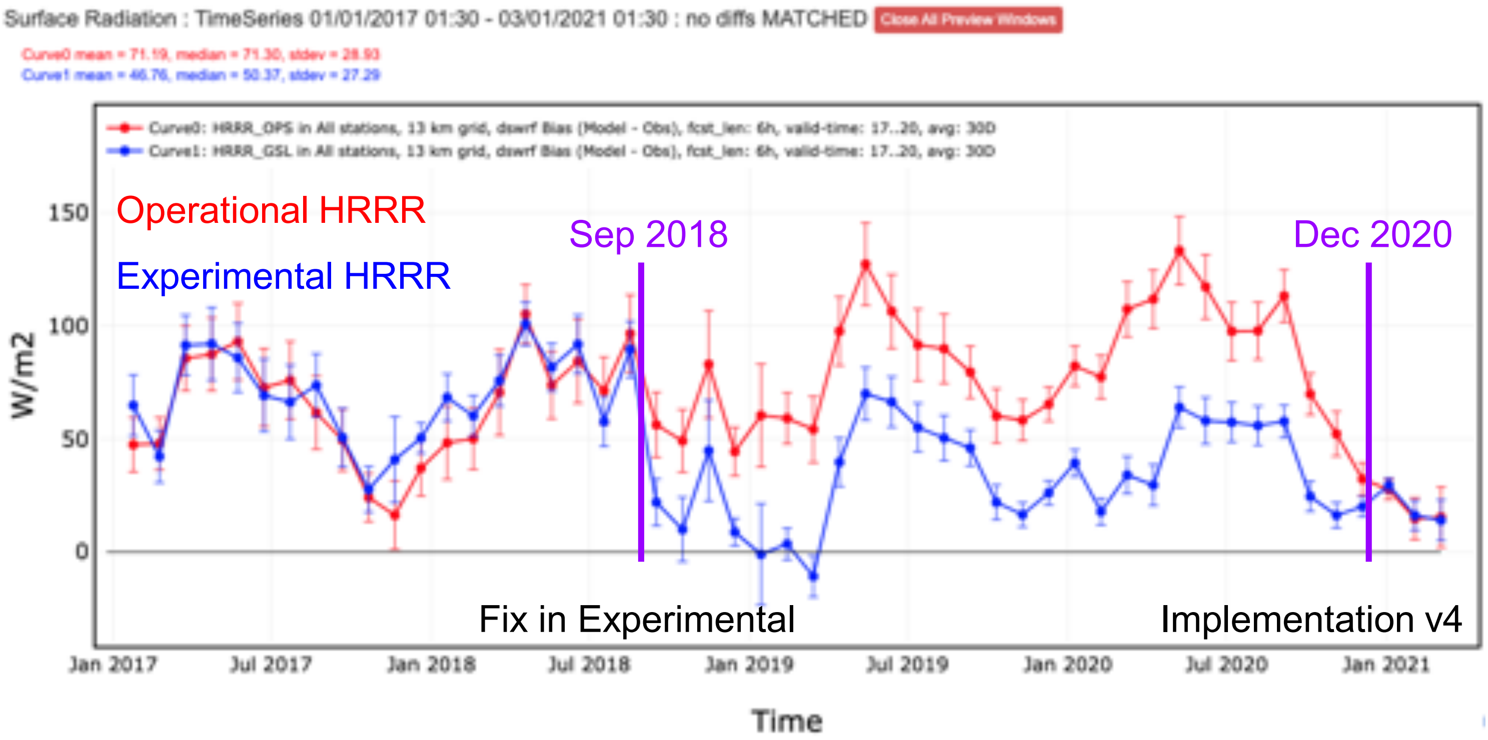




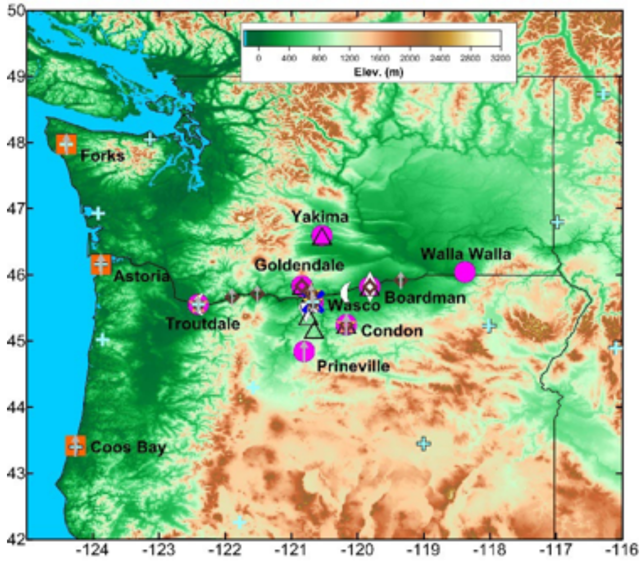


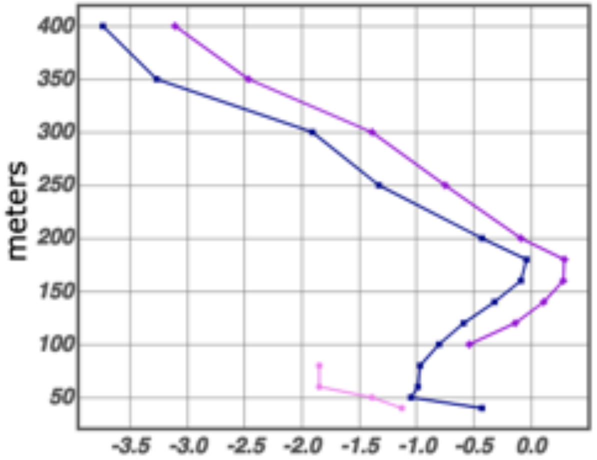
Verification System – R2O Example

- Issue with how clouds were treated at the sub-grid scale, below the scale at which the High-Resolution Rapid Refresh (HRRR) can explicitly resolve them
- Modelers utilized internal solar radiation verification at all stages of development
  - Diagnosis
  - Testing
  - Long-term viability
  - Operational implementation



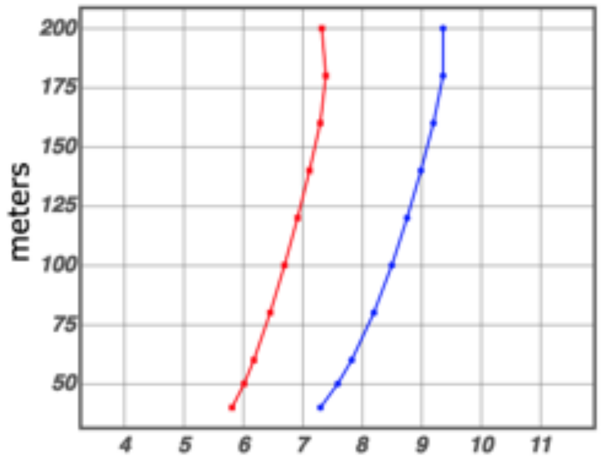
- Aggregated statistics over regions required to keep database size reasonable with rapid response times
- However, these objective metrics are hard to deconvolve to determine what atmospheric (physical) process(es) are the source of model forecast errors
- Process-Oriented Verification is a possible solution - WFIP-2 prototype





Vertical Profile of model bias when compared to Sodar (pink), Lidar (blue), and Profiler (purple) networks

List of Discriminators
Surface Bulk Richardson Number
Surface Heat Flux
Latent Heat Flux
Stability Parameter
Friction Velocity
Max Cloud Fraction
10-m Wind Speed
10-m Wind Direction
80-m Wind Speed
80-m Wind Direction
PBL Height
Wind Shear (200-10m)
Wind Shear (50-10m)
Wind Shear (200-50m)
Vertical Temperature Difference (200m-surface)
Downward Shortwave Radiation



Vertical Profile of model wind speed filtered by downward solar radiation (red) and unfiltered (blue)

Verification System – Process Oriented Verification 2

Process-oriented verification places stringent requirements on the Vx system

- Allow the users to select the filtering of the data to help isolate the physical process being investigated via the GUI
- Store orders of magnitude more data - scalable
- Store other variables to use as “discriminators” to filter the dataset
- More efficient database approach to be performant
- Currently prototyping new data model within a No-SQL system

Example MySQL table

sta_id	fcst_len	time	ndiff	press	temp	dp	wd	ws	rh	vgtyp
77	0	1600905660	0	9030	938	382	271	14	145	7
208	0	1600905840	0	9679	696	608	130	4	736	14
230	0	1600905600	0	10051	759	702	349	7	828	8
302	0	1600905960	3	10109	795	692	354	12	710	12
394	0	1600905600	0	9642	738	623	84	6	676	10
411	0	1600905660	0	9988	666	617	60	5	845	8
422	0	1600905660	0	9990	604	577	151	9	908	1
438	0	1600905720	0	9757	724	617	85	7	692	10
552	0	1600905600	0	10051	643	613	133	4	901	12
558	0	1600905960	0	10032	620	603	97	5	943	12

Example Couchbase document

DD:V01:METAR:HRRR_OPS:1608163200:1

Data Metadata


```
{
  "dataSourceId": "HRRR_OPS",
  "docType": "model",
  "model": "HRRR_OPS",
  "fcstValidBeg": "2020-12-17T00:00:00Z",
  "fcstValidEpoch": 1608163200,
  "subset": "METAR",
  "type": "DD",
  "version": "V01",
  "data": {
    "CWAE": {
      "Ceiling": 485,
      "DewPoint": 293,
      "RH": 845,
      "Reported Time": 1608163200,
      "Surface Pressure": 8603,
      "Temperature": 320,
      "VGTYP": 5,
      "Visibility": 1056,
      "WD": 2,
      "WS": 5,
      "name": "CWAE"
    },
```

Verification System – Future Efforts

Collaboration with the Unified Forecast System Effort*

- Requires a unified, distributed verification system
- The DTC's "METplus" software package has been adopted for this purpose
- GSL is incorporating our "lessons learned" into the METplus framework

** 2015 Recommendation B6.7*

Advanced and new observation systems

- Satellites, ground-based remote sensors, chemical and aerosol obs, etc
- Machine learning approaches (e.g., "feature-based verification")
- Object-oriented approach (automated subjective verification)

Cloud resources

- Shared data, storage, and web interface for collaboration (e.g. METexpress)

NOAA Global Systems Laboratory

Impact-Based Verification

Tanya Peevey
Research Scientist, Forecast Impact and Quality Assessment Services Branch

Global Systems Laboratory

Impact-Based Verification Overview

- Aim is to evaluate forecast products in the context of how they are used
  - Who is using the forecast?
  - What decisions do they need to make?
- Verification must consider the different decisions being made by a multiplicity of users

Aviation Weather Research Program (AWRP)

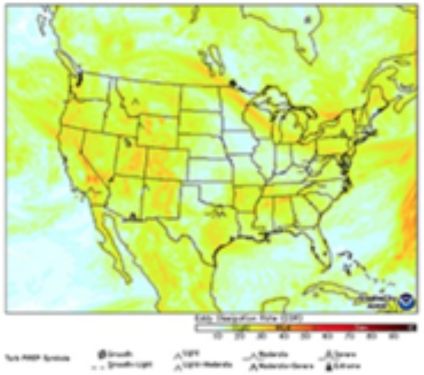
FAA's AWRP: "The goal of the research is to transition new or improved weather capabilities…[to] enhance aviation safety and efficiency."

GSL provides third-party, independent, evaluations of FAA-funded weather products to inform R2O decisions.

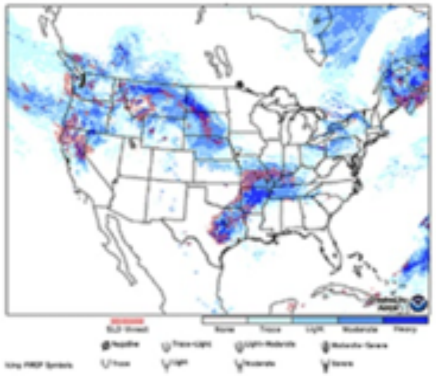
Assessments drove decisions to...

- Select the Localized Aviation MOS Product (LAMP) to provide ceiling and visibility data for the Helicopter Medical Emergency Services (HEMS) Tool.
- Delay implementation of the Offshore Precipitation Capability (OPC) to improve product performance.

Graphical Turbulence Guidance

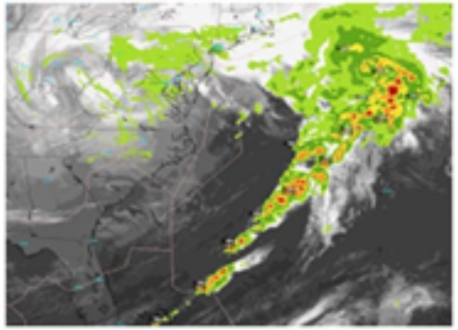


Forecast Icing Product



Helicopter Emergency Medical Services (HEMS) Tool

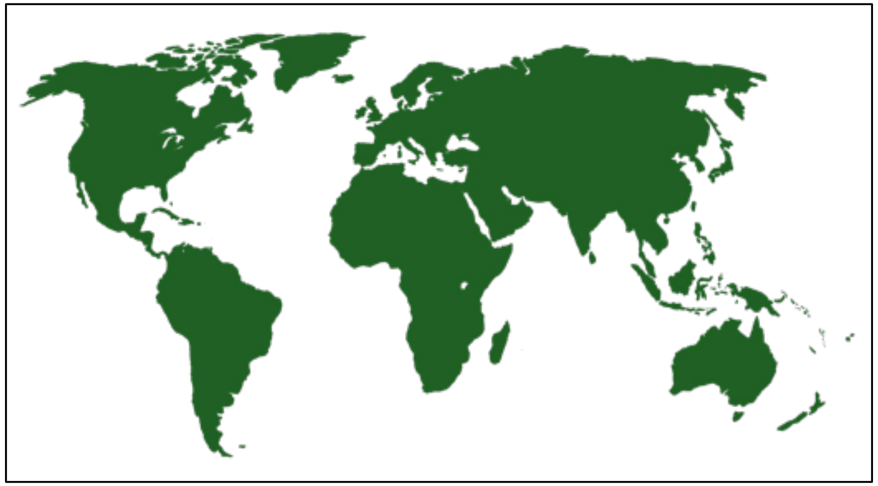
Offshore Precipitation Capability



Impact-Based Verification – Accomplishments

In the <u>last 5 years</u>, we have conducted numerous AWRP assessments

Users include commercial aviation, general aviation, helicopters & drones

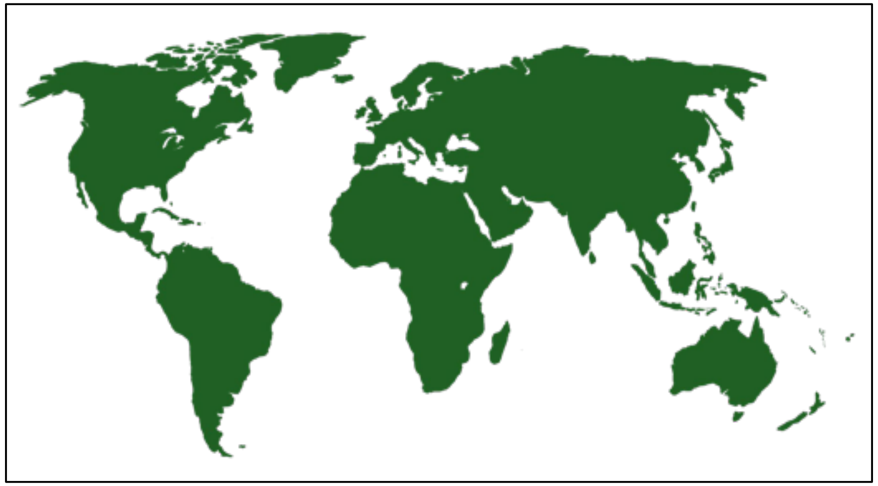


Impact-Based Verification – Accomplishments

In the last 5 years, we have conducted numerous AWRP assessments

Users include commercial aviation, general aviation, helicopters & drones

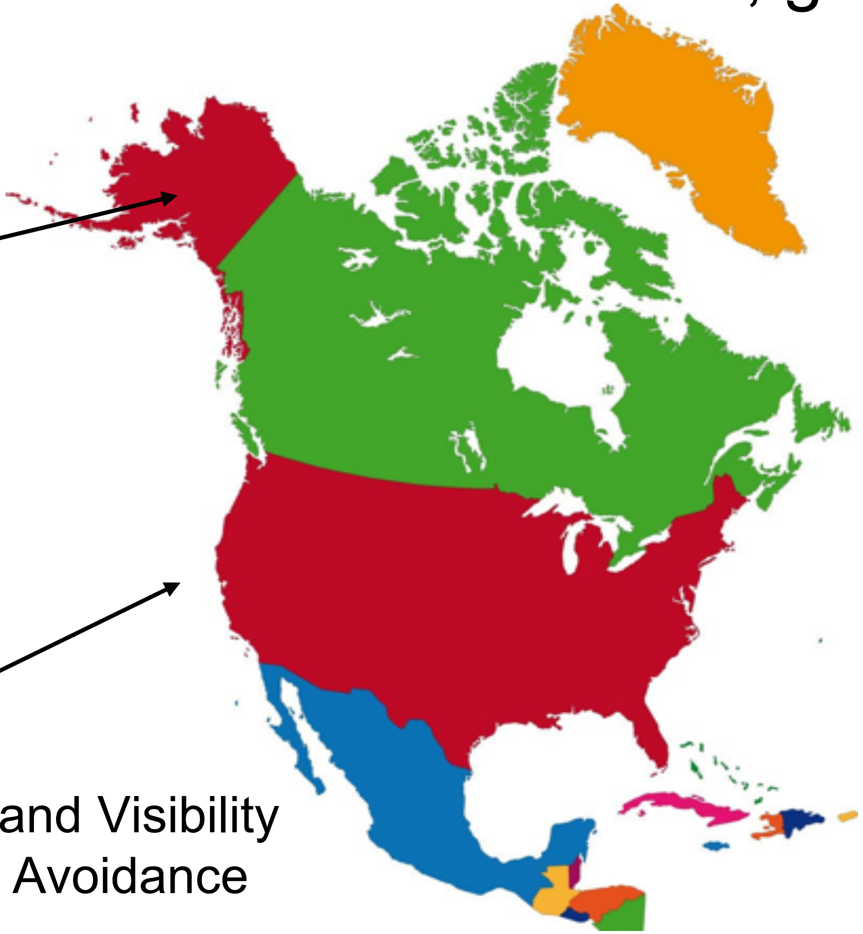
Alaska
In-flight Icing
Camera Visibility



Impact-Based Verification – Accomplishments

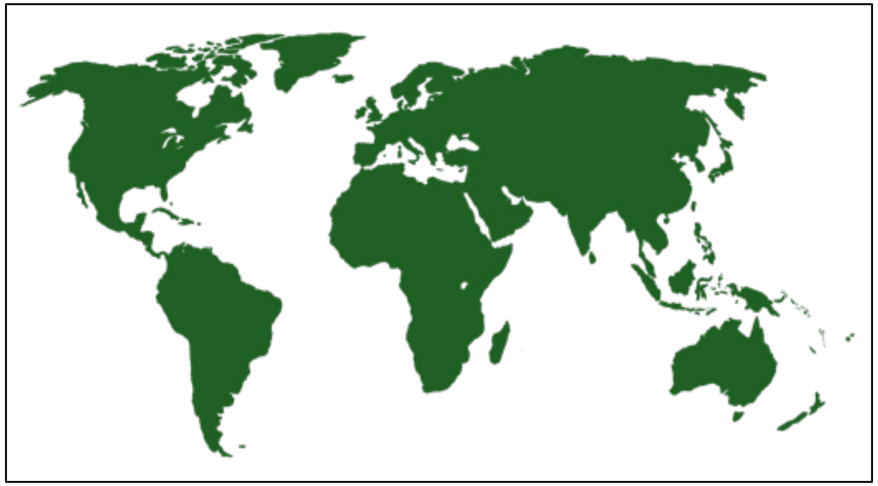
In the last 5 years, we have conducted numerous AWRP assessments

Users include commercial aviation, general aviation, helicopters & drones



Alaska
In-flight Icing
Camera Visibility

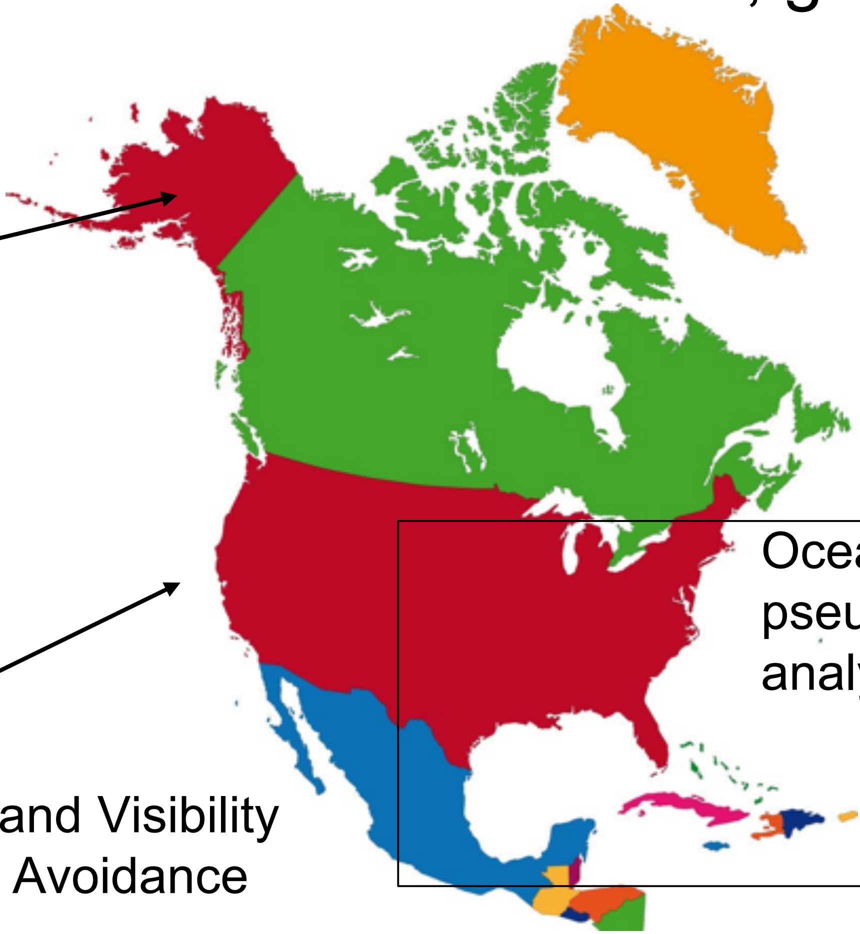
CONUS
RAP upgrade
HRRR upgrade
HEMS Tool: Ceiling and Visibility
Convective Weather Avoidance Model



Impact-Based Verification – Accomplishments

In the last 5 years, we have conducted numerous AWRP assessments

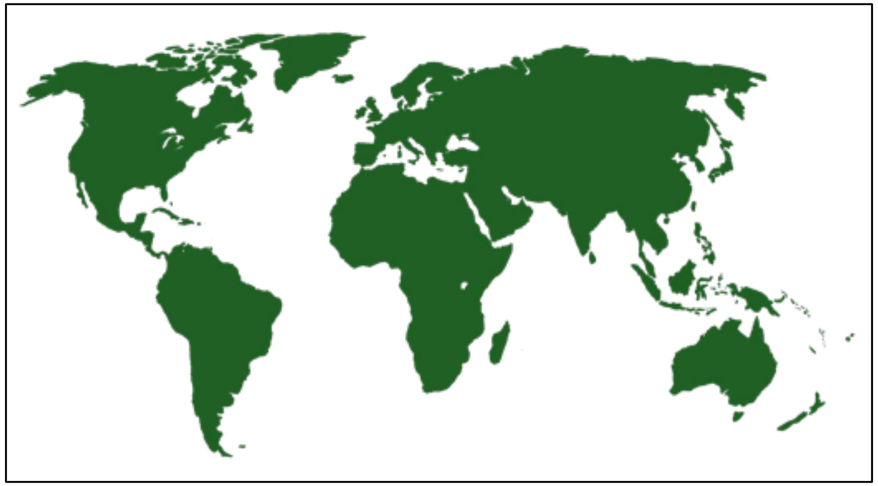
Users include commercial aviation, general aviation, helicopters & drones



Alaska
In-flight Icing
Camera Visibility

CONUS
RAP upgrade
HRRR upgrade
HEMS Tool: Ceiling and Visibility
Convective Weather Avoidance Model

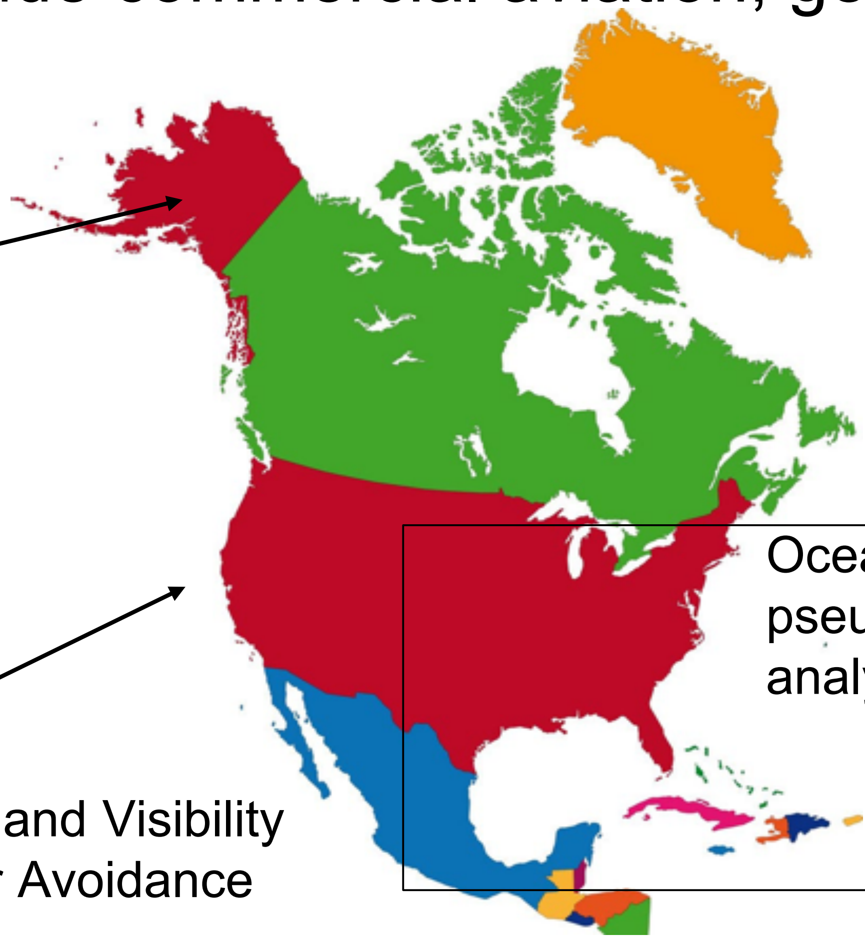
Oceanic pseudo-radar analysis



Impact-Based Verification – Accomplishments

In the last 5 years, we have conducted numerous AWRP assessments

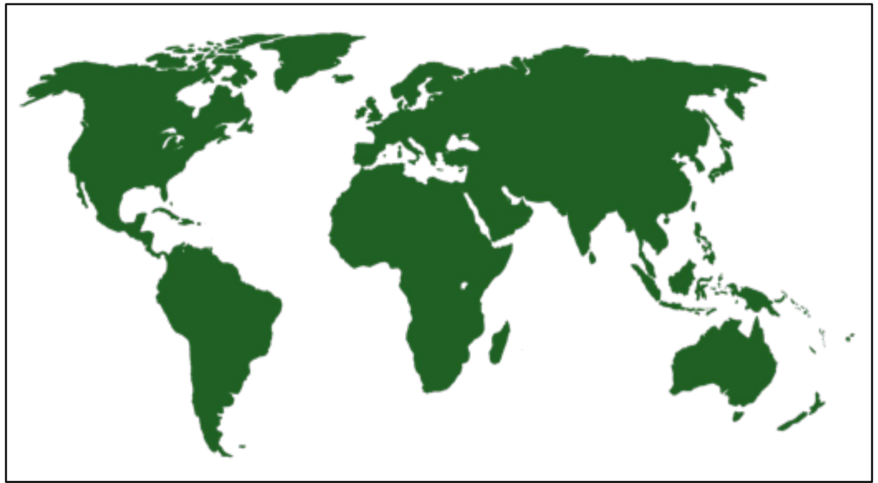
Users include commercial aviation, general aviation, helicopters & drones



Alaska
In-flight Icing
Camera Visibility

CONUS
RAP upgrade
HRRR upgrade
HEMS Tool: Ceiling and Visibility
Convective Weather Avoidance Model

Oceanic pseudo-radar analysis



World Area Forecast System: Global turbulence and convective forecasts

Assessment Components

Investigate new observations and verification techniques

Assessment Components

Discussions with FAA and product developers

Assessment Components

Incorporate ideas from Core R&D; Adjust verification to specifics of user needs

Assessment Components

Share details of approach; Adjust based on stakeholder feedback, where appropriate

Assessment Components

Develop code; Adapt verification tools

Assessment Components

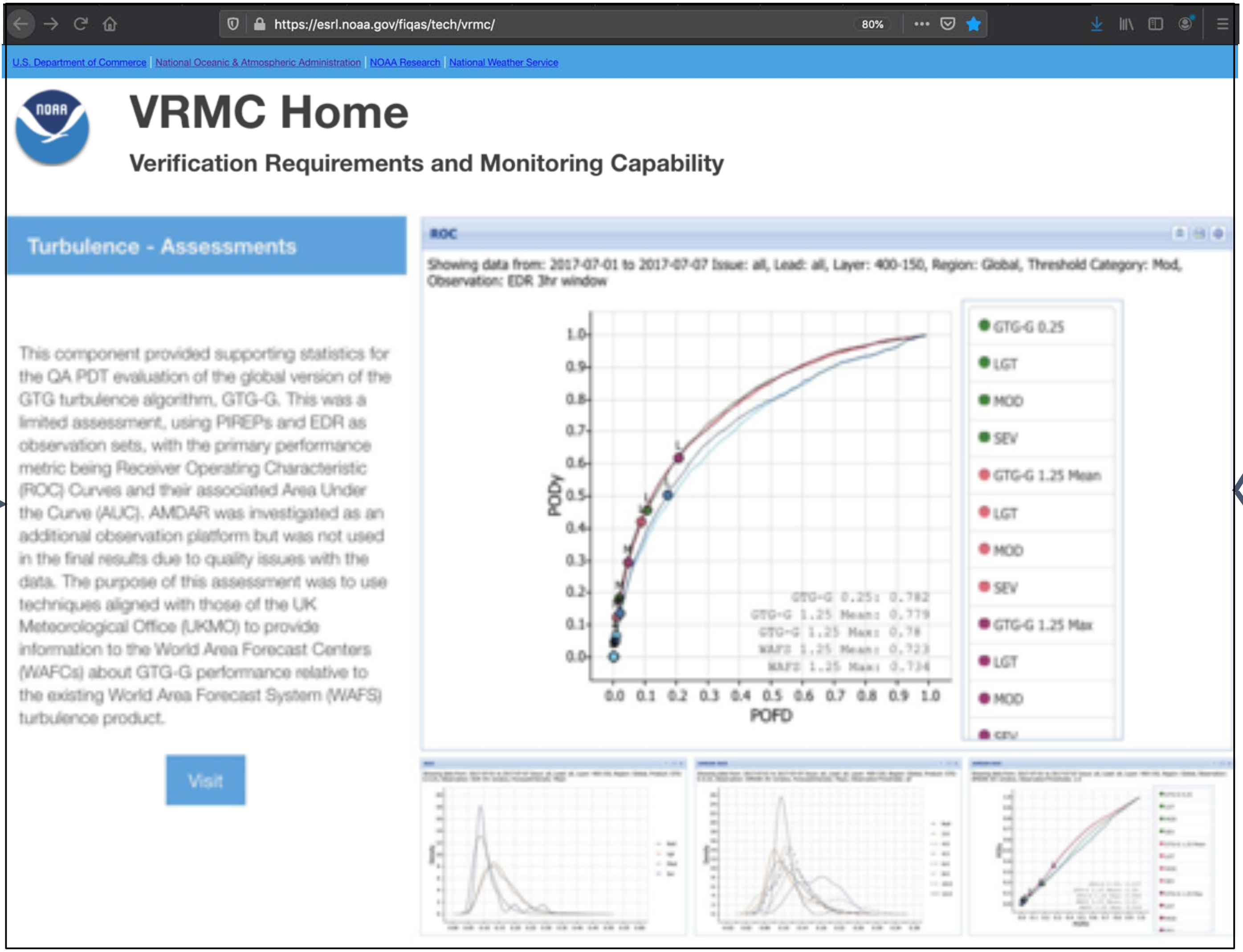
Iterative between analyst and engineers

Assessment Components

Present findings and submit written report

Verification Tools – FAA

Assessment Concepts



Assessment Support

Monitoring

Verification Tools – NWS

CWVS

Convective Weather Verification Service

Impact-based verification of TCF

TFVT

TRACON Gate Forecast Verification Tool

TRACON Gate forecasts for convection at the terminal

CBVT

CWSU Briefing and Verification Tool

Forecaster briefings to TMU for wind shift, C&V events

EVENT

Event-based Verification and Evaluation of NWS Gridded Products Tool

Gridded products for thunderstorm events, terminal and en-route

Impact-Based Verification – Example

- Use of probabilistic forecast information for decisions on snow removal operations at Denver International Airport
- Work done in collaboration with the Boulder Weather Forecast Office and social scientists in Weather Risks and Decisions in Society (WRaDS) program at NCAR
- *2015 recommendation D4.2

Impact-Based Verification – Example

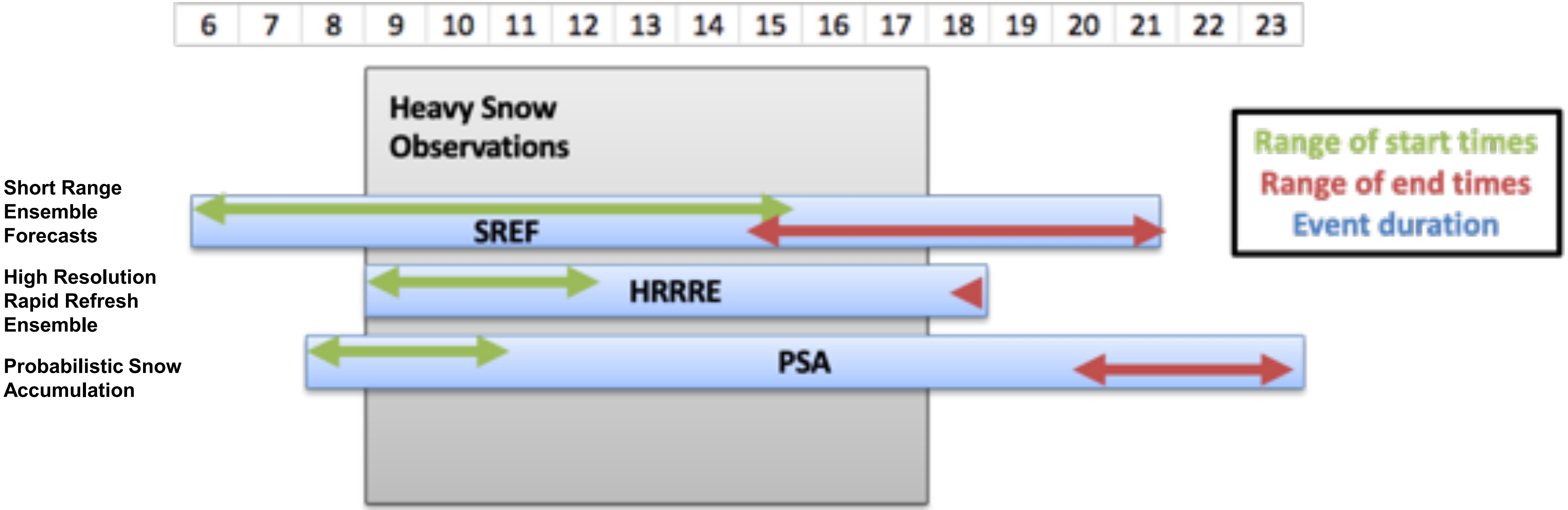
From interviews with the ground operations staff at Denver Int’l Airport

- Interested in event...
  - onset--when do we need to get emergency staff to the airport
  - duration--how long will they be here (do we need to provide food, cots, etc.)
  - intensity--how many staff do we need to bring in
- Need advance planning, not short-term forecasts
  - Forecast information useful in advance of event
  - During the event, rely primarily on observations
- Misses and underforecasts are worse than false alarms and overforecasts

Impact-Based Verification – Example

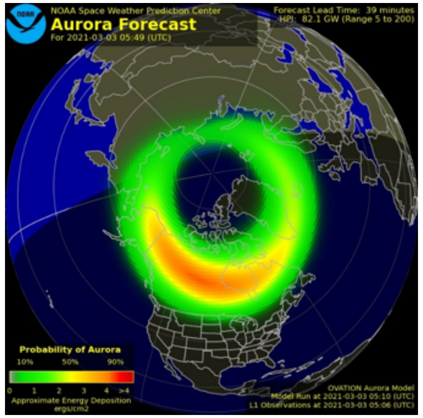
Present results in context of decisions--onset and cessation of heavy snow

Answers … how does the forecast perform in the context of their decisions and how can they make best use of the product?



Impact-Based Verification – Future Work

- Aviation products switch from (13-km) RAP-based to (3-km) HRRR/RRFS-based
- Probabilistic forecasts of global turbulence and icing
- Fire weather/smoke forecasts
- Verification to inform Impact-based Decision Support Services (IDSS)
- Boundary layer forecasts for unmanned aerial systems (UAS) operations
- Boundary layer forecasts for power generation and transmission
- Space weather in the context of aviation



NOAA Global Systems Laboratory

Economic Impact Studies

Dave Turner
Chief, Verification and Assessment Branch

Economic Impacts (EI) – Overview

- Developing NWP models is expensive; is the Nation getting a good return on its investment?
- EI from some tool can only occur if a decision is changed based upon that tool (e.g., the new forecast changes behavior)
- EI is only important for certain weather conditions
- EI is usually regionally dependent
- Working with CSU Economics Dept. to evaluate EI of the regular updates to our regional model HRRR
- Estimating EI of a decision is seldom easy

Yellow: Major metropolitan areas
Black/blue: Precipitation

Economic Impacts – Our Study

- CSU has developed Computable General Equilibrium models to translate weather-based decisions into financial impacts
- Focused economic analysis on three areas:
  - Morning commuting
  - Specialty crop agriculture
  - Wind energy production
- Financial impacts are rarely symmetric
  - Costs associated with "doing" usually quite different than for "not doing"

Turner et al., *BAMS*, under review

Economic Impacts – Analysis Periods

- Need to use the same weather events for a fair comparison between the baseline and new models
- Focused on overlap periods between (v1 & v2) and (v2 & v3)

Version	Primary update from previous version	Overlap period
V1 → V2	First inclusion of subgrid-scale clouds, aerosol particles included in cloud and precipitation processes, full-cycling of the land surface model	1 Jun 2015 to 1 Aug 2016
V2 → V3	Updated turbulence scheme to use non-local mixing, more realistic treatment of subgrid-scale clouds, improved vertical coordinate for simulation above complex terrain, improved data assimilation approach to help retain stratiform clouds	1 July 2017 to 1 June 2018

Economic Impacts – Wind Energy Example

Only focused on “overprediction” errors → largest financial impact

	New ”better” than old		
	Electricity Generated [MW]	Extra costs [$M]	Potential Savings [$M]
“Actual”	579,260		
Based on HRRR1	3,328,406	75.6	$59.0M
Based on HRRR2	1,170,927	16.6	

14 month period

Economic Impacts – Wind Energy Example

Only focused on “overprediction” errors → largest financial impact

	New ”better” than old		
	Electricity Generated [MW]	Extra costs [$M]	Potential Savings [$M]
“Actual”	579,260		
Based on HRRR1	3,328,406	75.6	$59.0M
Based on HRRR2	1,170,927	16.6	
“Actual”	873,022		
Based on HRRR2	2,162,395	36.7	$17.4M
Based on HRRR3	1,547,034	19.3	

14 month period

11 month period

Economic Impacts – Continued Research

- Will evaluate the economic impact in other areas too
  - HRRR version 3 vs version 4 (big improvements in clouds and mixing)
  - Dynamic line rating
  - Aviation forecasts

- Gauging economic impact in disasters (e.g., floods, tornadoes, etc) extremely challenging due to low number of events and assumptions

- Quantifying overall economic impact of a new model version is huge lift (so many different sectors)

NOAA Global Systems Laboratory

Summary: Verification and Evaluation

Matt Wandishin, Tanya Peevey,
Jeff Hamilton, and Dave Turner

Global Systems Laboratory

Verification and Evaluation – Summary

- Developed range of tools for objective NWP model evaluation
  - Goal: to help the model developers improve the models more efficiently
  - Many different statistics over numerous variables and regions
  - Working to identify physical processes that aren't represented correctly
- Impact-based assessment
  - Goal: demonstrate improved functionality and accuracy of new models for specific, high-visibility stakeholders
  - Evaluate the forecast systems from the stakeholder's perspective
  - Primarily objective based, but subjective interpretation also
- Economic assessment
  - Using economic models to translate model improvements into savings

Verification and Evaluation – Summary

- Automated verification tools and metrics to quantify model improvements demonstrate *performance*
- Innovative techniques such as impact-based and process-oriented evaluations demonstrate *quality*
- Independent model assessments used by NWS and FAA as decision points for transitioning models and tools to operations within NWS demonstrate *relevance*
- Significant economic benefits in several sectors demonstrate *quality and relevance* of the HRRR model to society

Thank you!