

Background

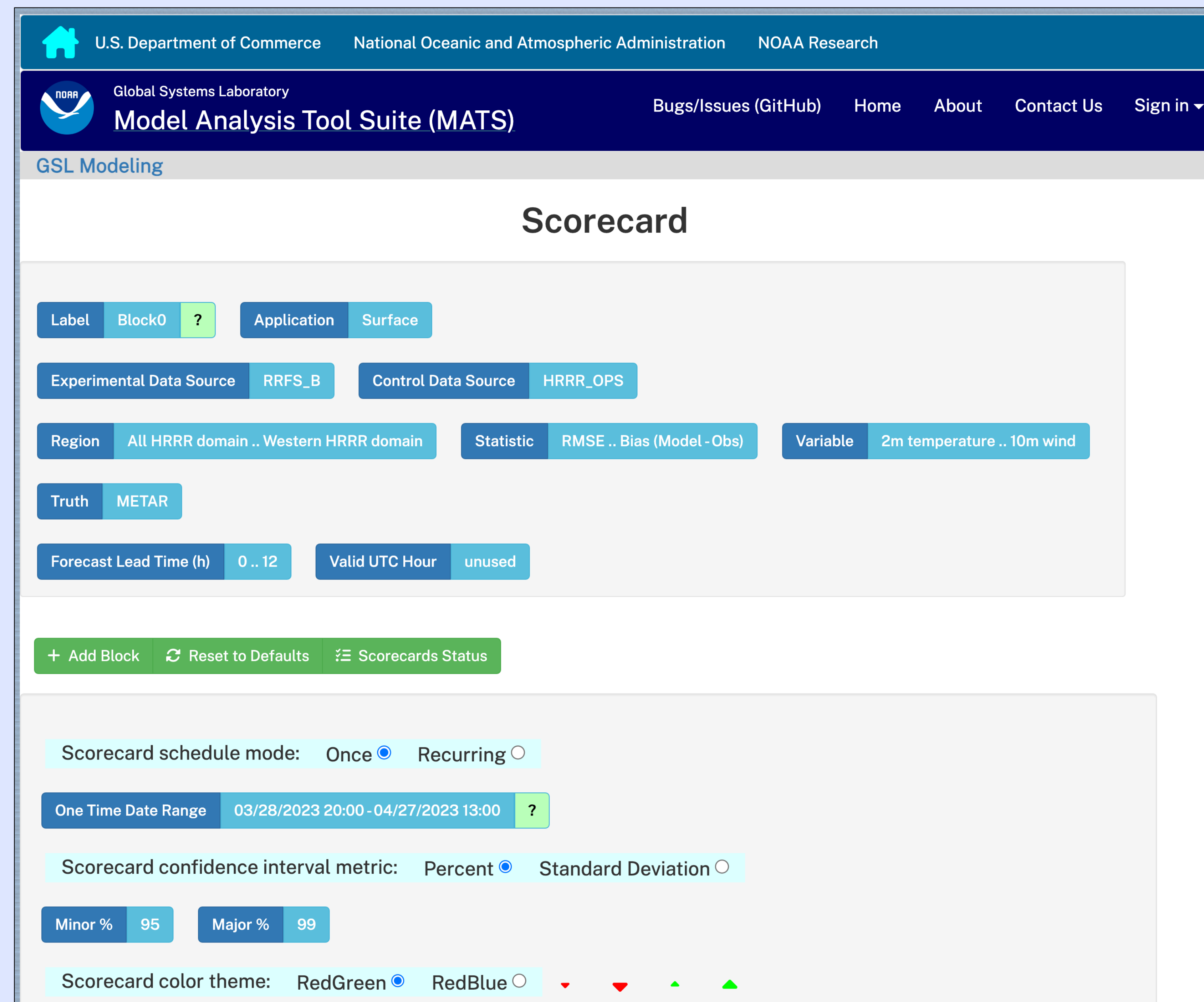
- The Model Analysis Tool Suite (MATS) is a software product developed in-house by the National Oceanic and Atmospheric Administration (NOAA)'s Global Systems Laboratory (GSL) to assess the forecast skill of regional and global weather models.
- MATS consists of individual applications, which are each 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.
 - The first is an interface that allows the user to select all parameters relevant to the model comparison.
 - The second is a data processing service that allows for timely calculations of all the individual cells' model data.
 - 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.

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...

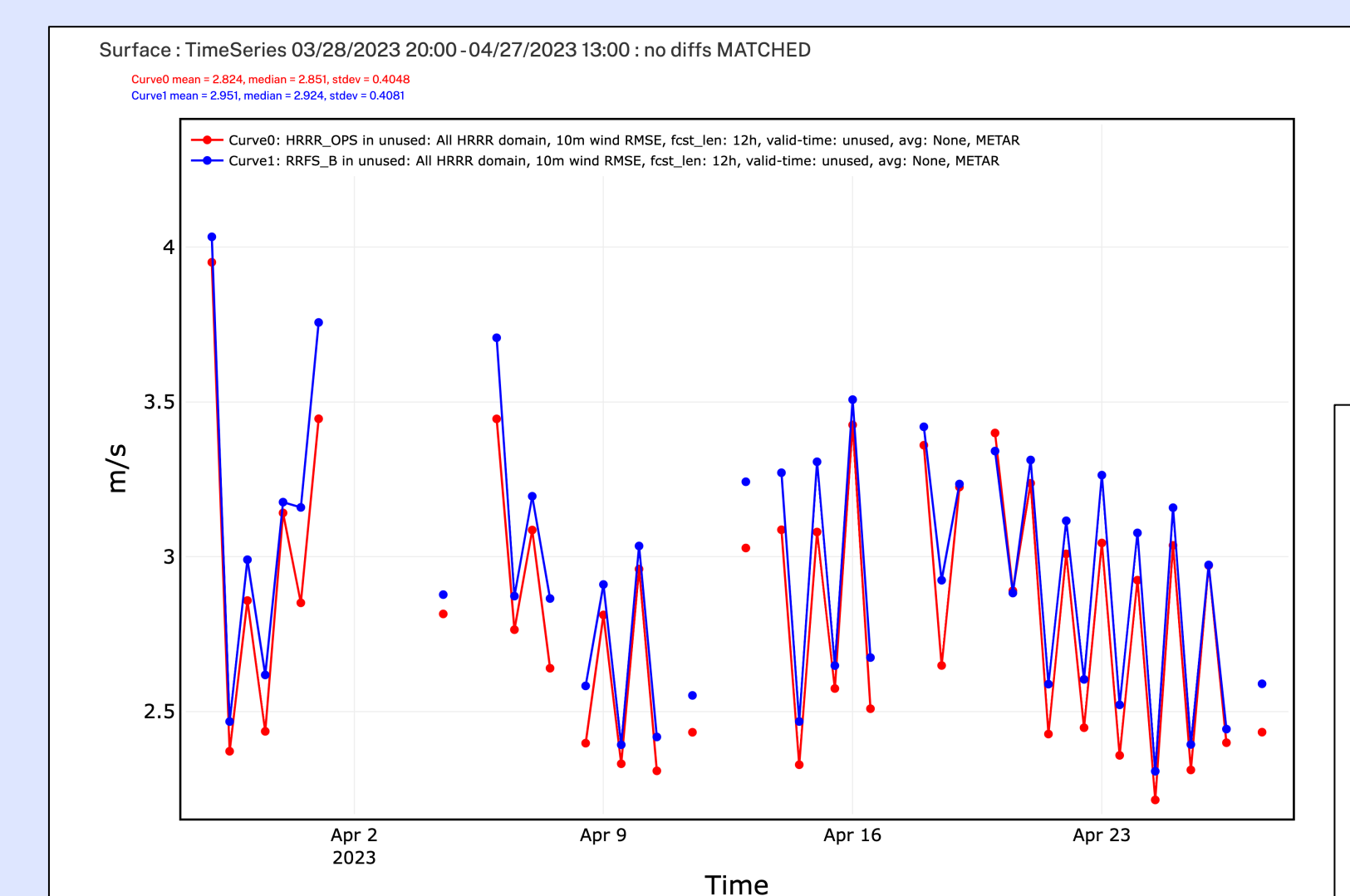
And a scorecard is produced! In this example, we are comparing the experimental RRFS_B against the operational HRRR.



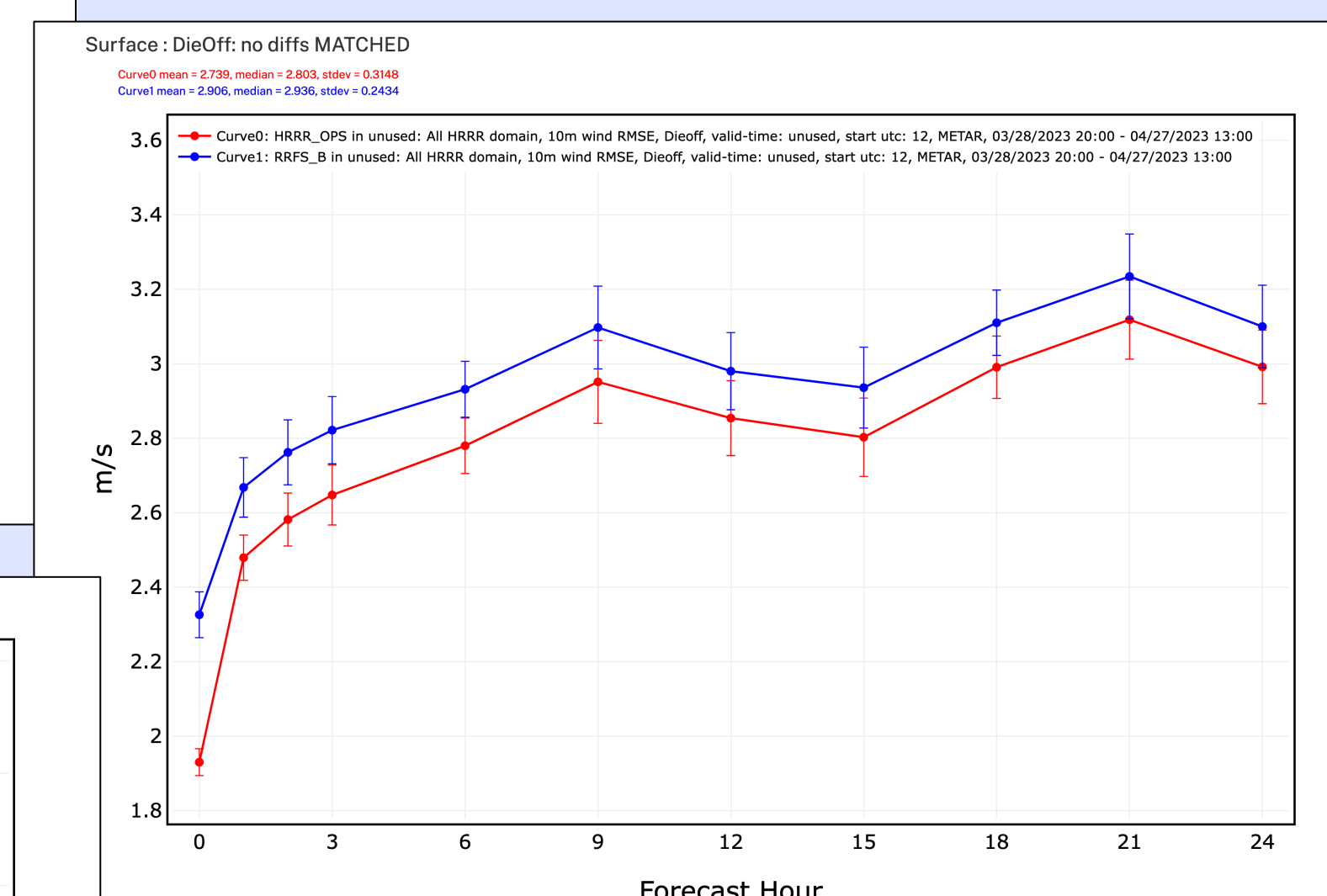
Scorecard Block0: Experimental Data Source = RRFS_B, Control Data Source = HRRR_OPS
Constant fields: truth = METAR; valid-time = unused; dates = 03/28/2023 20:00 -04/27/2023 13:00

| 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 | 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 | ▲▲▲▲▲▲▲▲ | ▲▲▲▲▲▲▲▲ | ▲▲▲▲▲▲▲▲ |

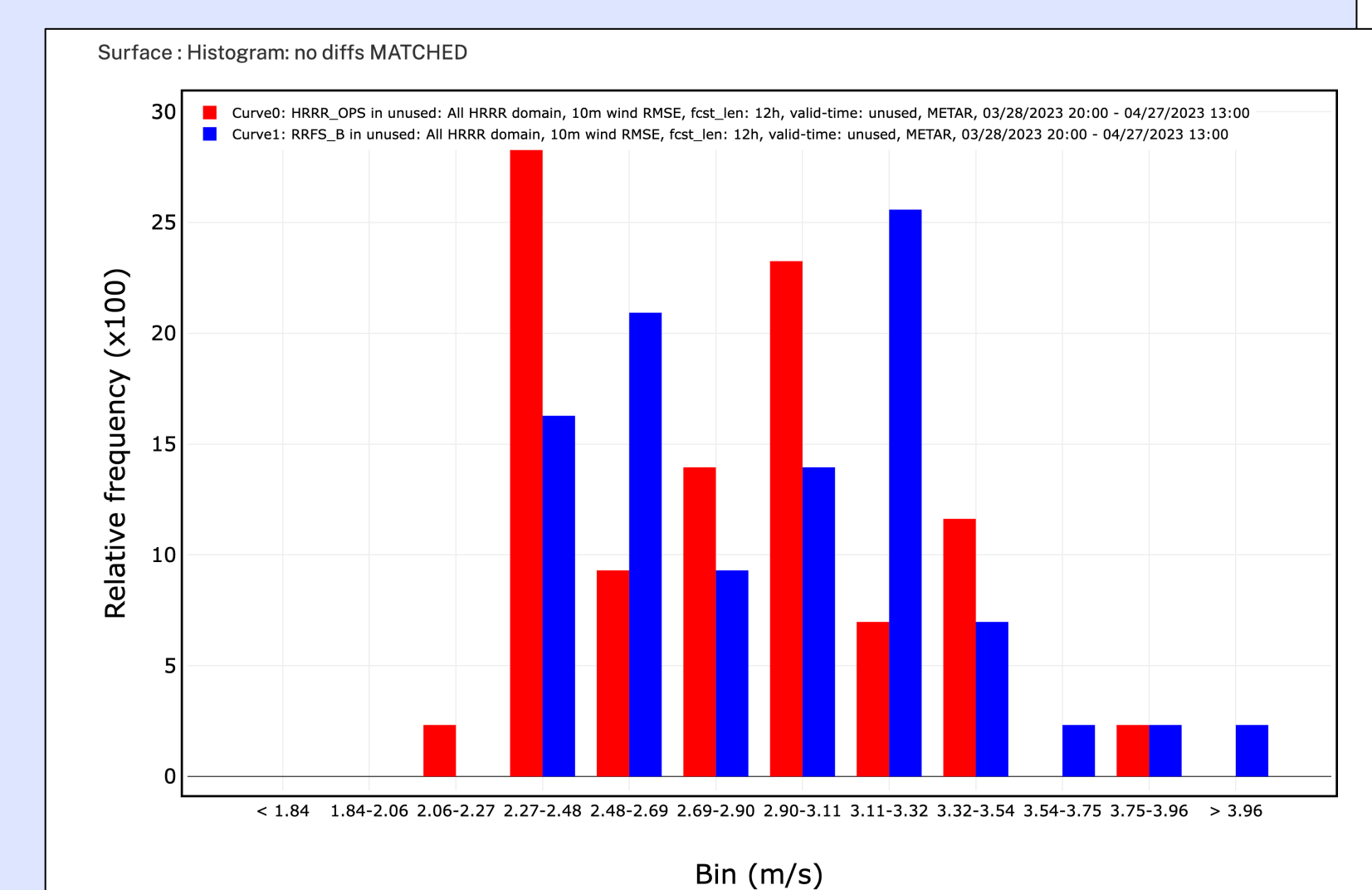
This particular scorecard that we've made compares surface variables, although users can also examine upper level fields, ceiling heights, radar reflectivities, and many other quantities. If we click on a cell, for example the one comparing the two models' wind RMSE over the entire HRRR domain at a 12 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.



Higher RMSE on blue curve (RRFS_B) for most dates and all lead times.



Overall distribution of RMSE values is shifted larger for RRFS_B (blue) than HRRR_OPS (red).



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 as dieoff (forecast lead time on the x-axis), valid time (hour of day on the x-axis), contour plots, or histograms.

