The Diurnal Cycle of Winter Season Temperature Errors in the Operational Global Forecast System (GFS)

Ronak N. Patel[1], Sandra E. Yuter[1], Matthew A. Miller[1], Spencer R. Rhodes[1], Lily Bain[1], Toby Peele[1]

[1]Department of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC

Key Points:

- NOAA's GFS model struggles to adequately represent the diurnal cycle of temperatures under observed conditions of $\leq 50\%$ and 25% cloud cover
- NOAA's HRRR model uses a different physics suite and does not have a strong diurnal cycle of temperature errors under the same conditions
- Examination of errors using similar weather conditions helps to constrain the portion of model physics that can yield larger forecast errors

Corresponding author: Sandra Yuter, seyuter@ncsu.edu

Abstract

Forecasts from NOAA's Global Forecast System (GFS) and the High-Resolution Rapid Refresh (HRRR) weather models are matched to surface observations for the winter season of November 2019 to March 2020 at 210 airports across the United States. The 2-m temperature errors, conditioned on observed weather conditions such as cloud cover amount and wind speed, are used to determine the nature of systematic model biases. We observe a strong diurnal cycle in 2-m temperature errors in the GFS in conditions with $\leq 50\%$ and $\leq 25\%$ sky cover, with a 1°C warm bias at night and a 2°C cold bias during the day. The HRRR, which uses a different set of physical parameterizations, does not have a clear diurnal cycle in errors under the same conditions. These results highlight the utility of weather-conditional comparisons across the diurnal cycle to diagnose sources of model weaknesses and to target model improvements.

Plain Language Summary

We evaluate the output from weather forecast models compared to observations at 210 airports across the United States during the November 2019 to March 2020 winter season. We focus on near-surface air temperature errors in the Global Forecast System (GFS) and High-Resolution Rapid Refresh (HRRR) weather models for different times of day and subsets of observed weather conditions. The GFS is 1°C too warm at night and 2°C too cold during the day in conditions with $\leq 50\%$ and $\leq 25\%$ cloud cover. The daily high and low temperatures have smaller errors in the HRRR model, which has different algorithms than the GFS model. Model refinement and development efforts would benefit from a focus on accurate representation of the diurnal cycle of temperatures as this basic characteristic of weather can reveal strengths and weaknesses in the model physics.

1 Introduction

Numerical weather prediction (NWP) models involve a suite of physical parameterizations, including convection, microphysics, land surface, boundary layer, and radiation schemes. The joint interactions among these parameterizations often yield difficulties in diagnosing sources of error within a model (e.g., Fovell et al., 2010; Bu et al., 2017; Caron & Steenburgh, 2020). The National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) and the High-Resolution Rapid Refresh (HRRR) models undergo a detailed testing and verification process before new operational versions are released (e.g., EMC Model Evaluation Group, 2019, 2020a, 2020b). NCEP's verification process focuses on aggregate statistics at the hemisphere, conterminous United States (CONUS), or CONUS sub-region scales and uses case studies to illustrate specific model strengths and weaknesses. For example, NCEP has documented a cold bias of approximately 0.5°C in CONUS East and 0.7°C in CONUS West within the GFS v15 at approximately 36 hours that increases in magnitude with lead time (EMC Model Evaluation Group, 2020a).

We use a relational database to facilitate analyses for specific forecasts and observed conditions. Examination of hourly model output across the diurnal cycle combined with conditioning on specific weather conditions provides a robust test of several aspects of model physics and aids error diagnosis by constraining conditions when the errors occur. Currently, publicly accessible evaluations of the NOAA models (e.g., EMC Model Evaluation Group, 2019, 2020a, 2020b) and the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting System (e.g., Haiden et al., 2021) do not utilize conditional analysis of model errors. We show that examining model temperature errors by similar weather conditions provides substantial additional value to diagnosing model inadequacies.

Our verification methodology compares the model forecasts to *observations*, not to reanalyses. A key weakness of reanalyses is that their accuracy is less well understood than the uncertainties in observations (Parker, 2016). Data assimilation methods often weigh observations less when they differ more from the model's solution (e.g., Houtekamer et al., 2005). Hence, uncertainties in reanalysis products are likely larger in locations and in weather conditions where numerical forecast models struggle—the very set of circumstances where information is most critical for model evaluation and refinement. The downside of using observations is that they are not available everywhere. If a certain bias is present throughout much of the United States, it is more likely to be the result of a model physics weakness than an observation issue.

2 Data and Methods

Data from Automated Surface Observing System (ASOS) sites and GFS and HRRR models are compared for the period of 1 November 2019 to 31 March 2020. Use of this 5-month season constrains characteristics such as number of daylight hours and land surface conditions. ASOS observations and matched model point data are stored in a MySQL relational database, which allows for easy querying of the data for analysis and requires much less space than storing the full model gridded files. For matching to model output, we use the following observed meteorological variables: 2-m temperature, 10-m wind speed, and sky condition. For each ASOS site, we obtain the corresponding GFS and HRRR model values of 2-m temperature, 10-m wind speed, and snow depth at each model run's set of valid times and lead times.

2.1 Observations

We used hourly Meteorological Terminal Air Reports (METAR) from 210 ASOS sites at airports in the CONUS to compare to model output. After data processing and quality control, variable values for each airport are uploaded to the database. The top-of-the-hour observations (i.e., no special observations) are compared to the model forecast valid at that hour. When the magnitude of the model temperature error was greater than 20°C, that specific forecast hour and observation pair is not used in the analysis even though that point passed NOAA's Meteorological Assimilation Data Ingest System (MADIS) quality control and our ingest determined the temperature value to be physically plausible (i.e., within range of temperatures observed on Earth's surface). Sky cover conditions are delineated by ASOS as CLR $\leq$ 5%, 5% $<$ FEW $\leq$ 25%, 25% $<$ SCT $\leq$ 50%, 50% $<$ BKN $\leq$ 87%, and OVC $>$ 87% (NOAA, 1998). We utilize this information to group cloudiness conditions into All conditions ($\leq$ 100% cloud cover, includes OVC, BKN, SCT, FEW, and CLR) , $\leq$ 50% cloud cover (includes SCT, FEW, and CLR), $\leq$ 25% cloud cover (includes FEW, and CLR), and $\leq$ 5% cloud cover (includes CLR).

The specific location chosen for each airport site was the approximate center of the airport property. This was a compromise between the ASOS site and the other associated sensors used to make meteorological measurements at different points across the airfield (NOAA, 1998). Choosing a central location accounts for the unknown variation in exact locations used for measurements. For example, an airport may have multiple wind sensors but only report the value from the active runway. For 201 out of the 210 airport sites, we found the approximate center of the airport property to be within 2 km of the ASOS station. For the other 9 airports, the ASOS site was within 3 km of the airport's center.

2.2 Model Output

2.2.1 GFS

We used the operational versions of NOAA's GFS model for analysis. GFS v15.1 changed to v15.2 on 7 November 2019 at 1200 UTC (Maxson, 2019), so we used GFS v15.1 before 7 November 2019 at 1200 UTC and GFS v15.2 after. The absence of any major model changes with this update (Maxson, 2019) allows the entire date range to be analyzed in aggregate. All GFS initialization times (0000, 0600, 1200, and 1800 UTC) were ingested into the database. We used the hourly GFS output for forecast hours 1 to 120. Since no long-term archive of the hourly output was known to exist, our own archive had to be created using the rolling 30-day archive on Amazon Web Services (AWS).

The 0.25-degree gridded GFS data were downloaded from the NOAA AWS cloud, which is part of their Big Data Program. Spatial linear interpolation is used to obtain model values within the 0.25-degree grid boxes at the 210 airport sites. The coarse resolution can yield mismatches between actual and modeled surface types for airports with runways adjacent to water. For example, the New York airports JFK and LGA are classified as water surface type rather than land.

2.2.2 HRRR

We compared the HRRR v3 (NOAA, 2020) to the GFS from 1 November 2019 to 31 March 2020 for forecast hours 0 to 36 based on the 0000, 0600, 1200, and 1800 UTC initialization times. Since the effective grid length in this model is approximately 3 km (NOAA, 2020), the nearest model grid point was chosen as being representative of the conditions at each of the airport sites. HRRR grids were downloaded from the University of Utah's HRRR archive (Blaylock et al., 2017), and data at the nearest grid point to the 210 airports were used to populate the database for this study.

2.3 Diurnal Cycle

To address the diurnal cycle of temperature errors, we examine hourly data at the time of the winter climatological daily low and high temperatures at 7 a.m. and 3 p.m. local standard time (LT), respectively. We approximate 7 a.m. and 3 p.m. LT using longitude bands: Eastern time between 67.5° W and 82.5° W as 1200 UTC and 2000 UTC, Central time between 82.5° W and 97.5° W as 1300 UTC and 2100 UTC, Mountain time between 97.5° W and 112.5° W as 1400 UTC and 2200 UTC, and Pacific time between 112.5° W and 127.5° W as 1500 UTC and 2300 UTC.

The local times of the climatological daily low and high temperature often do not coincide with the four times a day where exact 36-hour forecasts exist. We select among the 31 to 36 hour forecasts, picking the one closest to but less than the target time. The target time is the daily low or high temperature for each location (as in Table 1 and Fig. 1), or a specific UTC hour (as in Fig. 2). Since we use forecasts initialized every 6 hours, a target time of 0100 UTC corresponds to a lead time of 31 hours, 0200 UTC to 32 hours, 0300 UTC to 33 hours, 0400 UTC to 34 hours, 0500 UTC to 35 hours, and 0600 UTC to 36 hours. This pattern is repeated every 6 hours. For brevity, we call all of these a 36-hour lead time.

2.4 Cloud Cover and Wind Conditions

We condition the data on observed weather conditions to determine the dependence of temperature errors on cloud cover amount, winds, and snow cover. Our initial analysis uses all observed sky conditions and wind speeds to compare our results with NOAA's. We then perform a conditional analysis using observed $\leq 50\%$, $\leq 25\%$, and $\leq 5\%$ sky cover. Observed low 10-m wind speeds (≤ 2.57 m/s or 5 kt) are also used. We use 5 knots

Table 1. Average CONUS temperature errors for the GFS and HRRR models for different cloudiness conditions at 36-hour lead time.

	All conditions	$\leq$ 50% cloud cover	$\leq$ 25% cloud cover	$\leq$ 5% cloud cover
GFS Day (3 p.m.)	-1.3°C ± 0.1	-1.9°C ± 0.1	-1.9°C ± 0.1	-1.8°C ± 0.1
GFS Night (7 a.m.)	0.0°C ± 0.1	1.0°C ± 0.1	1.1°C ± 0.1	1.2°C ± 0.1
HRRR Day (3 p.m.)	0.2°C ± 0.1	-0.6°C ± 0.1	-0.5°C ± 0.1	-0.5°C ± 0.1
HRRR Night (7 a.m.)	-0.5°C ± 0.1	0.1°C ± 0.1	0.2°C ± 0.1	0.2°C ± 0.1

Data are for the time of the diurnal high temperature (3 p.m. LT) and diurnal low (7 a.m. LT). Uncertainty estimates are sample standard deviations of the mean errors.

as our low wind speed threshold since below 5 knots, ASOS wind directions are not reliable (NOAA, 1998).

3 Results

Our analysis yielded a GFS daily average cool temperature bias based on the airport locations of -0.70°C at 24 hours increasing in magnitude to -0.9°C at 120 hours, which closely matches the daily average biases found by NCEP (EMC Model Evaluation Group, 2020a) (not shown). We more closely examine the daily average biases in surface temperatures within the GFS and HRRR over CONUS at 36-hour lead time for the 210 airports in our relational database (Figure 1). To help diagnose conditions when these biases are more frequent, we examine biases at the times of the climatological low and high temperatures (7 a.m. and 3 p.m. LT). Table 1 shows the clear diurnal variation in average CONUS temperature errors within GFS for all conditions, with a bias at 3 p.m. LT of -1.3°C. The HRRR has a bias of approximately -0.5°C for the nighttime-low and a slight warm bias during the daytime-high for all conditions.

We examined various weather conditions to determine circumstances where stronger biases were more likely to occur. We found that 7 a.m. LT warm biases were usually larger in conditions with less cloudiness for both the GFS and HRRR. Figure 1 shows the average model biases at 36-hour lead time for the subset of conditions when $\leq$ 50% cloud cover and when $\leq$ 25% cloud cover is observed for each airport. For the $\leq$ 50% cloud cover subset of data, the CONUS average 7 a.m. LT temperature bias in the GFS is 1.0°C compared to 0.1°C for the HRRR (Table 1). When the data for $\leq$ 50% cloud cover are further conditioned for low winds ($\leq$ 2.57 m/s or 5 kt), the CONUS average 7 a.m. LT low temperature error increases to 1.7°C for the GFS and increases to 0.6°C for the HRRR (not shown). Strong nocturnal inversions are typically associated with conditions of low sky cover and light winds (Ahrens & Henson, 2019). The spatial pattern of temperature errors is very similar between conditions for $\leq$ 50% cloud cover and $\leq$ 25% cloud cover. At individual airports, some of the error magnitudes are higher for $\leq$ 25% cloud cover (Figure 1, Table 1).

At the approximate time of the high temperature (3 p.m. LT), the spatial patterns of errors are markedly different between the GFS and HRRR. There is a GFS cold bias at a 36-hour lead time in conditions of $\leq$ 50% cloud cover (Figure 1a), with a CONUS average of -1.9°C (Table 1). At the time of the daytime high, for periods with $\leq$ 50% cloud cover, the HRRR tends to have a cold bias east of the Great Plains and a warm bias to the west (Figure 1b). The HRRR also tends to have a slight cool bias at 7 a.m. LT for coastal sites. Some of the largest cold biases in the GFS and warm biases in the HRRR are at airport locations in the Intermountain West.

Comparison of diurnal high (3 p.m. LT) and diurnal low (7 a.m. LT) temperatures indicate a strong diurnal cycle of temperature errors under conditions of $\leq$ 50% and $\leq$ 25%

sky cover in the GFS. These biases are present throughout much of the United States, making it unlikely they are solely the result of a mismatch between model terrain and ASOS station elevations, which would be more prevalent in mountainous regions. In their analysis, Fovell and Gallagher (2020) found that elevation error had essentially no association ($r \approx 0.08$) with temperature bias in the HRRR once sites with large elevation discrepancies (>80 m) were removed.

Further details on the hour-by-hour temperature errors for conditions of $\leq 50\%$ sky cover are shown in Figure 2a,b,c,d for Detroit, MI (DTW) and Oklahoma City, OK (OKC), which are representative of many other airports across the US that are distant from mountainous terrain. In the GFS, after sunset (∼0000 UTC) the temperature errors trend warmer overnight until the time of the climatological daily minimum temperature (Figure 2a,c). Once the sun comes up, the sign of the error switches to negative (cool bias) during the day. For OKC, the GFS temperatures are approximately 3°C too high at night and 1°C too low during the day. In contrast, the errors in the HRRR do not yield much of a diurnal cycle in conditions with $\leq 50\%$ cloud cover (Figure 2b,d). Specifically, the median bias throughout the entire day in the HRRR is close to 0°C for both OKC and DTW.

The spatial pattern of errors at 7 a.m. LT when observed cloud cover is $\leq 50\%$ in the GFS indicates that some airports in the northern tier of the US have cold biases (Figure 1c). We considered the role of model snow cover in these errors by extracting the subset of data with an observed cloud cover of $\leq 50\%$ and a model forecast of least a 1-cm snow depth (i.e., snow already on the ground). ASOS does not automatically record snow cover as it is typically augmented by a human observer at select airports (NOAA, 1998). Based on webcam footage, we found that if the model indicated snow depth > 1 cm, then snow cover was usually observed. We examined the full diurnal cycle of errors in Minneapolis, MN (MSP), Sioux Falls, SD (FSD), and Aberdeen, SD (ABR) during conditions of cloud cover $\leq 50\%$ and snow on the ground and found cold biases at all times of day in both the GFS and the HRRR. The data for MSP are shown in Figure 2e,f, which indicates a larger median cold bias at 3 p.m. LT in the GFS (-3.6°C) compared to the HRRR (-1.7°C). Based on these findings, we speculate that some interactions between the snow-covered surface and near-surface temperatures within the models are not being simulated properly.

4 Discussion and Summary

Accurate forecasts in conditions of low sky cover have important practical applications. Errors of a few degrees in forecasts for temperatures near 0°C are especially impactful for aviation and road transportation (e.g., Ballesteros & Hitchens, 2018). Surface overnight low temperatures in winter are crucial for predicting frost formation and morning commute road conditions. Daytime high temperatures in conditions of low sky cover are important for predicting to what extent existing snow and ice will melt. Additionally, planning for de-icing operations for roads and at airports benefits from 36-hour or more lead times.

By conditioning the NWP model error analysis on various cloud cover and wind speed conditions and times of day, we have revealed some major systematic issues in the GFS temperature forecasts that are not as obvious in CONUS daily averages or in case studies. In particular, the GFS struggles to adequately represent the diurnal cycle of temperatures at 36-hour lead times under the mundane condition of $\leq 50\%$ cloud cover. Typically, the GFS is too warm at 7 a.m. LT by 1°C and too cool at 3 p.m. LT by almost 2°C. Overnight errors grow in magnitude when the $\leq 50\%$ cloud cover subset of data is further conditioned by low wind speeds for both the GFS and HRRR. The strong diurnal cycle in errors needs to be considered when making further model improvements. Otherwise, a parameterization refinement indicating an improvement only in daily averages may produce degraded accuracy at different times of the day.

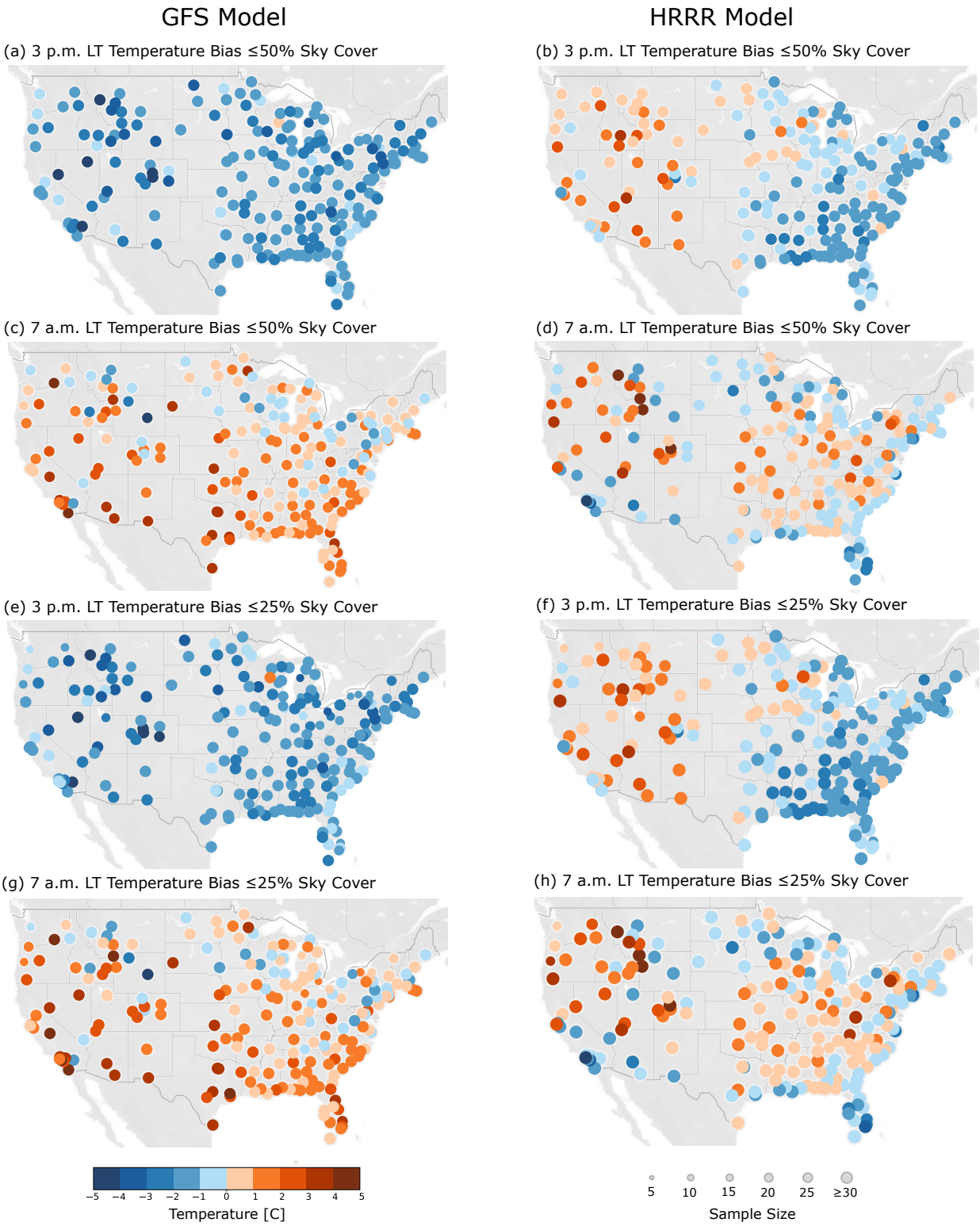


Figure 1. CONUS map of 210 airport sites showing the magnitude and sign of 36-hour lead time temperature biases (model - observation) with conditions of $\leq$ 50% cloud cover (a-d) and $\leq$ 25% cloud cover (e-h) at the times of the winter climatological daily low temperature 7 a.m. LT (c,d,g,h) and the daily high temperature 3 p.m. LT (a,b,e,f). Red shading indicates the model is too warm, and blue shading indicates the model forecast is too cold in the 1 November 2019 to 31 March 2020 time frame. Marker sizes depict sample sizes used in mean bias calculations.

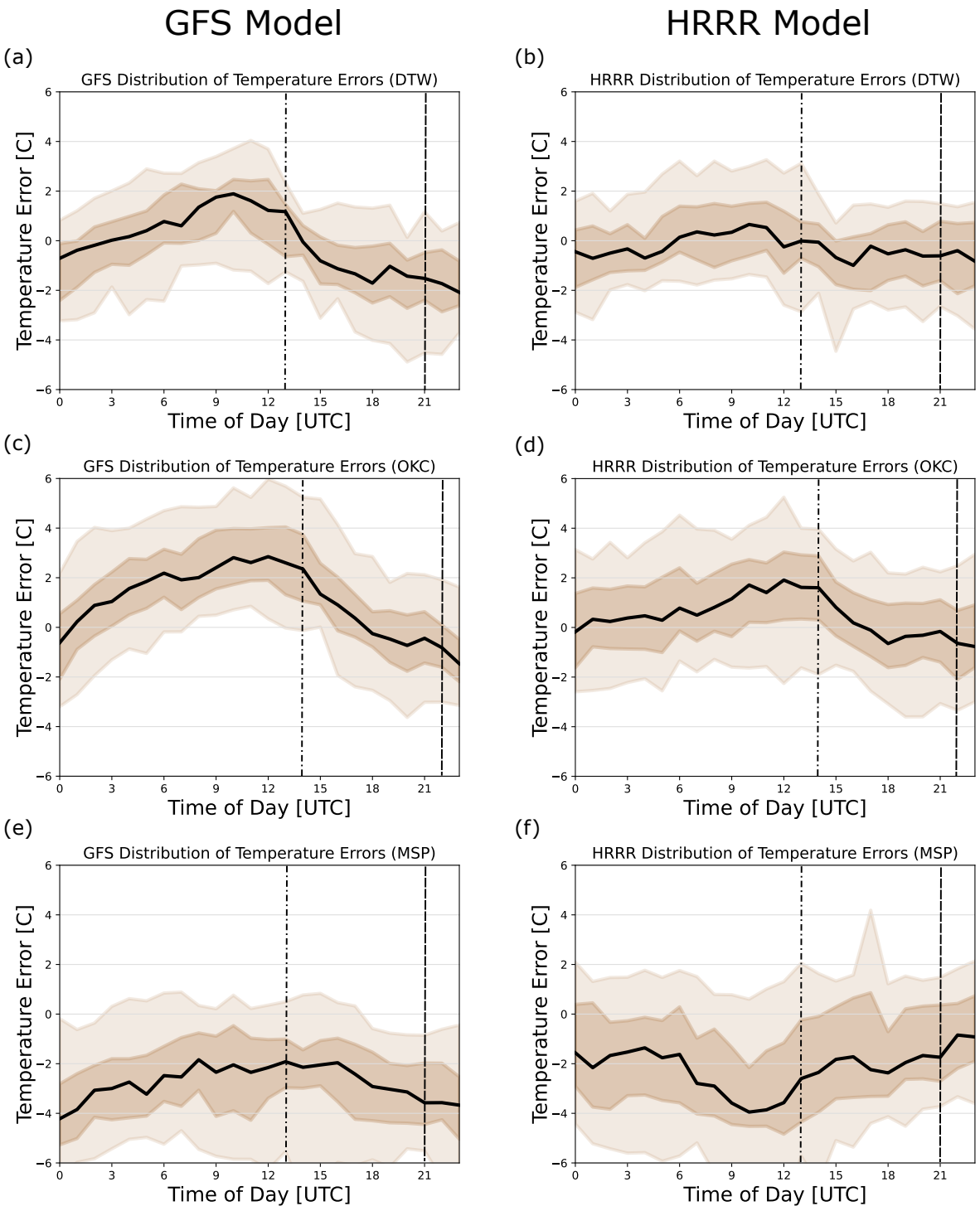


Figure 2. Distribution of temperature errors (model - observed) for conditions with observed $\leq$ 50% cloud cover during the 36-hour lead time by time of day for (a,b) Detroit, MI, (c,d) Oklahoma City, OK, and (e,f) Minneapolis, MN. Errors in the GFS are shown in the left column (in a,c,e) and the HRRR errors are shown in the right column (in b,d,f). The approximate time of the daily low temperature (7 a.m. LT) is indicated by the vertical dash-dotted line, and the approximate time of the daily high temperature (3 p.m. LT) is indicated by a vertical dashed line. Dark beige shading extends from the 25th to 75th percentiles with the median indicated by the solid black line. Light beige shading spans the 5th to 95th percentiles. MSP data (in e,f) were further restricted to hours with forecast snow cover. Sample size distributions vary by location and time of day. Median hourly sample sizes were OKC=65, DTW=29, MSP-GFS=36, and MSP-HRRR=30.

The HRRR does a better job representing the diurnal cycle for $\leq 50\%$ cloud cover conditions and has smaller errors overall at 36-hour lead times. There are regional patterns in HRRR errors, including a daytime-high cool bias in the eastern US, a daytime-high warm bias in the western US, and a nighttime-low cool bias at many coastal sites, that merit further investigation. Fovell and Gallagher (2020) document a warm bias in HRRR 2-m temperature forecasts for 89% of sites located more than 500 m above sea level. Many sites in the Intermountain West are above 500 m in elevation, so their finding is consistent with our observed warm bias in the HRRR during the day and night under clear skies in that region.

Examining errors by similar weather conditions on many days, rather than simple date ranges, helps constrain the portion of model physics in which larger forecast errors are more likely to occur. Strong nocturnal radiation inversions are more common under conditions of light winds and fairly clear skies (Ahrens & Henson, 2019). The pronounced nocturnal warm bias under $\leq 50\%$ cloud cover worsens when observations are further conditioned by low wind speeds. These findings strongly suggest a deficiency in representing nocturnal temperature inversions, which require interactions among radiation, boundary layer, and land surface parameterizations, for the Unified Forecast System medium-range physics used in GFS v15.

While the results presented here address only temperature errors, we demonstrate the utility of examining the diurnal cycle of model errors and conditioning model verification on weather characteristics to identify specific conditions when NWP model errors are larger. A model's ability to accurately represent diurnal variability is a robust test of model physics. Resources for model refinement are limited. These types of analyses can aid in targeting investigations of error sources and revisions to physics packages to ensure models produce the right answer for the right reasons.

Acknowledgments

A rolling 30-day archive of GFS model output is available from NOAA via Amazon Web Services at https://noaa-gfs-bdp-pds.s3.amazonaws.com/index.html. The operational GFS version was used, which is version 15.1 before November 7, 2019 at 12 UTC and version 15.2 after. HRRR v3 model (NOAA,2020) output is available through the University of Utah's HRRR archive (Blaylock et al., 2017). The operational HRRR version 3 was used in this analysis. METAR data are archived within the NOAA MADIS system (https://madis.ncep.noaa.gov/index NOAA maintains an archive of MADIS data, available at ftp://madis-data.ncep.noaa.gov/archive/. Netcdf files of METAR are indexed using the following format:

ftp://madis-data.ncep.noaa.gov/archive/yyyy/MM/dd/point/metar/netcdf/

where yyyy is the 4 digit year, MM is the 2 digit month, and dd is the 2 digit day.

The Open Science Foundation archive at https://doi.org/10.17605/OSF.IO/YTHR2 contains the data presented in figures as csv files and the model and observational data used in the analysis as Parquet files. Christina Cartwright edited the manuscript. This work was supported in partnership with Delta Air Lines and by Office of Naval Research grant N000142112116.

References

Ahrens, C. D., & Henson, R. (2019). *Meteorology today: An introduction to weather, climate, and the environment* (12th ed.). Cengage.

Ballesteros, J. A. A., & Hitchens, N. M. (2018). Meteorological factors affecting airport operations during the winter season in the midwest. *Weather Clim. Soc.*, *10*, 307-322. doi: 10.1029/2010GL042691

Blaylock, B. K., Horel, J. D., & Liston, S. T. (2017). Cloud archiving and data min-

ing of high-resolution rapid refresh forecast model output. *Comput. Geosci.*, *109*, 43–50. doi: 10.1016/j.cageo.2017.08.005

Bu, Y. P., Fovell, R. G., & Corbosiero, K. L. (2017). The influences of boundary layer mixing and cloud-radiative forcing on tropical cyclone size. *J. Atmos. Sci.*, *74*, 1273-1292. doi: 10.1175/JAS-D-16-0231.1

Caron, M., & Steenburgh, W. J. (2020). Evaluation of recent NCEP operational model upgrades for cool-season precipitation forecasting over the western conterminous united states. *Wea. Forecasting*, *35*, 857-877. doi: 10.1175/WAF-D-19-0182.1

EMC Model Evaluation Group. (2019). *FV3GFS (GFSv15.1) Status Update: Addressing Excessive Snowfall and Low-Level Cold Bias.* (Accessed 27 May 2020, `http://www.emc.ncep.noaa.gov/users/Alicia.Bentley/fv3gfs/updates/MEG_4-04-19_FV3GFS_COLD.pptx`)

EMC Model Evaluation Group. (2020a). *GFSv16 MEG Evaluation Overview.* (Accessed 28 September 2020, `https://www.emc.ncep.noaa.gov/users/meg/gfsv16/pptx/MEG_9-24-20_GFSv16_MEG_Eval_Overview.pptx`)

EMC Model Evaluation Group. (2020b). *The MEG Perspective on the Implementation of RAPv5/HRRRv4.* (Accessed 4 February 2021, `https://www.emc.ncep.noaa.gov/users/meg/rapv5_hrrrv4/updates/MEG_2020-12-03_RAPv5_HRRRv4_Implementation.pptx`)

Fovell, R. G., Corbosiero, K. L., Seifert, A., & Liou, K.-N. (2010). Impact of cloud-radiative processes on hurricane track. *Geophys. Res. Lett.*, *37*, L07808. doi: 10.1029/2010GL042691

Fovell, R. G., & Gallagher, A. (2020). Boundary layer and surface verification of the high-resolution rapid refresh, version 3. *Wea. Forecasting*, *35*, 2255-2278. doi: 10.1175/WAF-D-20-0101.1

Haiden, T., Janousek, M., Vitart, F., Bouallegue, Z. B., Ferranti, L., Prates, F., & Richardson, D. (2021). *Evaluation of ECMWF forecasts, including the 2020 upgrade* (Technical Memo No. 880). doi: 10.21957/6njp8byz4

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., & Hansen, B. (2005). Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. *Mon. Wea. Rev.*, *133*, 604–620. doi: 10.1175/MWR-2864.1

Maxson, B. (2019). *Upgrade NCEP global forecast systems to v15.2.0* (Service Change Notice No. 19-84). (`https://www.weather.gov/media/notification/scn19-84gfs15_2.pdf`)

NOAA. (1998). *Automated Surface Observing System (ASOS) user's guide* (User's Guide). (`https://www.weather.gov/media/asos/aum-toc.pdf`)

NOAA. (2020). *The High-Resolution Rapid Refresh (HRRR).* (Accessed 13 November 2020, `https://rapidrefresh.noaa.gov/hrrr/`)

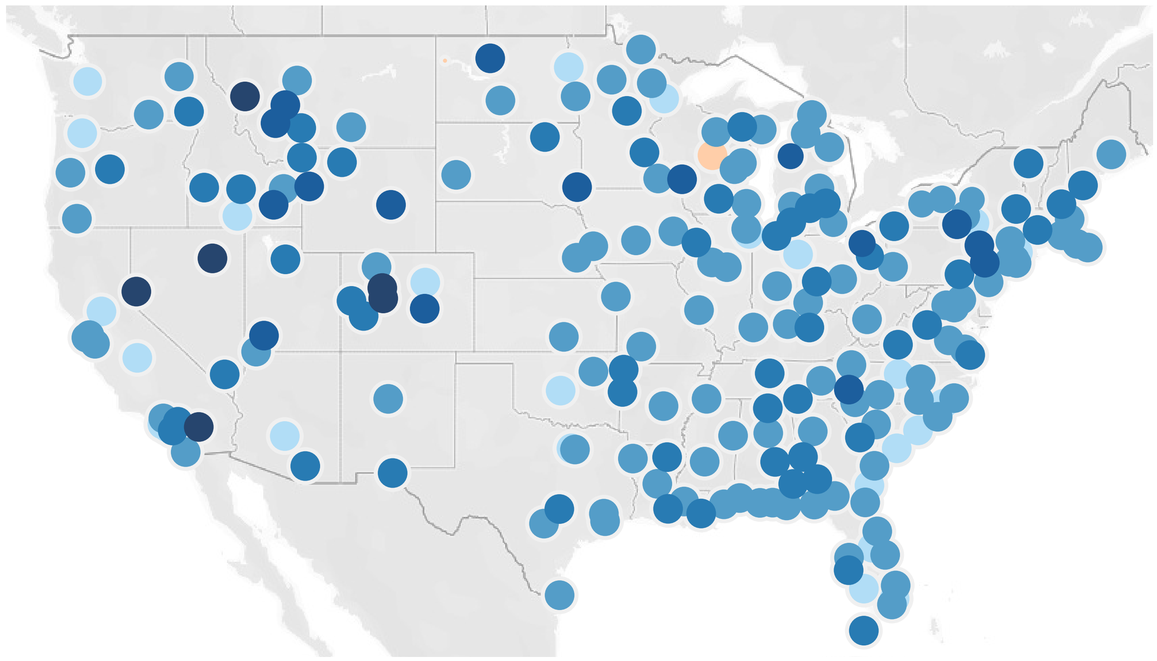
Parker, W. S. (2016). Reanalyses and observations: What's the difference? *Bull. Amer. Meteor. Soc.*, *97*, 1565–1572. doi: 10.1175/BAMS-D-14-00226.1

Figure 1.

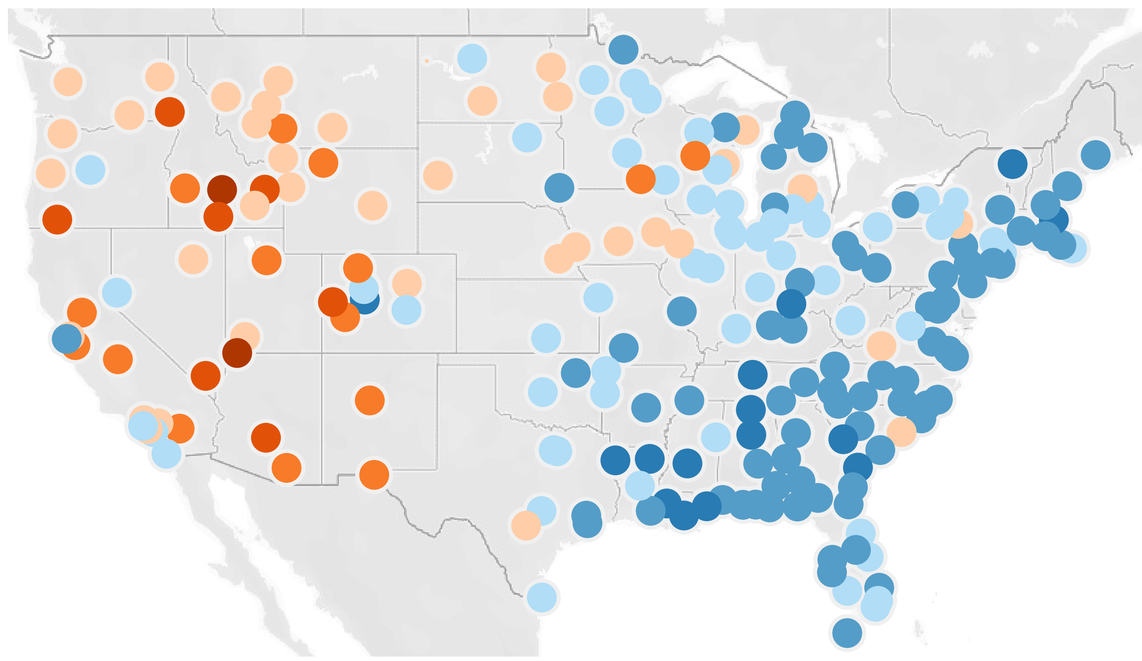
GFS Model

HRRR Model

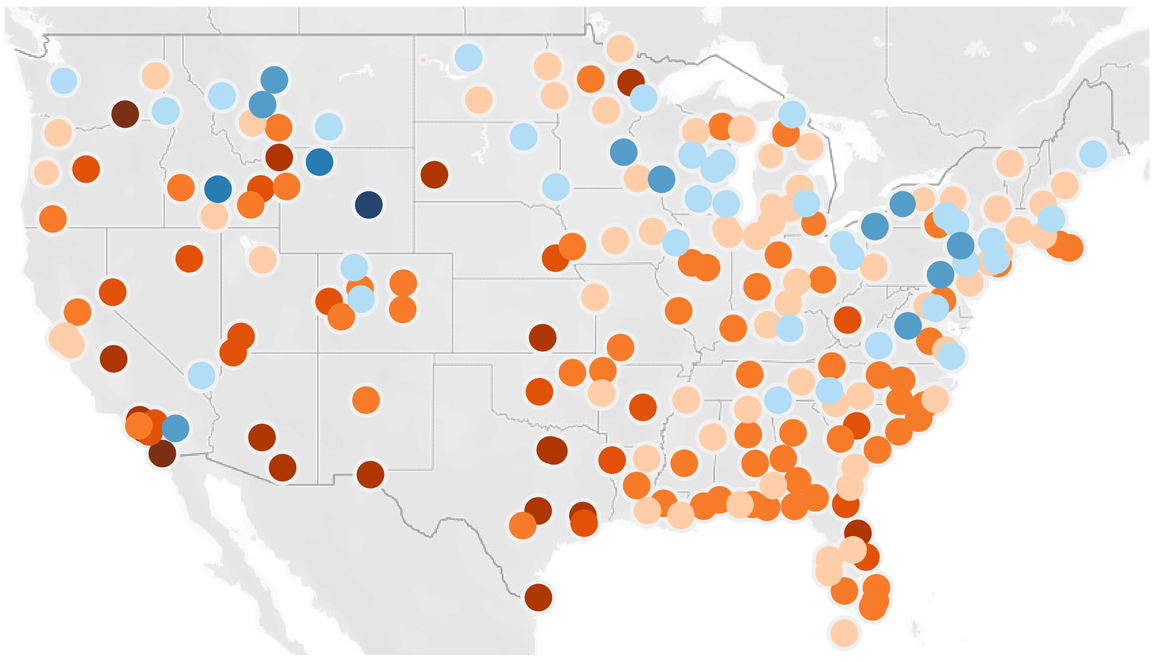
(a) 3 p.m. LT Temperature Bias ≤50% Sky Cover



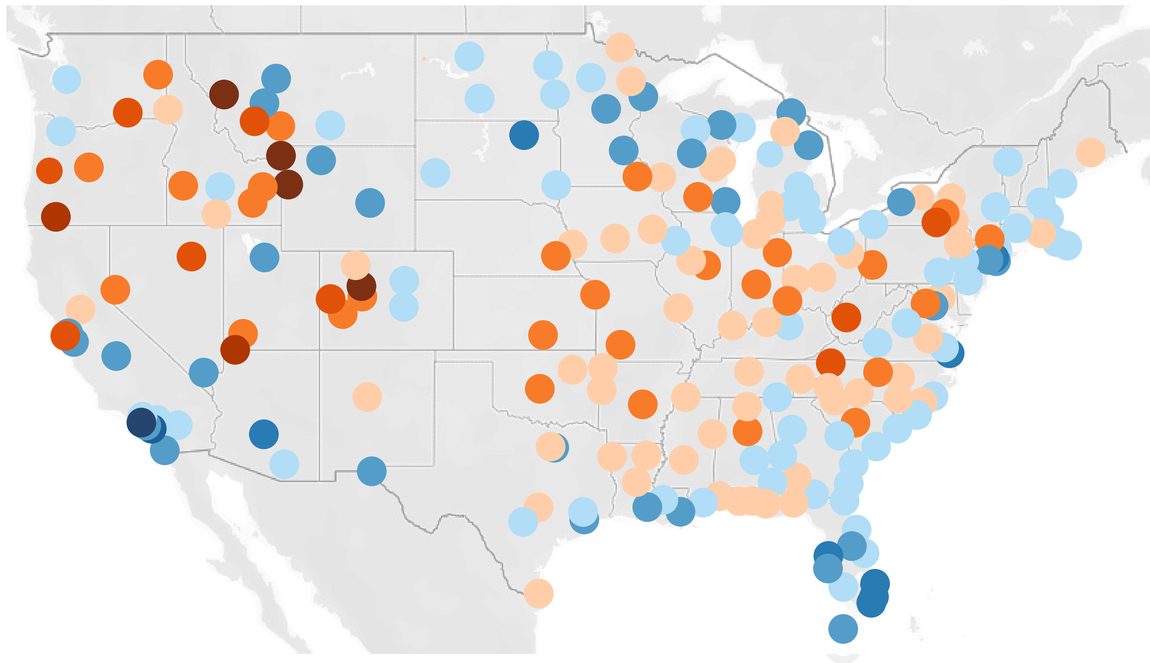
(b) 3 p.m. LT Temperature Bias ≤50% Sky Cover



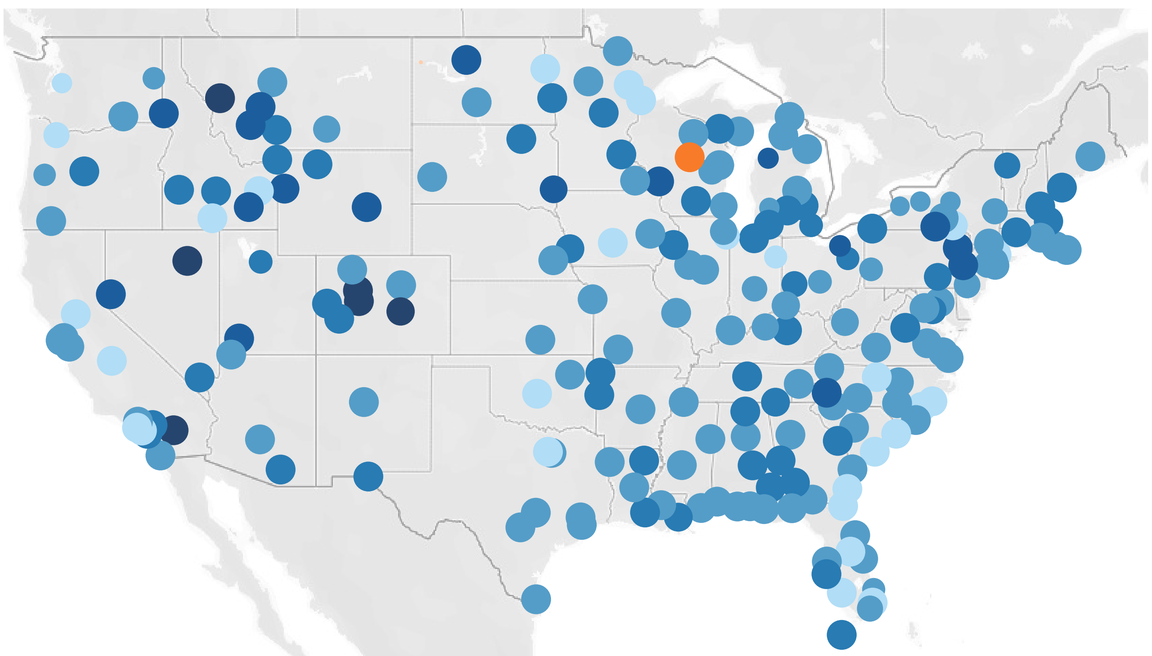
(c) 7 a.m. LT Temperature Bias ≤50% Sky Cover



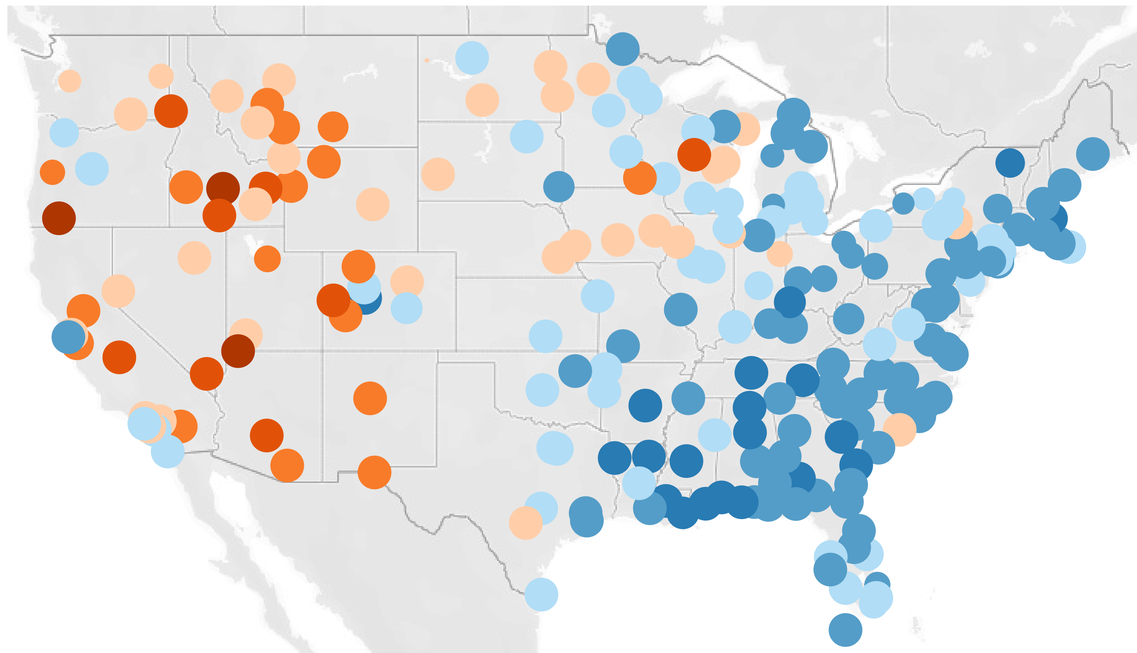
(d) 7 a.m. LT Temperature Bias ≤50% Sky Cover



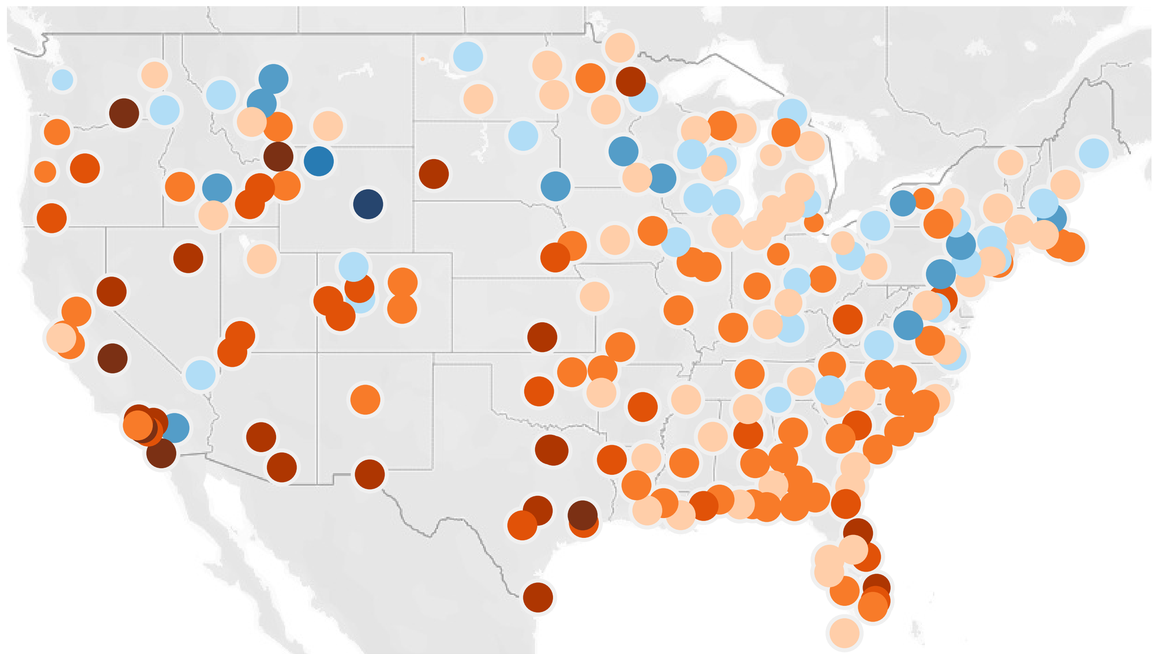
(e) 3 p.m. LT Temperature Bias ≤25% Sky Cover



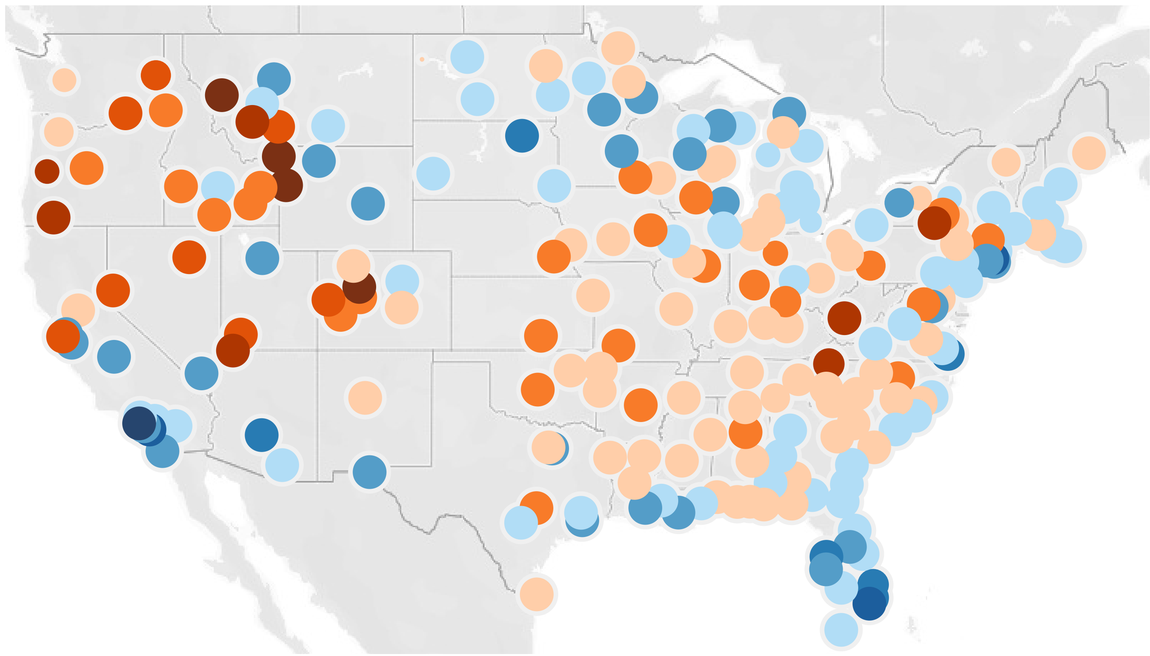
(f) 3 p.m. LT Temperature Bias ≤25% Sky Cover

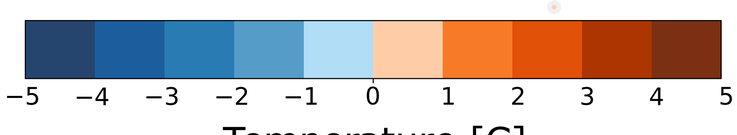


(g) 7 a.m. LT Temperature Bias ≤25% Sky Cover



(h) 7 a.m. LT Temperature Bias ≤25% Sky Cover





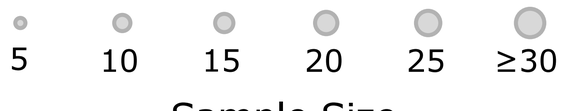
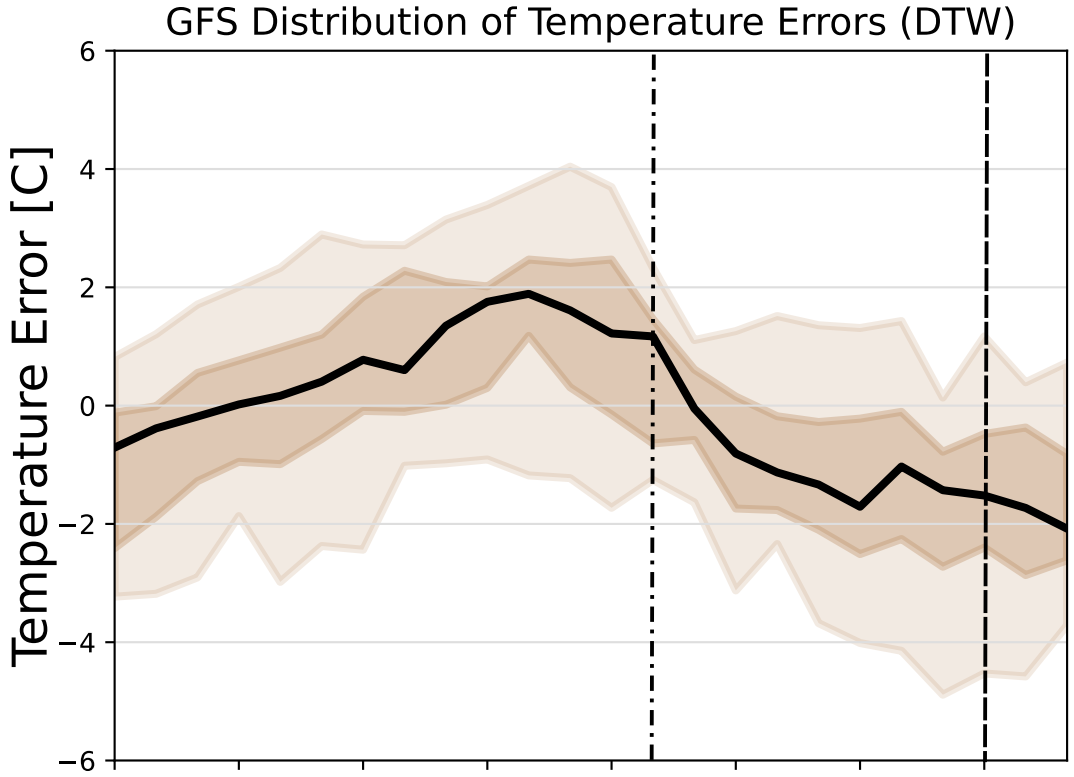


Figure 2.

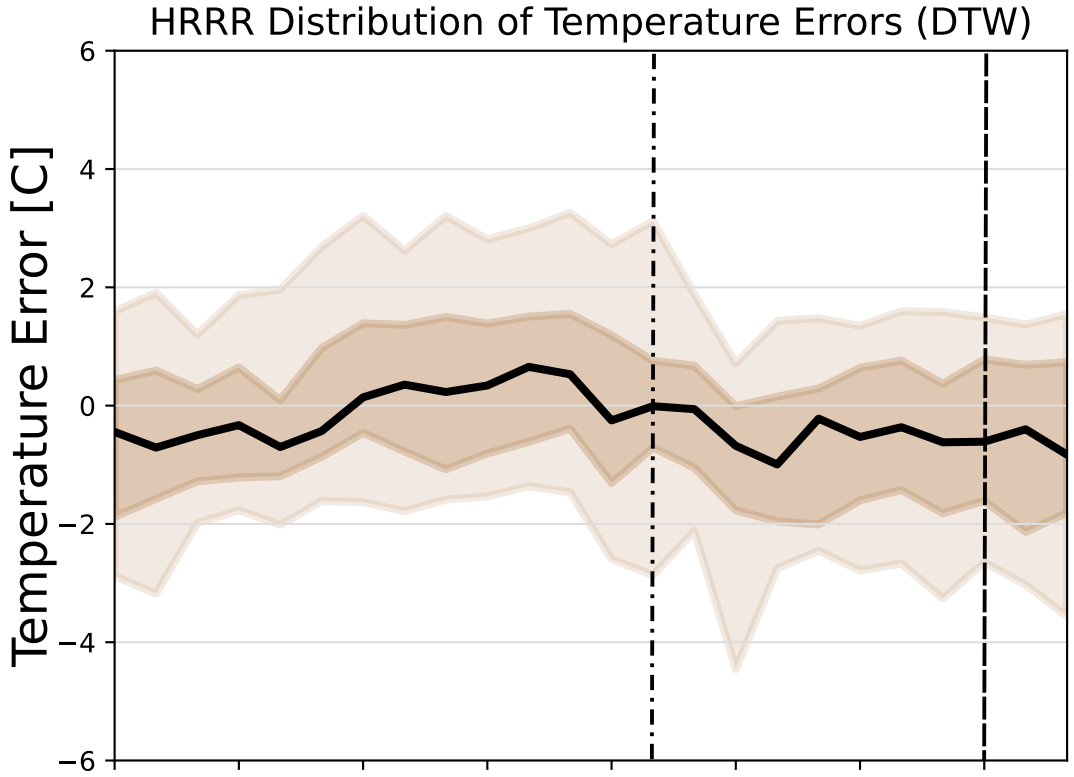
GFS Model

HRRR Model

(a)

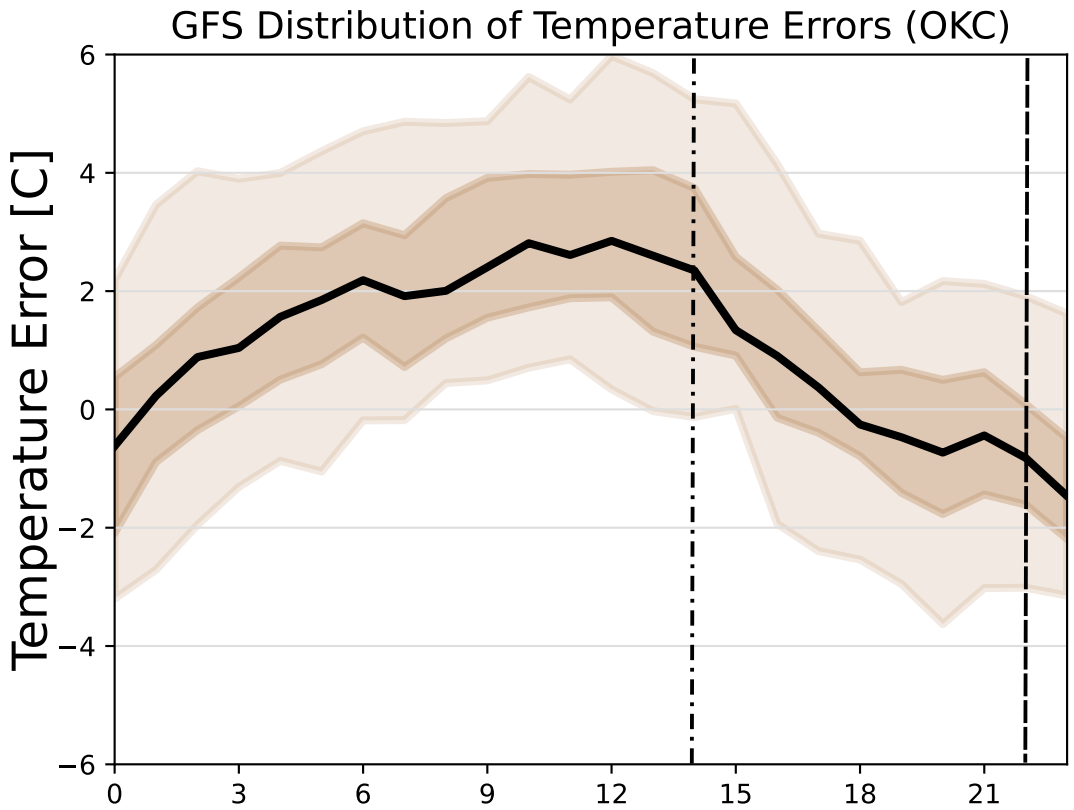


(b)

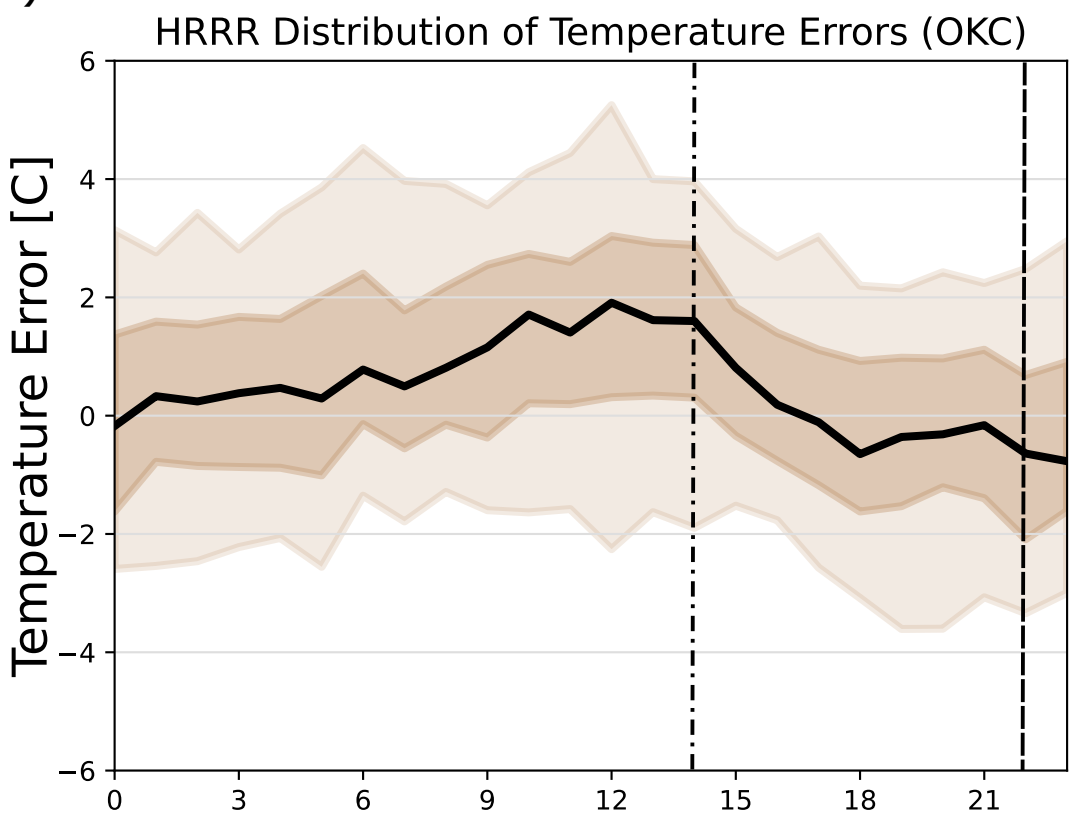


Time of Day [UTC]

(c)

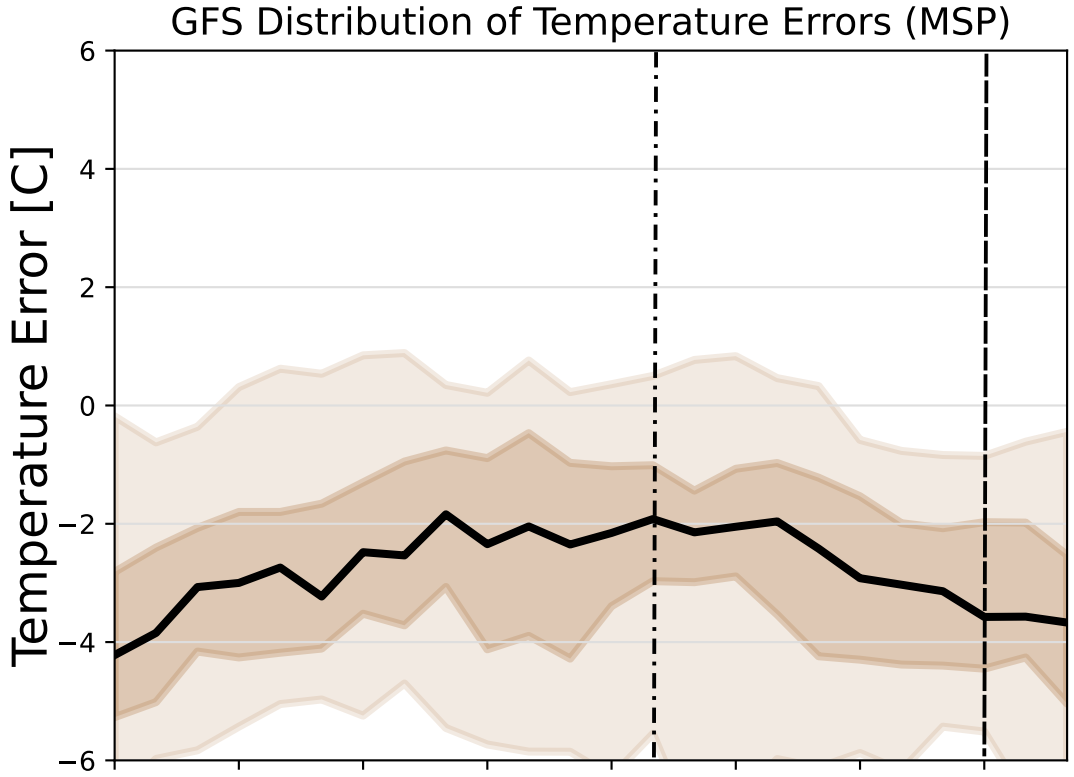


(d)



Time of Day [UTC]

(e)



(f)