Characterizing Error in the GFS Model

Weather and Storms Final Paper

Kai Marshland

December 2018

1 Background

Hurricanes killed 147 people last year in the US alone [1]; many of those deaths came from people who didn't evacuate. Building trust through high-quality predictions of hurricane strength and direction, is therefore critical to saving lives. Authorities often rely on models like GFS (Global Forecast System). The National Oceanic and Atmospheric Association (NOAA) produces GFS using numerous data sources, including satellite data, ground-based weather stations, and weather balloons [2]. These weather balloons, however, only provide data over land. At sea, where hurricanes form, the GFS model relies heavily on satellite data. In this paper, we aim to characterize the difference between GFS predictions and measured wind data over land and sea alike.

To measure wind data, we used measurements from flights of ValBal, a long-endurance high-altitude balloon platform. Where a typical weather balloon, such as those NOAA uses, flies for only 2-3 hours, ValBal flies for upwards of 120 hours for a comparable cost [3]. Developed by the Stanford Student Space Initiative, it maintains altitude within a certain band throughout the flight [4]. As the balloon drifts with the wind, position and time information provide wind data. We focus on data from two flights in particular, SSI-52 and SSI-77, as those flew both over land and over sea, allowing us to compare the relative accuracy of GFS in both cases. These flights both crossed the continental US and spent significant time over the Atlantic Ocean. Both flew primarily at altitudes between 13 and 16km. We also include data from SSI-63 and SSI-67, both long-lasting flights over the continental US.

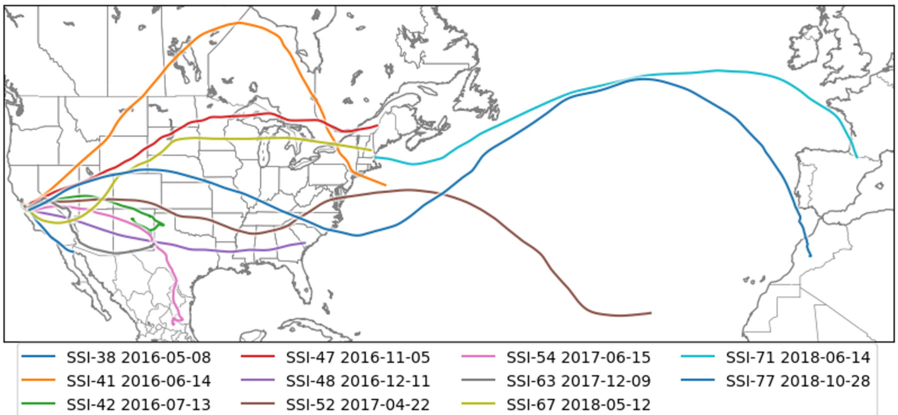

Figure 1: Selected ValBal Flights

2 Methodology

Note: All code is available at `https://github.com/KMarshland/wind-error-analysis`

To calculate wind data from the balloon path, we first had to acquire the GPS values and timestamps. These were downloaded from `https://habmc.stanfordssi.org/`, Stanford SSI's high-altitude balloon mission control suite. This only includes data downlinked via satellite communications, which is much lower resolution than the data logged to the SD card (once every few minutes compared to several times per second). It also potentially suffers from timing issues, as the time at which data is sent through the Iridium satellite network is not necessarily the same as the time at which it was measured. To verify that the satellite transmissions provided an accurate picture, we also ran the analysis on flights for which we had recovered the SD card data. These analyses showed that the satellite data corresponded relatively well to the logged SD card data. For example, for SSI-71, the average speed according to the SD card was 24.48 m/s and 23.17 m/s according to the satellite data. This shows that, while the satellite data may be off by approximately 5%, it is accurate enough to provide valid analysis for those flights for which we do not have SD card data, namely those flights which landed in the ocean.

Once position and time data were processed out from the dataset, we could calculate distance and bearing between two subsequent points. We calculated distance with the Haversine formula, where ϕ_1 represents the latitude of the first point in radians, λ_1 represents the longitude of the second point in radians, and likewise for the second point:

$$a = \sin((\phi_2 - \phi_1)/2) * \sin((\phi_2 - \phi_1)2) + \cos(\phi_1) * \cos(\phi_2) * \sin((\lambda_2 - \lambda_1)/2) * \sin((\lambda_2 - \lambda_1)/2)$$

$$\text{distance} = 2 * \text{atan2}(\sqrt{a}, \sqrt{1-a}) * R_{earth}$$

In order to calculate bearing, we used the following formula:

$$x = \cos(\phi_1) * \sin(\lambda_2 - \lambda_1)$$

$$y = \cos(\phi_1) * \sin(\phi_2) - \sin(\phi_1) * \cos(\phi_2) * \cos(\lambda_2 - \lambda_1)$$

$$\text{bearing} = \text{atan2}(y, x) + 90^\circ$$

We then extracted wind data from the GFS 004 (0.5°) model, downloaded from `https://nomads.ncdc.noaa.gov/data/gfs4/`. Four times for each day, a new forecast is released, forecasting from 0 hours relative to the forecast time to 384 hours in 3 hour increments. When extracting the GFS wind data, we selected the two datasets closest to the specified time with the lowest possible offset. For example, for the time 2018-11-10 at 10am, it would select the dataset generated on 2018-11-10 at 6am forecasting three hours in the future and the dataset generated on 2018-11-10 at 12:00pm forecasting zero hours in the future.

After selecting the two nearest datasets, we find the eight closest datapoints in each one, as the dataset only has data for certain combinations of latitude, longitude, and pressure. We then select the u and v values from each of those eight datapoints and linearly interpolate by position, and then linearly interpolate again by time to deliver the final values. Bearing can be calculated as:

$$\text{atan2}(u, -v)$$

We then calculated both net error and RMS for both bearing and speed, over land, over sea, and across the flight as a whole. Net error was calculated by taking the average value in the data and the average value in the prediction, then taking the absolute value of the difference. RMS error was calculated by taking the square root of the mean of, for each point, the difference between predicted and measured squared.

3 Results

GFS appears to be relatively accurate over long distance scales, but less accurate over short ones: though net error tends to be low, RMS error is high. This could easily be explained by GFS placing pressure gradients in different places than reality. A high pressure gradient will cause faster geostrophic winds [5]. Where these high pressure gradients are placed won't matter to the net error (ie the difference between predicted and measured average speed and average bearing), as the balloon will still pass these high pressure gradients at some point, but it will have a strong effect on RMS error. For example, if it goes through the fast winds later than expected, the speed will first be low, then be high where it's predicted to first be high and then low. Despite average speed being about the same between predicted and measured over the course of the flight, the difference between predicted and measured speed will be high at any given point.

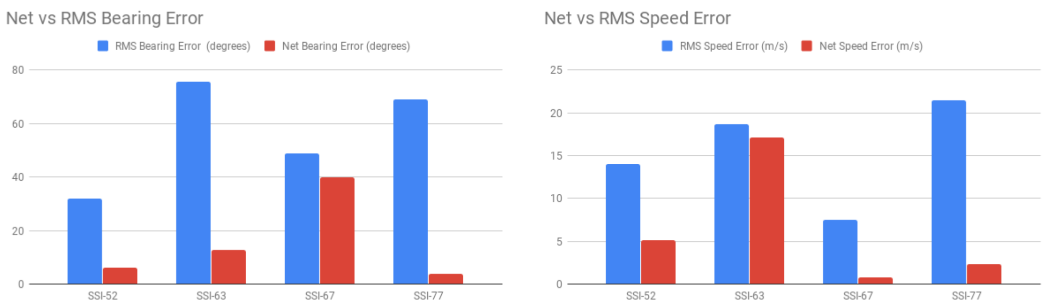

Figure 2: Comparison between net and RMS error

Another piece of the high RMS error might be the noise in the data. Part of the noisiness might be due to real variation in winds that GFS has too low a resolution to capture, but much of this is likely due to noise in the GPS measurements themselves. Note that there are some discontinuities in the data where the GPS was turned off during the flight.

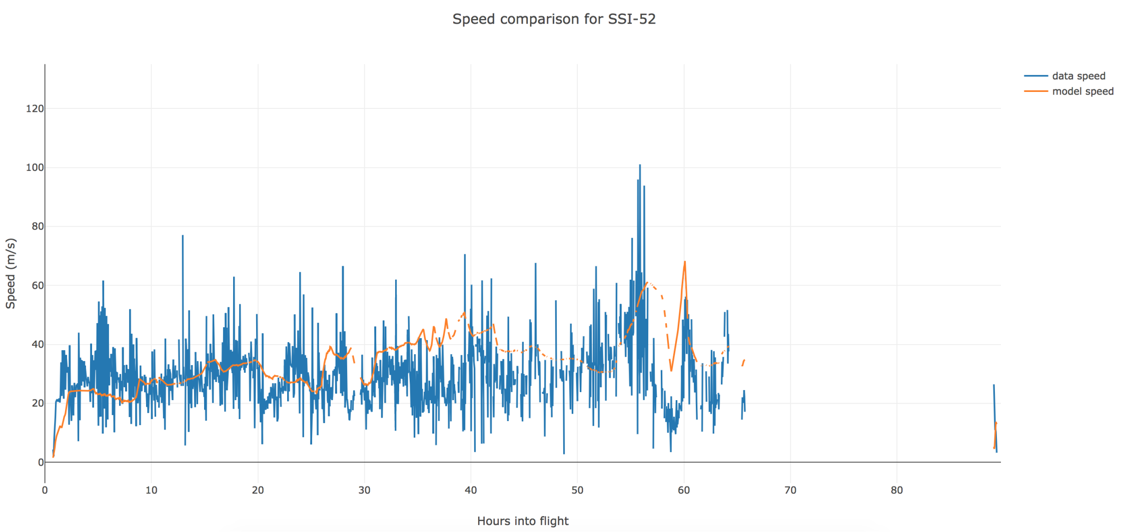

Figure 3: Comparison between measured and predicted speed, showing noise

GFS appears to be much less accurate over sea than over land, at least for speed. This fits the initial hypothesis. Any model is only as good as its inputs, and GFS simply has fewer inputs over the ocean that it does where ordinary weather balloons can be launched. That said, we only have data from two flights that spent significant time above both land and ocean; more data may prove that differences in GFS accuracy over land vs over sea are not statistically significant. Over the course of future flights, we hope to collect enough data to make a more accurate assessment.

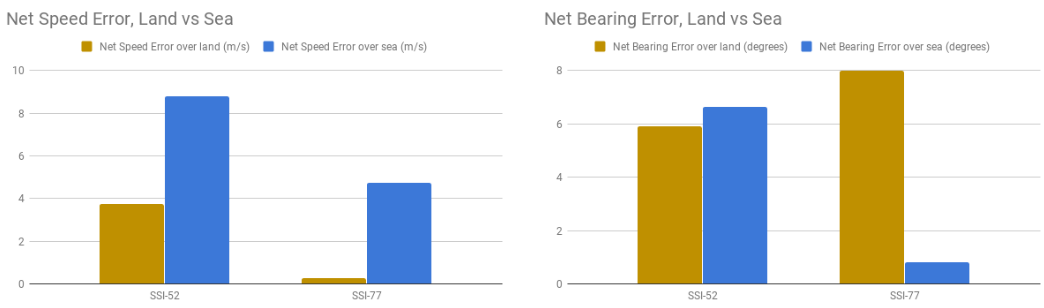

Figure 4: Comparison between data over land and data over the ocean

4 Future work

This methodology could easily be applied to alternative weather models. The ECMWF dataset is the primary alternative model to GFS, and considered to be more accurate as well [6]. Another potential weather model to consider is the NAM model. Unlike GFS or ECMWF, NAM covers only North America, but is at a higher resolution.

This analysis considered only the closest forecast to a given time. Given that GFS releases forecasts every 6 hours, this meant that it never used forecasts that were more than 3 hours old relative to the balloon data. One could easily compare the measured data with forecasts further in the future.

References

[1] Insurance Information Institute, "Facts + Statistics: Hurricanes" `https://www.iii.org/fact-statistic/facts-statistics-hurricanes#Hurricanes%20And%20Related%20Deaths%20In%20The%20United%20States,%201998-2017`

[2] Ferrel, Jesse, "The Secrets of Weather Forecast Models, Exposed", 2010, Accuweather. `https://www.accuweather.com/en/weather-blogs/weathermatrix/why-are-the-models-so-inaccurate/18097`

[3] A. Sushko et al., "Advancements in low-cost, long endurance, altitude controlled latex balloons (ValBal)," 2018 IEEE Aerospace Conference, Big Sky, MT, 2018, pp. 1-10.

[4] A. Sushko et al., "Low cost, high endurance, altitude-controlled latex balloon for near-space research (ValBal)," 2017 IEEE Aerospace Conference, Big Sky, MT, 2017, pp. 1-9.

[5] Ahrens, C. Donald and Robert Henson, *Meteorology Today: An Introduction to Weather, Climate, and the Environment*, 12th edition. C. D. Ahrens and R. Henson, Cengage Learning, Boston, 2019.

[6] Kerns, B. W., and S. S. Chen (2014), ECMWF and GFS model forecast verification during DYNAMO: Multiscale variability in MJO initiation over the equatorial Indian Ocean, J. Geophys. Res. Atmos., 119, 3736–3755, doi: 10.1002/2013JD020833.

[7] Riddle, E. E., P. B. Voss, A. Stohl, D. Holcomb, D. Maczka, K. Washburn, and R. W. Talbot (2006), Trajectory model validation using newly developed altitude-controlled balloons during the International Consortium for Atmospheric Research on Transport and Transformations 2004 campaign, J. Geophys. Res., 111, D23S57, doi: 10.1029/2006JD007456.

5 Appendix

5.1 Acknowledgements

The author would like to thank the Stanford Student Space Initiative as a whole, in addition to the ValBal team in particular for making this analysis possible. Joan Creus-Costa and John Dean deserve special thanks for their help in processing past ValBal data and suggesting improvements to the methodology.

5.2 Related work

A similar analysis was made in "Trajectory model validation using newly developed altitude-controlled balloons during the International Consortium for Atmospheric Research on Transport and Transformations 2004 campaign" [7]; it found that mean trajectory error in GFS was approximately 26%. In our analysis, we calculated mean trajectory error to be approximately 20%, which fits their analysis.

5.3 Predictive power of GFS

The result that GFS is relatively accurate over the distance and time scales of a flight is corroborated by Stanford SSI's prediction tools. This paper's analysis may feed back into the prediction tools by allowing us to simulate error in the wind data more correctly.

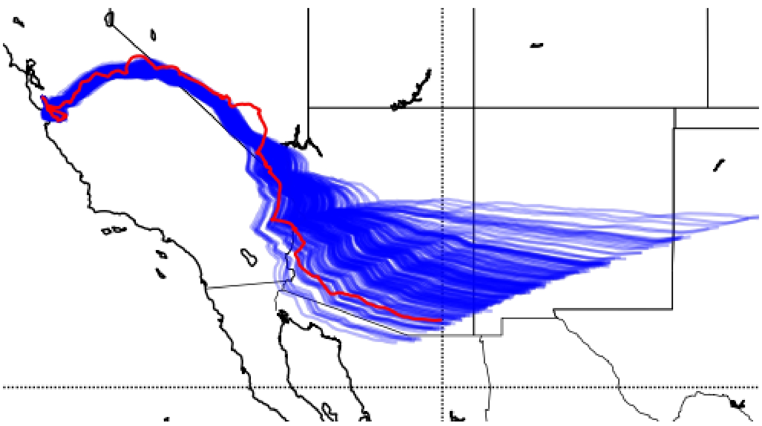

Figure 5: Blue is simulated flight paths; red is the actual flight of SSI-63