Statistical Analysis of Sailing Forecasts

C. Branning, Model Accuracy Inc., USA, chrisbranning@gmail.com.
G. Sutcliffe, Model Accuracy Inc., USA, 2geoff@pobox.com.
T. Beavers, Model Accuracy Inc., USA, thomas.beavers@gmail.com.
U. Visser, Model Accuracy Inc., USA, ubbo.visser@gmail.com.
R. Schutt, US Olympic Sailing Team, USA, rileyschutt@ussailing.org.

Knowing which weather forecast to rely on is a challenging problem for all racing sailors, from ocean navigators to Olympic campaigners. The *Model Accuracy* sailing software compares logged weather data with weather forecast GRIBs, to provide statistical analysis that determines which GRIB is the most accurate with respect to the log data. *Model Accuracy* helps the navigator understand trend errors in weather forecasts, and gives confidence that (s)he is navigating with the right GRIB for the current location on the water. Two test cases show the applicability of the method to distinct racing scenarios. The first shows how logs from onboard wind instrumentation in a long distance ocean race can be used to identify the best performing forecasts from the previous day, and to easily calibrate out persistent errors in wind speed and direction. The second shows how data from multiple ground and water-based wind data sources at the 2020 Olympic venue in Enoshima, Japan are used to assess the long term accuracy of competing forecast models. *Model Accuracy* is essentially the "consumer report" of weather forecast GRIB files.

1 INTRODUCTION

The demand for precisely correct navigation decisions in ocean racing on sailboats has generated a sustained demand for increasingly capable navigation software that generates optimized routes in real time onboard. Well known examples of such navigation software are Expedition[1] and Adrena[2]. Navigation software relies primarily on the boat's Velocity Prediction Program (VPP) and a selected weather forecast presented as a General Regularly-distributed Information in Binary (GRIB) file. Additional input to navigation software comes from the boat's telemetry, transmitted from the boats instruments (mainly the anemometer) to an onboard computer using a standardized protocol, e.g., National Marine Electronics Association (NMEA)[3]. Navigation programs typically convert the raw log data into a format that is easier to work with directly, e.g., Comma Separated Value (CSV) format. Other sources of weather data might also be used, e.g., data coming from weather buoys such as provided in the USA by the National Oceanic and Atmospheric Administration (NOAA)[4], and data coming from shore stations or instruments mounted on coach boats.

The VPP of a boat, commonly called the "polars", is critical to a navigator's ability to ensure that the routing software provides the most accurate projected performance. The VPP of a boat is often well calibrated, based on the boat manufacturer's testing and subsequent tuning from sailing experience. In contrast, weather forecasting to produce GRIBs that accurately predict future weather conditions, in particular wind speed and direction, is notoriously difficult. Without an accurate weather forecasting GRIB, the polars cannot be used to produce an effectively optimized route. As a result, there is a plethora of GRIB sources (GFS, EC, COAMPS, Predict Wind, etc.), all claiming they are the best source to trust when making navigation decisions.

GRIBs from a given source are produced every six to twelve hours - these are called GRIB *runs*. The time of a GRIB run is given in Universal Time Coordinated (UTC), e.g., if a GRIB was produced at 0600 UTC it is considered the "zero six" GRIB run for that source. Within each GRIB forecast there are *steps* of time at which the weather prediction is updated. Most steps are on the hourly mark, ranging from 12hr steps down to 1hr steps. The smaller the time steps within a GRIB the higher the resolution of the forecast, as there is less uncertainty about the forecast at times not at a step time.

Until now there has been no way to formally analyze the accuracy of GRIBs while on the water. The *Model Accuracy* sailing software allows a navigator to analyse GRIBs relative to what is being observed by the boat or buoy at the current location, and hence make an informed and confident decision about which GRIB to trust the most. With *Model Accuracy* being a stand-alone application, with no internet connection required, the navigator is able to complete an analysis of the available GRIBs while racing in the middle of the ocean. By knowing that navigation decisions are based on the best GRIB source, which has been verified relative to what is being observed at the current location of the boat or buoy in the ocean,

[1] https://www.expeditionmarine.com
[2] https://www.adrena-software.com/en/
[3] https://www.nmea.org
[4] https://www.noaa.gov

the navigator can spend less time in the nav' station. *Model Accuracy* gives the navigator confidence that the heading and tactics are based off the best GRIB source available.

The analysis provided by *Model Accuracy* also allows the navigator to identify the ways in which a GRIB is not correctly predicting the weather, e.g., the extent to which the GRIB is forecasting under/over the wind speed, and left/right of wind direction. Understanding the deviations over a selected time period allows the navigator to correctly raise/lower predicted wind speed and twist predicted wind direction inside the routing functions of other sailing software. This way the projected optimal course is based not only on the most accurate GRIB, but also the most accurate GRIB that has been calibrated to the current and past forecast errors.

2 INSIDE MODEL ACCURACY

Model Accuracy runs in three phases, controlled either through a Graphical User Interface (GUI) or using command line parameters that mimic the GUI elements. (This paper focusses on use via the GUI.) The first phase is data preparation, done in three parts: reading and cleaning the log data, reading the GRIB files, and data pre-analysis to establish possible and recommended analysis start and end times. The second phase is the heart of *Model Accuracy*, in which the log data is compared with the GRIB data, to produce statistical measures of the extent to which each GRIB source agrees with what is being observed in the log data. The third phase reports the statistical data, including a recommendation of which GRIB source to use - the one agrees best best with the log data. Optionally, plots that compare the log and GRIB data can be produced. These phases are described in more detail in the following subsections.

2.1 DATA PREPARATION

2.1.1 Log Data Preparation

Log data can be provided to *Model Accuracy* in Expedition format, Adrena format, NMEA format, or NOAA buoy format. Internally, all formats are converted to Expedition format, to retain the following information about each data record:

- Time the data was logged, in UTC.
- Latitude (LAT) and longitude (LON).
- True Wind Speed (TWS) in knots.
- True Wind Direction (TWD) in degrees.
- Barometric pressure (BAR) in millibars.

Data that falls between the specified analysis start and end times (see Section 2.2) is extracted for cleaning. The extracted data is cleaned to remove inconsistencies, noise, etc. Log data cleaning does the following:

- Remove data records with duplicate times.
- Remove data records whose values are all zero. Such zero records are the result of logging instruments being turned off or disconnected, sometimes momentarily due to boat activity.
- Disable analysis of BAR analysis if too many records have BAR values outside the reasonable range of 900mb to 1100mb. This is necessary because a lot of logged barometric data is faulty.
- Remove data records that have impossible values: negative TWS, TWD outside 0°to 360°, or BAR outside the reasonable range.
- Remove data records that form a plateau (all readings are the same). These are caused by faults in the logging instruments or data collection processes.
- Remove data records that have outlier data values, i.e., data values that are extremely different from the others in a sliding window of time around them. These are typically caused by faults in the logging instruments or data collection processes.
- Optionally, smooth the data over a sliding window, to reduce the impact of very rapid fluctuations in data values.

The remaining "clean" data is saved in CSV format, in reverse chronological order. The reverse order is required for the statistical analysis phase (see Section 2.2).

2.1.2 GRIB Data Preparation

The preparation of the GRIB data is, in principle, quite simple (although the implementation is non-trivial!). *Model Accuracy* scans for files with a `.grb`, `.grb2`, `.grib`, or `.grib2` extension. For GRIB files found in a subfolder, the name of the GRIB source is taken from the name of the folder, otherwise regular expression matching is used on the GRIB filename to identify the GRIB source. The run times of the files are extracted and saved, then each file is processed to find the GRIB steps. For each GRIB step, the U and V components of each GRIB point are extracted - the latitude, longitude, TWS height, and U or V component of the TWS. The GRIB points for each step are saved in CSV format in a separate time-stamped file, ready for the statistical analysis phase (see Section 2.2).

2.1.3 Data Pre-analysis

For the statistical analysis the user has to provide analysis start and end times. Necessarily, these must fall within the limits of the available log and GRIB data. To assist the user, the GUI offers functionality to examine the log and GRIB data, to provide possible and recommended analysis start and end times. The possible times are constrained by the time range of the log data and the maximal time range covered by any one of the GRIBs. The recommended times are constrained by the time range of the log data and the time range covered by all of the GRIBs.

If this pre-analysis is performed, the possible and recommended time ranges are provided to the user in the GUI, and the prepared log and GRIB data is cached for use in the statistical analysis. If pre-analysis is not performed then no possible or recommended start and end times are provided, and data preparation is done for the analysis time range specified by the user (if possible, otherwise analysis is aborted).

2.2 STATISTICAL ANALYSIS

The statistical analysis requires the user to set the parameters of the analysis. The parameters are:

- Level of log data smoothing to apply (default: None).
- Logged wind data is True or Magnetic (default: True).
- Anemometer height (default: 0m).
- Data components to analyse - TWS, TWD, BAR (default: All three).
- Analysis mode - see below (default: *ALL*).
- Analysis granularity - see below (default: 1min).
- Analysis start and end times (No defaults)

Model Accuracy offers three analysis modes: *ALL*, *SINGLE*, and *STEPPED*.

- *ALL* mode compares all the available GRIB sources against the log data. For each GRIB source, at each time point in the analysis, the latest GRIB run and GRIB step before the time point is used for comparison.
- *SINGLE* mode uses only the latest GRIB run before the analysis start time. This allows analysis of the longevity of the single GRIB run.
- *STEPPED* mode breaks the analysis down into user-specified periods, and runs *ALL* mode separately on each period. This allows analysis of the overall longevity of a GRIB source.

Regardless of the analysis mode, the same basic algorithm is used for analysis. The analysis runs backwards in time, starting at the analysis end time (the end of the period in *STEPPED* mode), moving back in time increments of the analysis granularity specified by the user, continuing back to the analysis start time (the start of the period in *STEPPED* mode). At each analysis time point the latest log data point before the time point is extracted from the prepared log data. If the wind direction is specified as being Magnetic, it is converted to True (as TWD) using the 2020 World Magnetic Model produced by the United States National Geospatial-Intelligence Agency. Then for each GRIB being analysed, a two-dimensional bucket search is used to find the GRIB point closest to the latitude and longitude of the log data point, from the latest GRIB run and step before the analysis time point. The log–GRIB data pair is saved for statistical analysis.

After all the data pairs have been accumulated, statistical analysis is performed.

- Correlation between the log and GRIB data values. A Pearson's correlation is used for all except the TWD, which uses a circular correlation.
- The average RMS error of the GRIB values wrt the log values.
- The average absolute error of the GRIB values wrt the log values.
- The average difference error of the GRIB values wrt the log values.
- The overall *Model Accuracy* error of the GRIB data wrt the log data.

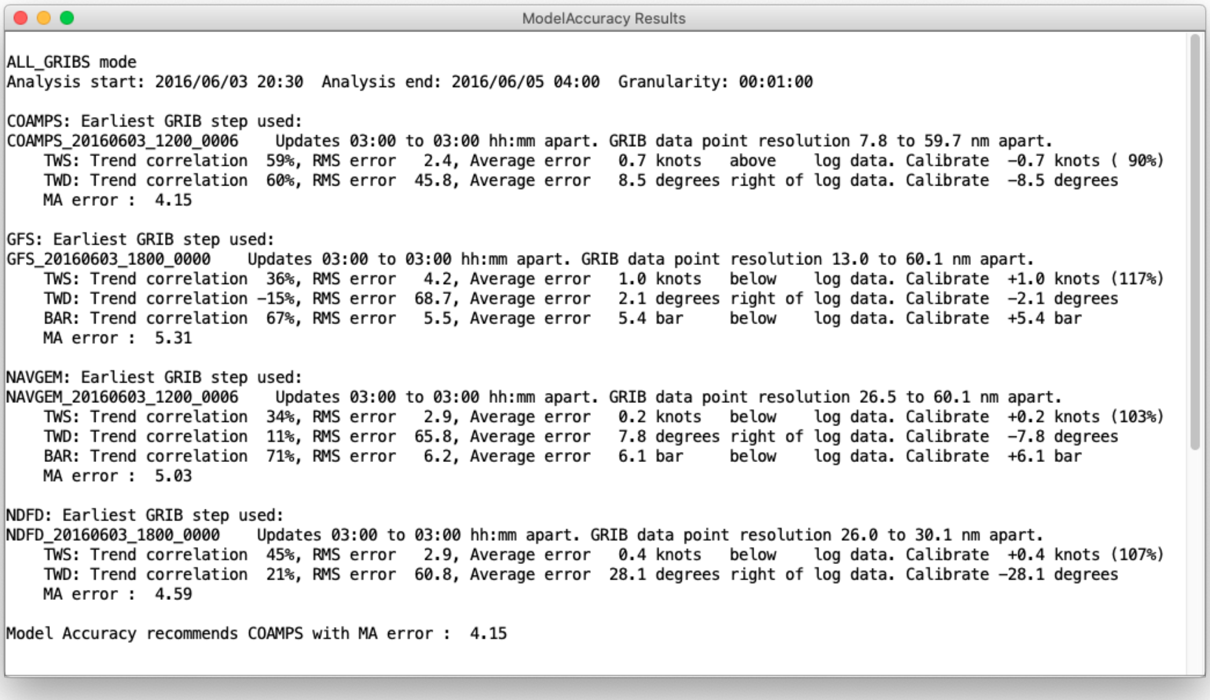
The RMS, absolute, and difference errors are in the units of the data component, i.e., knots for TWS, degrees for TWD, and millibars for BAR. The difference errors are reported as knots/millibars above/below for TWS and BAR, and degrees left/right for TWD. For the TWS the average difference error is also reported as a percentage (for easy calibration in Expedition). The overall *Model Accuracy* error is a proprietary secret, but can be thought of as being in knots.

2.3 RESULTS PRESENTATION

The results of the statistical analysis are output as text, as shown in Figure 1. The report provides:

- The analysis mode.
- The analysis start and end times, and the time step granularity.
- For each GRIB source:
  - The name of the GRIB source.
  - The earliest GRIB run and step used.
  - The range of times between GRIB steps.
  - The range of distances between GRIB forecast points.
  - The statistical analysis results.
  - The overall *Model Accuracy* error.
- A recommendation of which GRIB source to use - the one that agrees best with the log data.
- The analysis parameters (not shown in Figure 1).

In addition to the results of the statistical analysis, plots of the log data and the GRIB data can be output. Examples are shown in Figures 2, 3, and 4. The black line shows the log data values, while the coloured lines show the GRIB values. A red square on a GRIB line indicates the start of a new GRIB run, and a small green square indicates the start of a new GRIB step.



```
ModelAccuracy Results

ALL_GRIBS mode
Analysis start: 2016/06/03 20:30  Analysis end: 2016/06/05 04:00  Granularity: 00:01:00

COAMPS: Earliest GRIB step used:
COAMPS_20160603_1200_0006    Updates 03:00 to 03:00 hh:mm apart. GRIB data point resolution 7.8 to 59.7 nm apart.
    TWS: Trend correlation  59%, RMS error   2.4, Average error   0.7 knots   above    log data. Calibrate  -0.7 knots ( 90%)
    TWD: Trend correlation  60%, RMS error  45.8, Average error   8.5 degrees right of log data. Calibrate  -8.5 degrees
    MA error :  4.15

GFS: Earliest GRIB step used:
GFS_20160603_1800_0000    Updates 03:00 to 03:00 hh:mm apart. GRIB data point resolution 13.0 to 60.1 nm apart.
    TWS: Trend correlation  36%, RMS error   4.2, Average error   1.0 knots   below    log data. Calibrate  +1.0 knots (117%)
    TWD: Trend correlation -15%, RMS error  68.7, Average error   2.1 degrees right of log data. Calibrate  -2.1 degrees
    BAR: Trend correlation  67%, RMS error   5.5, Average error   5.4 bar     below    log data. Calibrate  +5.4 bar
    MA error :  5.31

NAVGEM: Earliest GRIB step used:
NAVGEM_20160603_1200_0006    Updates 03:00 to 03:00 hh:mm apart. GRIB data point resolution 26.5 to 60.1 nm apart.
    TWS: Trend correlation  34%, RMS error   2.9, Average error   0.2 knots   below    log data. Calibrate  +0.2 knots (103%)
    TWD: Trend correlation  11%, RMS error  65.8, Average error   7.8 degrees right of log data. Calibrate  -7.8 degrees
    BAR: Trend correlation  71%, RMS error   6.2, Average error   6.1 bar     below    log data. Calibrate  +6.1 bar
    MA error :  5.03

NDFD: Earliest GRIB step used:
NDFD_20160603_1800_0000    Updates 03:00 to 03:00 hh:mm apart. GRIB data point resolution 26.0 to 30.1 nm apart.
    TWS: Trend correlation  45%, RMS error   2.9, Average error   0.4 knots   below    log data. Calibrate  +0.4 knots (107%)
    TWD: Trend correlation  21%, RMS error  60.8, Average error  28.1 degrees right of log data. Calibrate -28.1 degrees
    MA error :  4.59

Model Accuracy recommends COAMPS with MA error :  4.15
```

Figure 1: A statistical report

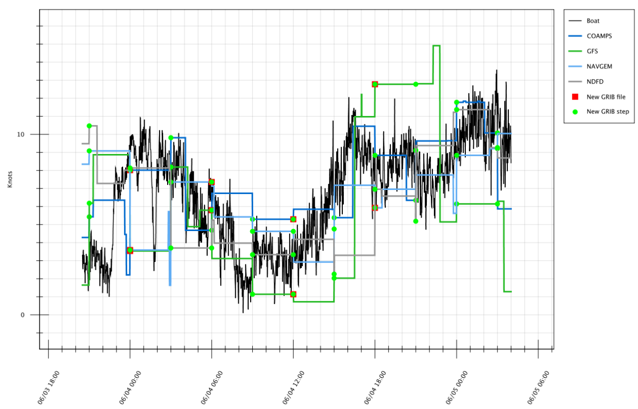


Figure 2: A True Wind Speed plot

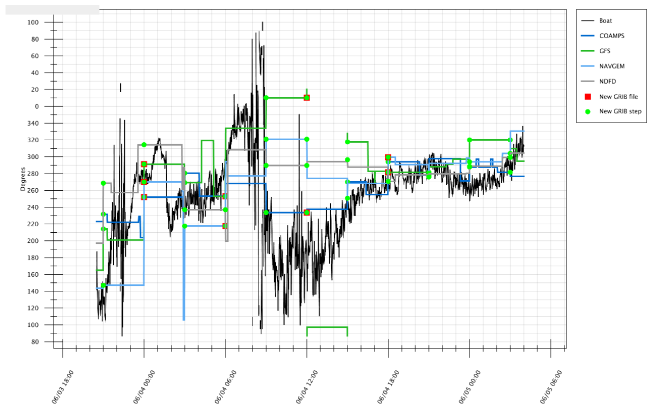


Figure 3: A True Wind Direction plot

3 THE GRAPHICAL USER INTERFACE

Figure 5 shows an example of the GUI, populated ready for statistical analysis. Pre-analysis has already been performed, resulting in possible analysis start and end times of `2016/06/03 12:00` and `2016/06/05 06:28`, and recommended analysis start and end times of `2016/06/03 18:00` and `2016/06/05 06:28`.

The UI elements are as follows:

- Race name - The race name, which is also used to name a folder within the *Model Accuracy* installation to save information about the race.
- New and Delete buttons - Used to start a new race or delete the artifacts of an existing race.
- Log data source - The format of the log data. If NOAA is selected, the NOAA buoy name is provided, and a check-box option to fetch updated buoy data from the NOAA web site.
- Log data directory and GRIB directory - The folders on

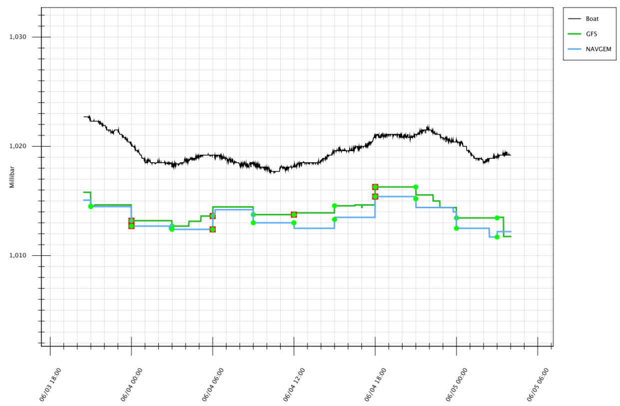


Figure 4: A Barometric Pressure plot

the computer where the raw log data and GRIB files are to be found.

- Wind direction type - Either True or Magnetic.
- Anemometer height - Height of the anemometer in metres.
- Log data smoothing - Size of the sliding window for log data smoothing during log data preparation.
- Prepare Data - The button to invoke data pre-analysis.
- Analysis components - Selection of which data components to analyse - TWS, TWD, BAR.
- Analysis mode - Selection of analysis mode.
- Analysis granularity - Time step size for the statistical analysis.
- Analysis start and Analysis end - The analysis start and end times, the possible start and end times (computed in pre-analysis), and the recommended start and end times (computed in pre-analysis).
- Use and Now buttons - To use the recommended analysis start and end times, or to use the current time as the analysis end time.
- Plots - Option to enable/disable production of plots.
- Analyse! - The button to invoke statistical analysis.
- Save and View - Option to save the report and plots on the computer, and view saved reports and plots.
- Progress bar and Abort - A progress bar that informs the user of activity, and an option to abort the activity.
- Version, license information, Help, and WWW - The version of *Model Accuracy* running, with access to the version history, the licensing state for this computer, access to help files, and access to the *Model Accuracy* web site[5].
- Not shown in Figure 5, but if the computer is not licensed to run *Model Accuracy*, a button to insert or buy a license key.

[5]https://www.ModelAccuracy.com

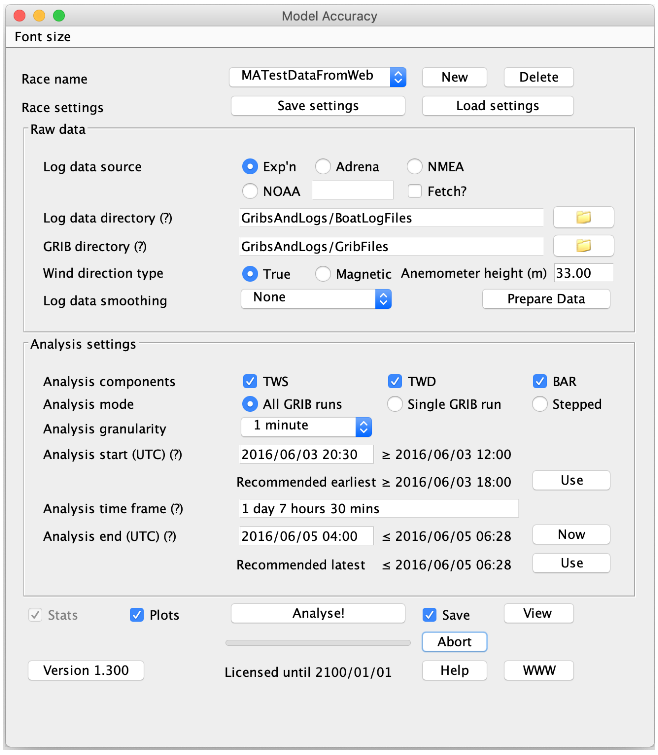


Figure 5: The graphical user interface

4 EXAMPLE USE CASES

4.1 THE 2016 PACIFIC CUP

Model Accuracy was used by the navigator[6] on a 100' super maxi during the 2016 Pacific Cup. Correctly calling when to jibe from starboard to port and nailing the approach to Hawaii is critical for this race course. During the first two days of the race, the navigator downloaded a series of weather forecast GRIBs. Using *Model Accuracy*, the navigator was able to statistically compare the GRIBs with the boat's logged wind instrument data. This revealed that the GFS weather model was clearly the most accurate. However, even though GFS was the most accurate for this part of the ocean, it still had trend errors. It was evident that GFS was forecasting the TWS higher than what the boat's log data was showing, the TWD slightly right of what the boat's log data was showing.

The statistical analysis provided by *Model Accuracy* enabled the navigator to focus his attention on when to jibe, using the GFS GRIBs calibrated within the Expedition software for the error trends revealed by *Model Accuracy*. The result was a new optimal route that encouraged the team to dig into the corner for another two to three hours on a 1,000-nautical mile layline. This was a difficult navigation and tactical call, but the *Model Accuracy* analysis available there in the middle of the ocean gave confidence in the navigator's recommendation. The boat broke the existing course record by nearly three

[6]C. Branning, one of the authors of this paper.

hours. As a collateral benefit, the navigator was able to stop downloading the GRIBs that were shown to be less accurate, saving the team a considerable amount of satellite time and money.

4.2 OLYMPIC SAILING PREPARATION

In preparation for the 2020 Tokyo Olympics, the US Sailing Team spent the summer of 2019 training and competing at the Olympic sailing venue in Enoshima, Japan. *Model Accuracy* was used to assess the performance of various GRIB sources each day, and importantly to track the statistical performance of each GRIB source over time.

Nine GRIB sources were evaluated, here numbered 1-9. GRIB 1 and GRIB 2 were evaluated for multiple release times per day, while the other GRIBs were evaluated using only the latest release before 10am local time each day. The nine GRIB sources fell into four spatial resolution categories, *Global* 50km+ (GRIBs 3, 6, and 9), *Medium* 5km - 10km (GRIBs 5 and 8), *High* 1km - 2km (GRIBs 1, 4, and 7), and *Ultra-High* sub-1km (GRIB 2). Wind data from three anemometer locations were used over the course of three months to assess each GRIB source. In each location, analysis was restricted to 10am - 4pm, isolating possible racing hours. The first anemometer logs came from the Enoshima marina, where TWS and TWD are publicly reported every five minutes. The second set of logs came from anemometers carried onboard the US Sailing Team's coach boats. These anemometers measure TWS and TWD multiple times per second, but data was available only from periods when the coaches were out on the water. The third set of logs were from an anemometer located on the beach to the north of the racing areas and to the east of Enoshima marina.

Figure 6 shows the average *Model Accuracy* errors for the three month period, for each GRIB-anemometer pair. Lighter red squares have a lower average *Model Accuracy* error, indicating better agreement in TWS and TWD between the GRIB forecast and the anemometer observations. Darker red squares have a higher error, and less agreement between the GRIB forecast and the anemometer observations. The lowest error (outlined in blue) occurred between GRIB 2 (18Z) and the coach boat anemometer. This is encouraging, if not unexpected, because it means that the most up-to-date, highest resolution model does the best job of predicting the wind observed on the race courses. In contrast, the anemometer logs from the beach location had poor agreement with all of the GRIBs over the analysis period, resulting in high *Model Accuracy* errors. Looking further into the components of the error, it was found that the beach anemometer was consistently reporting TWS 2kts to 3kts lower than all GRIBs. This likely means this anemometer or measurement location should not be used to assess GRIB performance on the race courses. Further analysis showed GRIBS 4 and 5 consistently predicted TWD more than 30°to the left of observations on the race courses, and should be adjusted. Based on this comparison, the US Sailing Team meteorology staff is likely to weigh the GRIB 2 forecasts more heavily than others.

5 CONCLUSION

This paper has described the *Model Accuracy* sailing software, which compares logged weather data with weather forecast GRIBs, to provide statistical analysis that determines which GRIB is the most accurate with respect to the logged data. *Model Accuracy* helps the navigator understand trend errors in weather forecasts, and provides confidence that (s)he is navigating with the right weather GRIB for the current location on the ocean. Understanding the deviations over a selected time period allows the navigator to correctly raise/lower predicted wind speed and twist predicted wind direction inside the routing functions of other sailing software. This way the projected optimal course is based not only on the most accurate GRIB, but also the most accurate GRIB that has been calibrated to the current and past forecast errors.

Future work includes an optimizer function that does a grid search with increasing resolution to find the best combination of time shift, adjustment of predicted wind speed, and twist of predicted wind direction, to align GRIB data with the observed log data. In this way a GRIB that might seem worse than others can be calibrated to the log data, to produce a GRIB that is better than others.

AUTHORS' BIOGRAPHIES

C. Branning is one of the world's most qualified offshore sailing navigators. He has amassed more than 60,000nm navigating offshore, including the Transpacific, Transatlantic, and Rolex Fastnet races. His podium finishes include the Middle Sea Race, Hong Kong-Vietnam, Barn Door for the TransPacific Race, and the elapsed time record for the Pacific Cup. He was the navigator for Morning Light and All American Ocean Racing. He has had a long time passion for meteorology.

G. Sutcliffe is a Professor in the Department of Computer Science at the University of Miami. Outside of his academic life, Geoff owns a 1973 Bristol 30', and is an active sailor on Biscayne Bay just south of Miami. He also crews (mainly on the foredeck) on various boats in both short and longer races. He served as Commodore of the Coconut Grove Sailing Club in 2018-2019.

T. Beavers received a Bachelor's in Computer Science from the University of Florida. Since then he has worked as a programmer for corporate clients, developing interfaces and background processes that help support clients' marketing efforts. Thomas grew up around sailboats and began racing at the age of 9 in the Miami area. He raced for many years on his boat Finesse, a Tartan 34C, in Biscayne Bay and offshore.

U. Visser is an Associate Professor in the Department of Computer Science at the University of Miami. He does research in several domains, including autonomous robotic systems. He has won several awards for the development of innovative AI software. Ubbo comes from a sailing family, and learned sailing early in his life on a 30' centerboarder on the North Sea between Germany and the Netherlands.

R. Schutt is the Innovation, Research & Development Performance Analyst at the US Sailing Team, working to prepare for the 2020 Tokyo Olympics. Riley received his PhD in Aerospace Engineering from Cornell University, studying the Unsteady Aerdynamics of Sailing. He has pursued a career in the performance modeling and design of racing yachts. In addition to his current Olympic work, he has worked as a designer in the America's Cup, Volvo Ocean Race, Vendée Globe, and on numerous record breaking ocean racers.

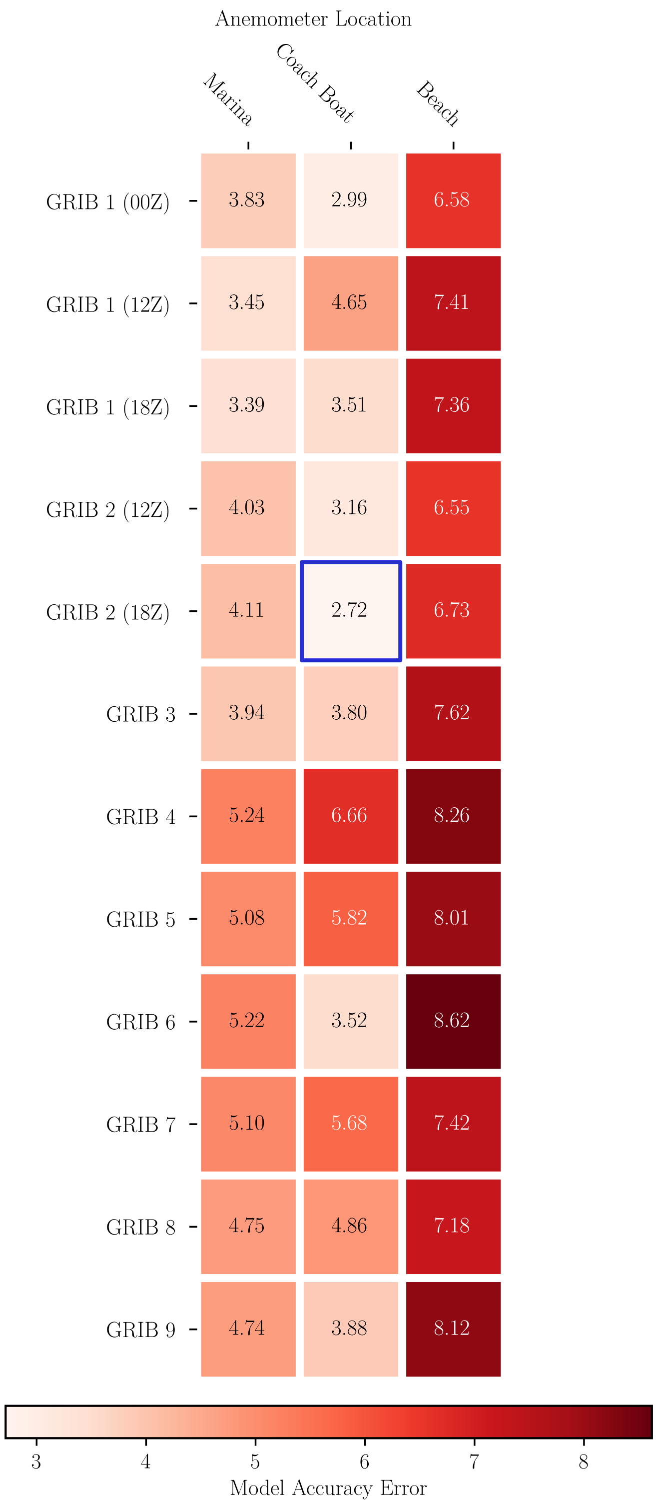


Figure 6: *Model Accuracy* errors in Enoshima, Japan.