Initial development of GSL's interactive verification scorecard tool



¹Cooperative Institute for Research in Environmental Sciences, Boulder, CO; ²NOAA Global Systems Division, Boulder, CO; ³Cooperative Institute for Research in the Atmosphere, Ft. Collins, CO

Background

- The Model Analysis Tool Suite (MATS) is a software product developed in-house by the National Oceanic and Atmospheric Administration (NOAA)'s Global Systems Laboratory (GSL) to assess the forecast skill of regional and global weather models.
- MATS consists of individual applications, which are each designed to verify a particular meteorological facet (e.g. ceiling heights, radar reflectivities, precipitation accumulations, etc.).
- Using MATS, model developers can produce interactive, publication-quality plots to quickly and easily compare the skill of individual models' forecasts as they occur, as well as assess the performance of models under development.
- In the past year, GSL added a new feature to MATS: interactive verification scorecards that provide a high-level overview of the relative performance of two modeling systems (e.g., the operational HRRR vs. the experimental RRFS).

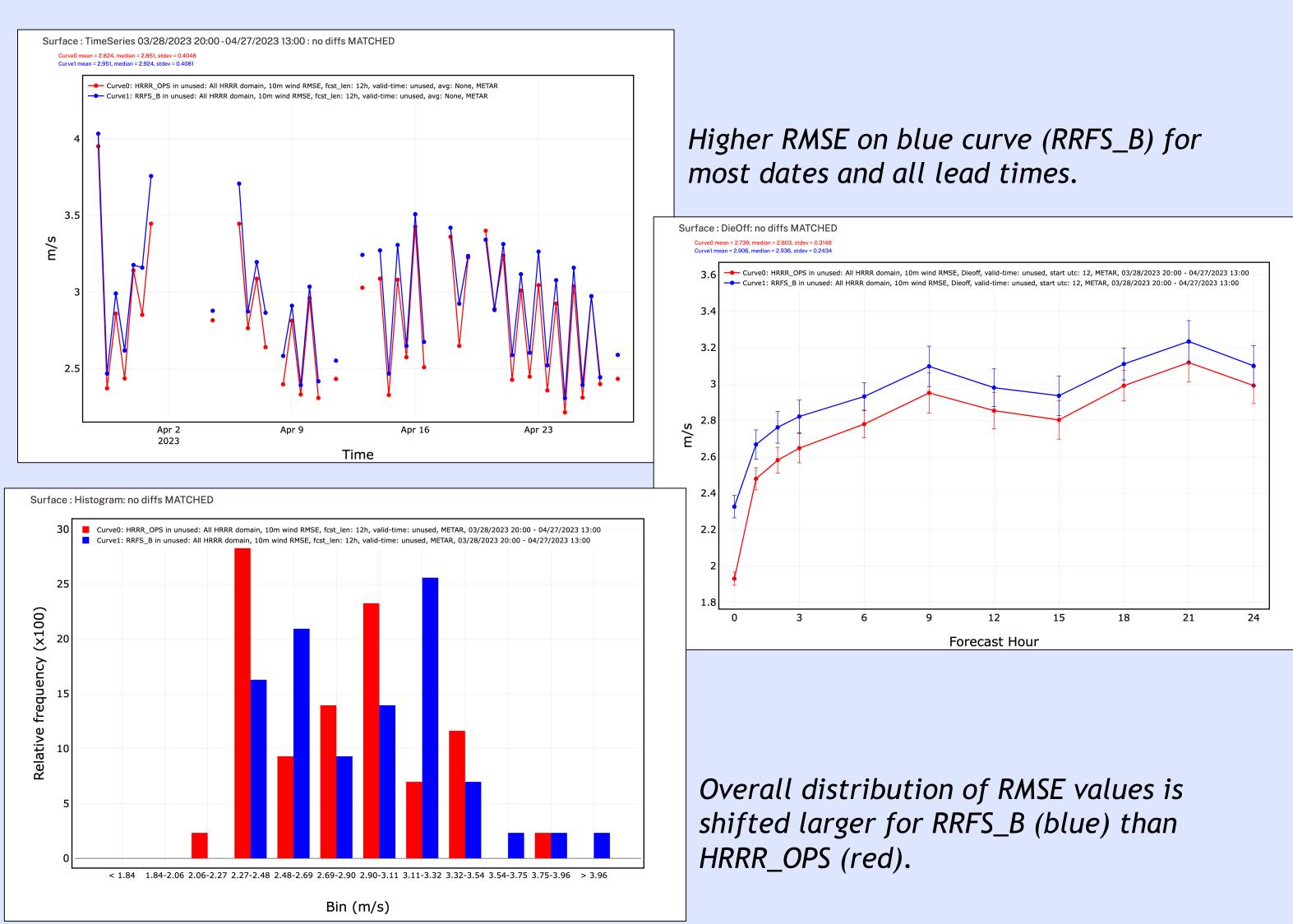
Scorecard Description

- In general, scorecards are made up of a matrix of cells, with each cell showing which model performs better for a set of given parameters. (e.g., time period, domain, variable, skill score, etc.), and whether that performance difference is statistically significant.
- MATS interactive scorecards consist of three components.
 - \succ The first is an interface that allows the user to select all parameters relevant to the model comparison.
 - \succ The second is a data processing service that allows for timely calculations of all the individual cells' model data.
 - \succ The third is a visualization application that displays the results in the scorecard, allowing the user to click on any of its cells.
 - When clicked, the data that was used to generate the score for that particular cell will be passed to the appropriate MATS application, thereby allowing the user to interactively plot the data using the range of MATS plot types, and understand the characteristics of the data.
- We envision that this interactive capability to visualize verification statistics in a number of different variables, forecast hours, and regions simultaneously will greatly speed up our understanding of model behavior and biases, which will improve the speed of model development.
- The scorecard is available online at apps-dev.gsd.esrl.noaa.gov/mats-dev/scorecard. Currently a GSL VPN is required, but we are working to make it publicly available by the end of the summer.

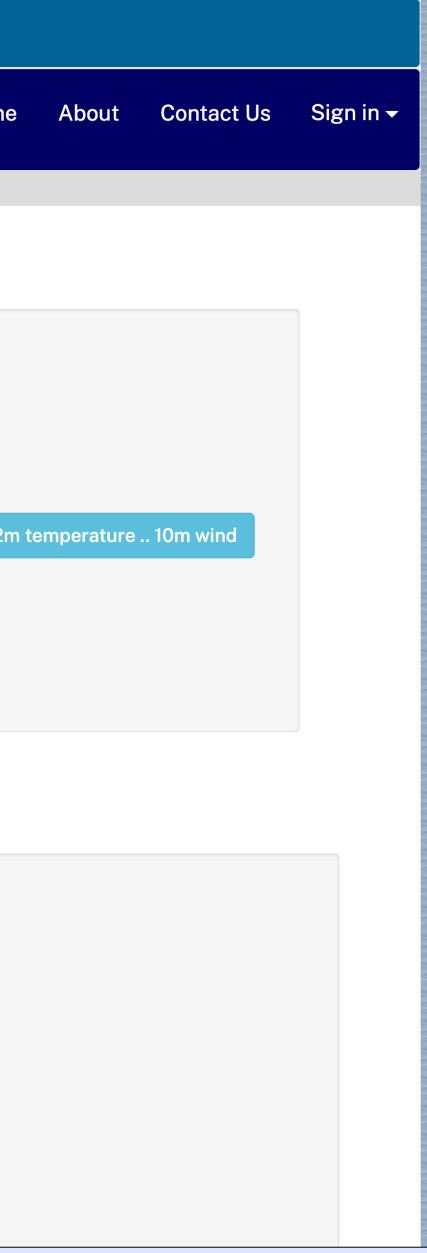
Molly Smith^{1,2}, Randy Pierce^{2,3}, Ian McGinnis^{2,3}, Gopakumar Padmanabhan^{1,2}, Jeff Hamilton^{1,2}, Bonny Strong^{2,3}, Dave Turner², and Matthew Wandishin²

Users select the two models they wish to compare, as well as the regions, statistics, variables, lead times, etc., over which the comparisons will be performed... -

	J.S. Department of Comm	nerce National (Oceanic and Atm	nospheric Admiı	nistration	NOAA Resea	arch
NORR	Global Systems Laboratory Model Analysi		(MATS)	E	Bugs/Issues	(GitHub)	Hon
GSL Mo	odeling						
			S	coreca	rd		
Label Experin		lication Surface FS_B Control	Data Source	HRRR_OPS			
Region	All HRRR domain We	stern HRRR domain	Statistic	RMSE Bias ((Model - Obs)	Variable	e 2
Truth	METAR						
Forecas	st Lead Time (h) 0 12	Valid UTC Ho	ur unused				
+ Add	Block <i>C</i> Reset to Defa	ults 🛛 若 Scorecard	ds Status				
Scor	recard schedule mod	e: Once 🔍 F	Recurring O				
One Ti	me Date Range 03/28/	2023 20:00 - 04/27/3	2023 13:00 ?				
Scor	recard confidence int	erval metric:	Percent 🔍 🤇	Standard Dev	viation O		
Minor Scor	% 95 Major % recard color theme:	99 RedGreen ©	RedBlue O	• •	•		



Once the graph is produced, users can take a closer look at the data that went into each cell on a scorecard in an interactive timeseries plot. They also have the option of switching to one of the other MATS plot types, such \blacktriangleleft as dieoff (forecast lead time on the x-axis), valid time (hour of day on the x-axis), contour plots, or histograms.



Scorecard Block0: Experimental Data Source = RRFS_B, Control Data Source = HRRR_OPS Constant fields: truth = METAR; valid-time = unused; dates = 03/28/2023 20:00 - 04/27/2023 13:00

Application: Surface		All HRRR domain	Eastern HRRR domain	Western HRRR domain	
stat	variable	0 0 0 0 0 0 1 0 1 2 3 6 9 2		0 0 0 0 0 0 1 0 1 2 3 6 9 2	
Bias	10m wind				
Bias	2m RH				
Bias	2m dewpoint				
Bias	2m temperature				
RMSE	10m wind				
RMSE	2m RH				
RMSE	2m dewpoint				
RMSE	2m temperature				

This particular scorecard that we've made compares surface variables, although users can also examine upper level fields, ceiling heights, radar reflectivities, and many other quantities. If we click on a cell, for example the one comparing the two models' wind RMSE over the entire HRRR domain at a 12 hour lead time, the scorecard will open the relevant MATS app (in this case Surface) and plot the data that went into the scorecard comparison in a pair of timeseries.

U.S. Department of Commerce National Oceanic and Atmospheric Global Systems Laboratory Model Analysis Tool Suite (MATS) GSL Modeling Plot Type: TimeSeries verage: None Forecast Lead Time (h): 12 Valid UTC Hour: unused Truth: verage: None Forecast Lead Time (h): 12 Valid UTC Hour: unused × Remove All 🛛 🚯 Save All Curve Setti 😙 Plot Unmatche Forecast Lead Time (h alid UTC Hour Average Non 🕂 Add Curve 🛛 🕹 Restore Settings 🛛 🤁 Reset to Defaults 28/2023 20:00-04/27/2023 13:00



And a scorecard is produced! In this example, we are comparing the experimental RRFS_B against the operational HRRR.

ministration	NOAA Research					
		Bugs/Issues (GitHub)	Home	About	Contact Us	Sign in 🗸
Sur	face					
		1				
Statistic: R	MSE Variable: 10m wind					
R		J				
Statistic: RMS	E Variable: 10m wind					
R						
ot Matched	🗆 Hide Gaps					
RRFS_B						