# Global inequalities in weather forecasts and the observation infrastructure *

Manuel Linsenmeier† Jeffrey Shrader‡

October 20th, 2023

**Abstract**

Global weather forecasts are of great economic value for society, but geographical differences in forecast accuracy can create new and potentially exacerbate existing economic inequalities. Regional differences in forecast accuracy are particularly relevant if weather forecasts are considered as an important tool to reduce some of the negative effects of future climate change such as mortality from extreme temperature events. In this paper, we provide a comprehensive global analysis of the accuracy of short-term temperature forecasts and relate our findings to existing economic inequalities and the role of the global infrastructure of weather stations. We report three main results: First, temperature forecasts are currently substantially more accurate in high income countries compared to low income countries. The average forecast accuracy for high-income countries is about 19% higher than the average accuracy in low-income countries. Second, while forecast accuracy has improved steadily since 1985, with the largest increases in the 1990s, the gap between high income and low income countries has closed only slightly. Third, similar to the accuracy of global weather forecast, the infrastructure for weather observations is highly unequally distributed across countries, with fewer land-based stations and radiosondes in poorer countries. These inequalities are even larger when lower reporting rates in poorer countries are taken into account.

## 1 Introduction

Weather forecasts can protect lives and provide multiple other benefits to society. The monetary value of these benefits can be enormous. For example, recent evidence on heat mortality (Shrader et al., 2023), labor supply decisions (Song, 2023), and the construction sector (Downey et al., 2023) suggests that economic benefits of accurate weather forecasts exceed multiple times their production costs. The value of weather forecasts critically depends on their accuracy but while this accuracy is routinely assessed by international and national meteorological and hydrological services (NMHS), outside this small community little is known about its temporal and spatial distribution. This is potentially problematic

*Acknowledgments: We are grateful for comments from Adam Sobel.

†Columbia University, Climate School. Contact: mpl2157@columbia.edu

‡Columbia University, School of International and Public Affairs. Contact: jgs2103@columbia.edu

because the methodologies and metrics used for the verification by NMHS are not necessarily sufficient for all relevant questions about weather forecasts (Casati et al., 2008).

Here we provide the first comprehensive analysis of global inequalities in weather forecasts. We focus on near-surface temperature due to the global importance of temperature-related mortality. Furthermore, we focus on forecasts several days ahead because this is the most widely used forecast horizon and because most adjustments to extreme temperatures do not need much anticipation (Shrader et al., 2023). In the last part of the paper, we examine existing inequalities in the infrastructure for weather observations.

No prior work has specifically examined the economic dimension of the distribution of forecast accuracy. Routine verifications usually focus on global averages or specific parts of the world, such as the extratropics, or they provide results with coarse spatial aggregation, such as the Northern and the Southern hemispheres (Haiden et al., 2021). Related prior research also differed in other ways from our study. For example, prior work focused only on trends over time (Magnusson and Källén, 2013), used coarse spatial aggregation or excluded parts of the world (Bauer et al., 2015), focused on longer forecast horizons (Barnston et al., 2010), and generally focused primarily on atmospheric pressure (Bauer et al., 2015) and sometimes rainfall (Wheeler et al., 2017). The work closest to ours is de Perez et al. (2018) who study the predictability of temperature extremes in different parts of the world but focus on forecast horizons of 3 days or more and do not explicitly examine the economic dimension of forecast inequalities.

For our analysis we combine several large dataset. This includes a global dataset of daily temperature forecasts from the digital archives of the ECMWF from 1985-2020 and a global dataset on daily station observations for the same time period from NOAA. We combine these data with economic data to examine the economic dimension of existing inequalities. For the analysis of the observation infrastructure, we also assemble global datasets of land-based weather stations that report sea-level pressure, a global dataset of radiosondes and pilot ballons, and a global dataset of drifting sea bouys. Together these datasets cover the most important components of the global in-situ weather observation system (Haiden et al., 2021).

## 2 Results

We first examine how the accuracy of 1-day-ahead temperature forecasts was globally distributed between 2011 and 2020. Following prior work (Bauer et al., 2015; Haiden et al., 2021) we quantify the accuracy of forecasts as the correlation of the anomaly of the forecast with the anomaly of the observation. All anomalies are calculated using climate normals for the period 1991-2020 from ERA5 reanalysis. To match forecasts with observations and reanalysis, we match every station with the closest grid cells based on its distance from all grid cell centroids. We then calculate the forecast accuracy separately for every station and every year, average over the years 2011-2020, and aggregate the results from stations to the country-level as described in the section Methods. To examine the economic dimension, we use data on GDP per capita at purchasing power parity and country income groups from the World Bank.

We find substantial variation in forecast accuracy across countries (Figure 1a). Forecasts generally tend to be less accurate in tropical countries, but also countries at high latitude in the Northern hemisphere exhibit low accuracy. Overall, countries with higher GDP per capita tend to have more accurate forecasts (Figure 1b). In terms of income groups, 1-day-ahead weather forecasts tend to be most accurate in high-income countries and least accurate in low-income countries (Figure 1b). The differences between countries are large also relative to the decline of forecast accuracy as the forecast horizon increases. For example, 1-day-ahead forecasts in low-income countries tend to be substantially worse than 5-days-ahead forecasts in high-income countries (Figure 1c).

These patterns are robust to using a different way of aggregating results from individual stations to countries (SI Figure S2b). We also obtain similar results if we restrict the sample to the stations with continuous reporting between 1985 to 2020 (SI Figure S2c). (SI Figure S2c). The pattern is also very similar if we use forecasts from the Global Forecast System (GFS) of NOAA, another popular global weather model (SI Figure S3a). We focus on air temperature in 2 metres because of its relevance for human activity, but we find a similar pattern for forecasts of surface pressure (SI Figure S3b).

Weather forecasts have become substantially more accurate over the last decades due to improvements in the availability and use of information, especially from satellites, a better understanding of physical processes, and increases in computing power (Bauer et al., 2015). This improvement since 1991 has been slightly larger in poorer countries than in richer countries (Figure 2a). The gap between low income and high income countries has accordingly decreased by 4 points from 1991-2000 to 2011-2020 to a value of 14.5 points (Figure 2b). While forecast accuracy increased more in the earlier decade than in the later decade, the convergence has been slightly more pronounced in the most recent decade (Figure 2a,b).

We also explore heterogeneous trends in different parts of the world. Consistent with prior work (Bauer et al., 2015), we find large improvements in forecast accuracy in both hemispheres (Figure 2c) and in both the tropics and extratropics (Figure 2d). These improvements were generally larger in the earlier decades. Futhermore, we find that forecasts improved only marginally more in the Southern hemisphere than in the Northern hemisphere

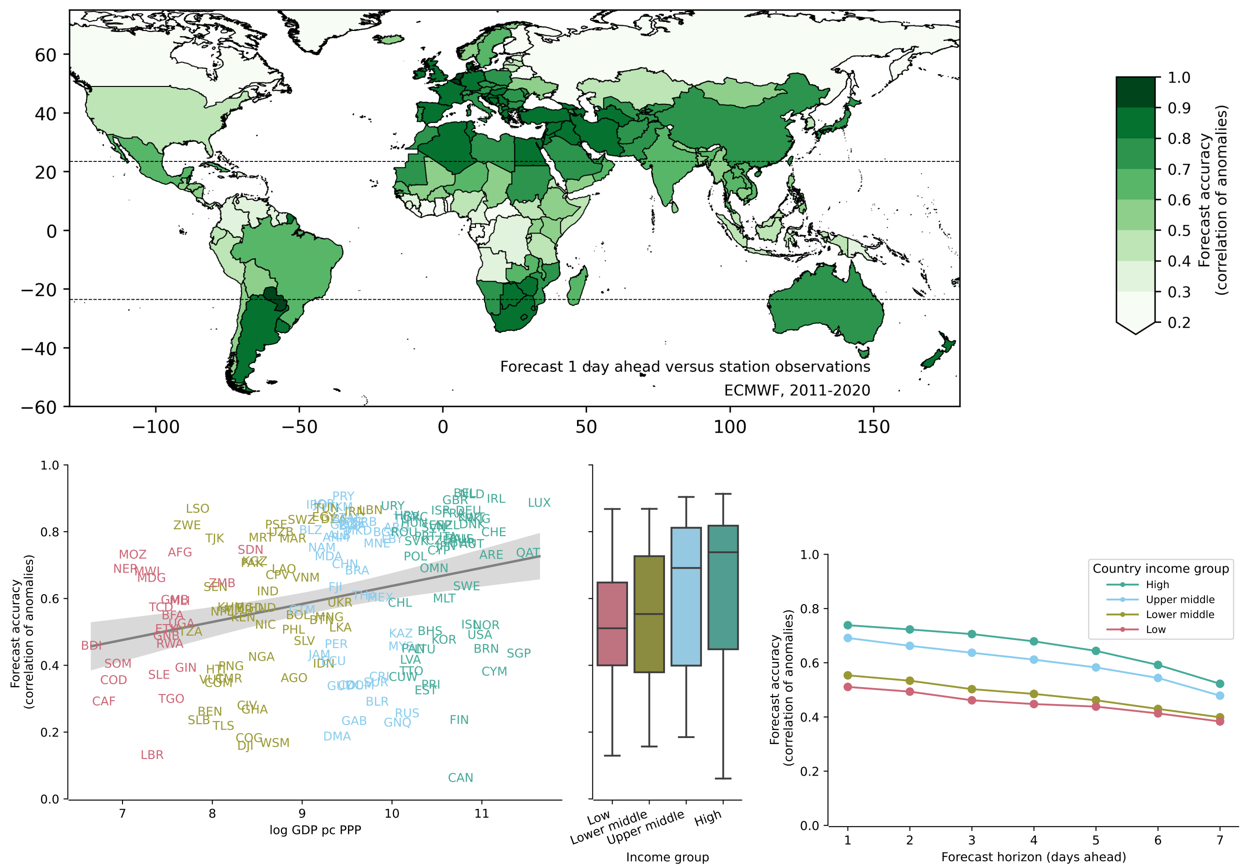


Figure 1. **Countries with higher GDP per capita tend to have more accurate weather forecasts**. All figures are based on an evaluation of daily forecasts from the ECMWF of air temperature in 2 metres above ground against station observations. Forecast accuracy is quantified as correlation of anomaly. Unless otherwise noted, 1-day ahead forecasts are used. Mean values over the period 2011-2020. a. Map of the geographical distribution of forecast accuracy. b. Scatter plot and linear fit with 95% confidence intervals of forecast accuracy and log GDP per capita. Colours indicate country income group of the World Bank. Boxplots show median values, interquartile ranges, and full ranges of values for different income groups. c. Median forecast accuracy for different forecast windows. See SI Figure S1 for a map of country income groups.

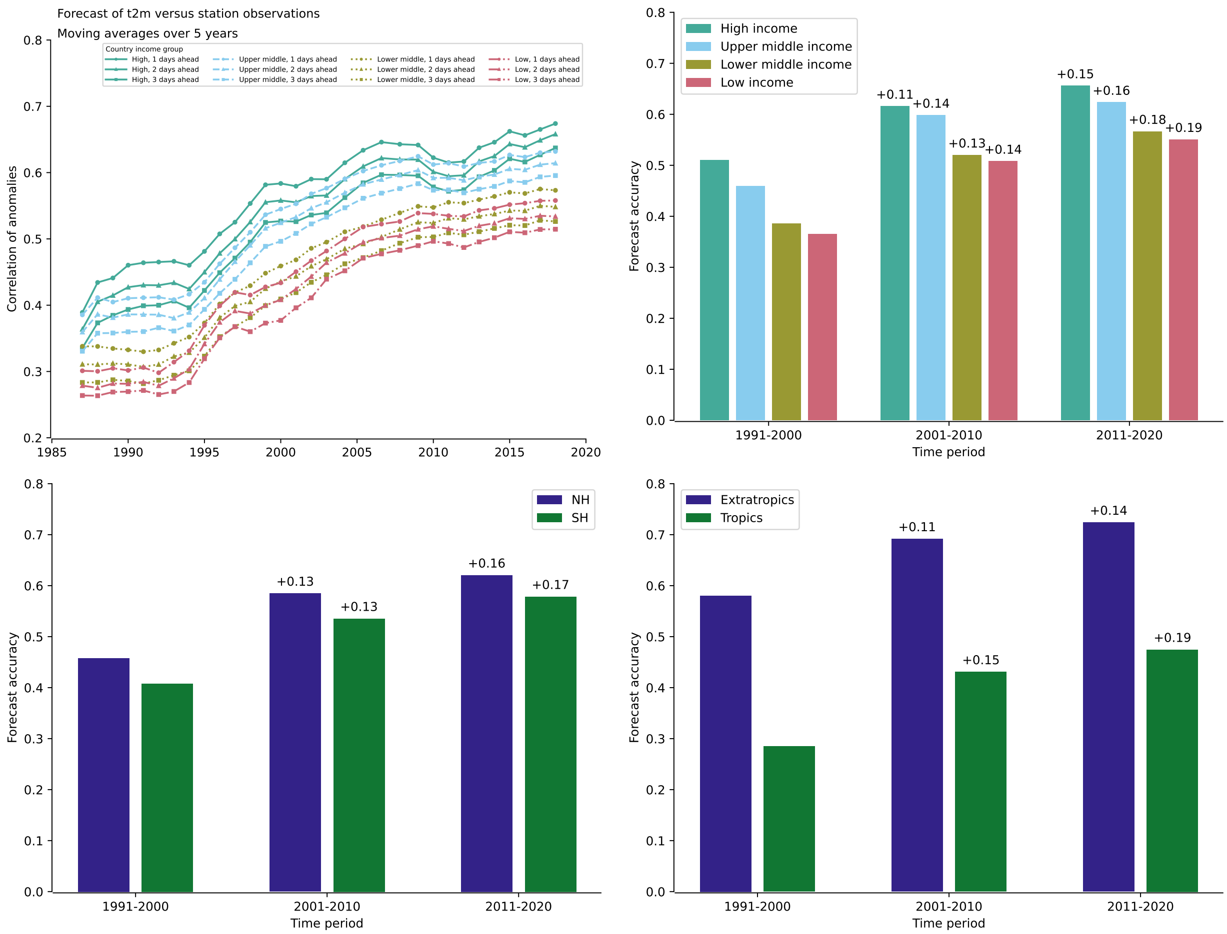


Figure 2. **Forecast accuracy has improved over time everywhere, but persistent differences remain**. The data are prepared in the same way as for for Figure 1. Unless otherwise noted, 1-day ahead forecasts are used. a. Timeseries of forecast accuracy for different country income groups and for different forecast windows, using a five-years running mean. b. Forecast accuracy by country income group and time period. c. Forecast accuracy by hemisphere (NH = Northern hemisphere, SH = Southern hemisphere) and time period. d. Forecast accuracy by world region (tropics, extratropics) and time period.

(Figure 2c), but substantially more in the tropics than in the extratropics (Figure 2d). Furthermore, in the tropics forecasts 1-days-ahead improved more than forecasts 3-day-ahead (SI Figure S5c). In the extratropics, larger improvements can be observed for the 3-days-ahead forecast than for the 1-day-ahead forecast (SI Figure S5d).

Our finding of almost no convergence between the two hemispheres is in contrast to prior work that reported evidence for a more substantial covergence (Bauer et al., 2015). Notably, we find similar evidence of convergence if we validate weather forecasts with the associated analysis, as was done in prior work, but not if the forecasts are validated against station observations (SI Figure S4). For all examination of trends, we use the same set of about 3500 stations that reported in all years. Furthermore, to make the results comparable, for the validation with analysis, we extract the analysis for each station location. Our results therefore suggest that the previously reported convergence between the two hemispheres may be due to the fact that weather forecasts and the associated analysis became more similar without proportional improvements in actual forecast accuracy. This may be especially so in the Southern hemisphere with less in-situ observations where the assimilation of satellite data had a stronger effect on analysis and forecasts than in the Northern hemisphere.

In the last part of the analysis, we examine inequalities in the observing system that can potentially explain the persistent inequalities in forecast accuracy. We focus on the period 2011-2020 and the geographical distribution of land-based weather stations that reported sea-level pressure, radiosonde and pilot balloons, and drifting sea buoys that operated during this time. For sea-buoys, we aggregate monthly location data to ocean hexagons and then count the mean number of buoys on ocean hexagons within 1000 km of all coastal land hexagons of a country. We find that the infrastructure for land-based weather stations and the other two types of observations are highly unequally distributed (Figure 3a; SI Figure S6a-c). The inequality of their distributions can be quantified with the Gini coefficient of observations per land area in different countries. We find the most unequal distribution for land-based stations (Gini = 0.95), followed by radiosondes and pilot balloons (0.94), and sea buoys (0.50).

Similar to the accuracy of weather forecasts, we find that the infrastructure of land-based stations and radiosondes tends to be more sparse in poorer countries (Figure 3b, c). For sea buoys, we do not find a significant relationship between their density close to the coast and the GDP per capita of a country (Figure 4a). Observation infrastructure is only useful if measurements are regularly made. To examine the availability of data, we count the average number of times per year on which land-based stations reported a value at 12 UTC and 0 UTC. Similarly, we count on how many days per year a radiosonde reported information. The results reveal that for both land-based stations and radiosondes poorer countries tend not only to have fewer stations, but also fewer observations per station (Figure 4a).

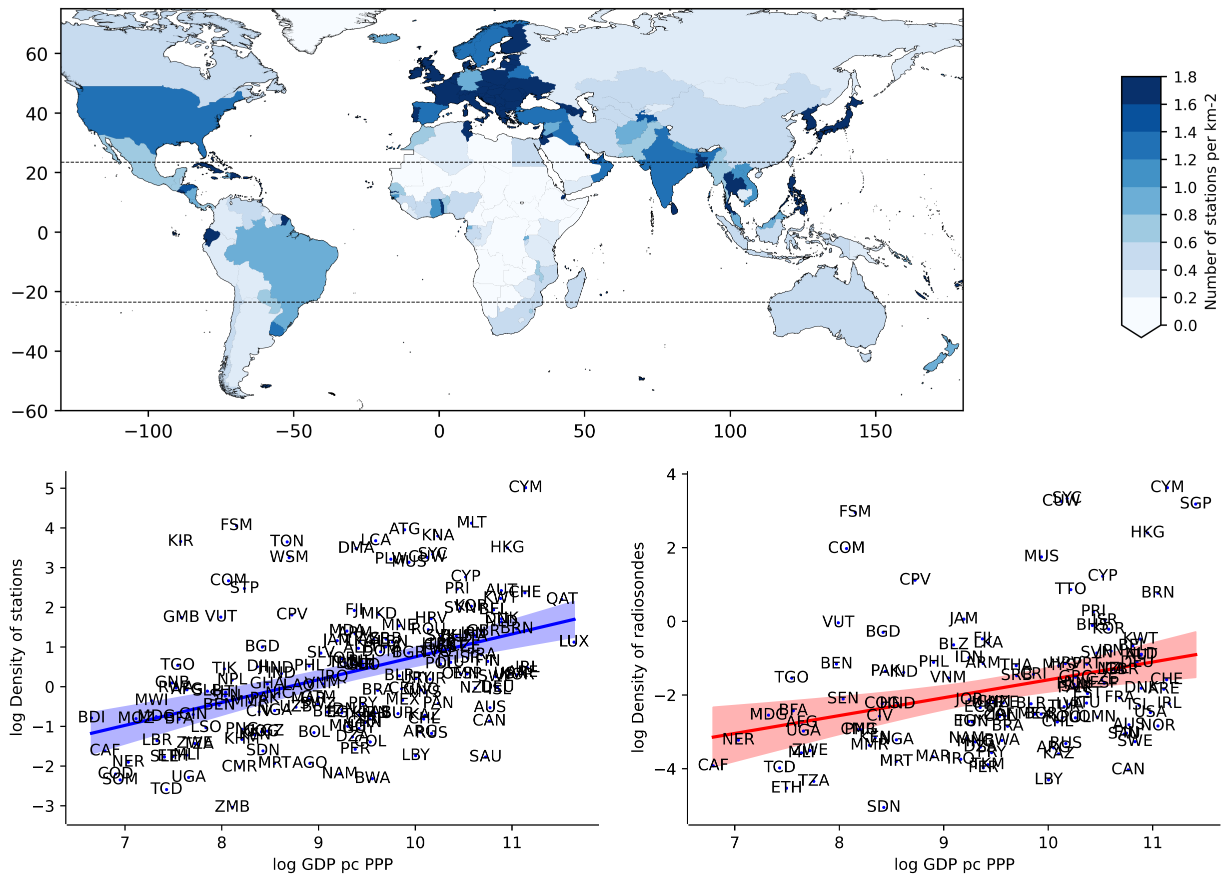


Figure 3. **The infrastructure for weather observations is highly unequal.**. Data on land-based weather stations as well as radiosondes and pilot balloons from NOAA. All figures show results for mean values over 2011-2020. a. Map of the density of land-based weather stations that reported sea-level pressure. b. Scatter plot of the density of land-based weather stations and GDP per capita of countries. c. Scatter plot of the density of radiodondes and pilot ballons and GDP per capita of countries. Errorbars show 95% confidence intervals obtained from heteroscedasticity-robust standard errors.

The predictions of numerical global weather models are an important source of information for national hydrological and meteorological agencies when they issue official forecasts. Because of limited capacity, not only agencies issue and disseminate their own forecast. To examine inequalities in the issuance of official forecasts, we download official national weather forecasts for countries' capitals from the WMO. The data show that about 80% of high-income countries issue an official forecast and report it to the WMO, but only about 20% of low-income countries do so.

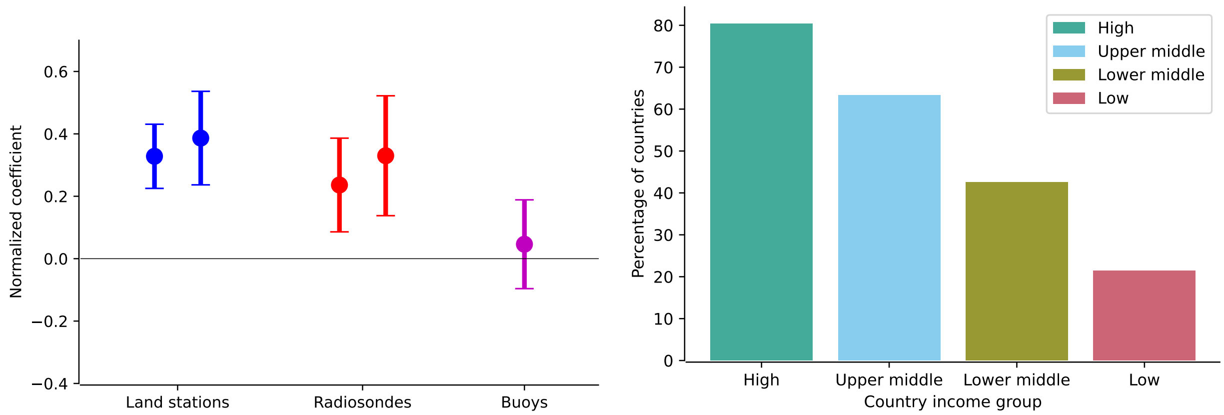


Figure 4. **Richer countries tend to have higher reporting rates of their observation infrastructure and they are more likely to issue official national forecasts than poorer countries.**. a. Linear regression coefficients illustrating the statistical associations between the density of the infrastructure or the frequency of observations and the GDP per capita of a country. b. Percentage of countries that report an official national forecast for their capital to the WMO. Source: WMO.

## 3 Discussion

In this paper, we provide a comprehensive analysis of the global distribution of the accuracy of short-term forecasts of near-surface temperature. We combine our results with economic data to illuminate the economic dimension of existing inequalities. The paper focuses on air temperature because of its relevance for population health (Gasparrini et al., 2015; Carleton et al., 2022), economic output (Katz and Lazo, 2011), and because forecast accuracy might play an important role in mitigating climate change impacts (Shrader et al., 2023). Our work differs from operational forecast verification reports of meteorological agencies and the prior peer-reviewed literature with its detailed analysis of the economic dimension of forecast accuracy and its analysis of inequalities in the observing infrastructure.

(Tzachor et al., 2023) radar, early warning systems

Global variation in forecast accuracy arises from differences in the pysically constrained predictability of weather (Goddard et al., 2001; Zhang et al., 2019), differences in the ability of models to approximate the most relevant physics (Magnusson et al., 2019), and differences in the quantity and quality of weather observations (Žagar, 2017). The importance of weather observations is routinely assessed with data-denial experiments or observing system experiments (Kelly et al., 2007; Bormann et al., 2019). The results of such computationally demanding assessments, which simulate counterfactual weather forecast that ignore some types of observations, suggest that satellites are most influential for forecast accuracy (Haiden et al., 2021) and advances in satellite data assimilation have been associated with large improvements in forecast accuracy especially in the late 1990s (Magnusson and Källén, 2013). However, the importance of different sources of observations differs by forecast horizon (Bormann et al., 2019) and by region of the world (Magnusson et al., 2019). Of the three types of observations considered in this paper, drifting sea buoys are usually considered as most important for forecast accuracy, followed by radiosondes, followed by land-based weather stations. Given that all types of observations contribute to higher accuracy, reducing inequalities in the observing infrastructure can be expected to reduce some of the inequalities in forecast accuracy. Better observations improve weather forecasts also elsewhere (Magnusson, 2017), which should be taken into account in any quantification of the benefits of better observations. Our results also suggest that some inequalities can already be reduced by increasing the recording and dissemination of data of the existing infrastructure (Ingleby, 2015; Dinku, 2019).

## 4 Methods

We use daily weather forecasts from the European Center for Medium-Range Weather Forecasts (ECMWF) from 1985-2020. The ECMWF forecast is widely considered to be the most accurate global, numerical weather forecast. We focus on the validity time 12 UTC and download all weather forecast initialised 0, 24, 48, 72, 96, 120, 144, and 168 hours before a given date, which we refer to as "X-day-ahead" forecast (X = 1 for 24 hours, X = 2 for 48 hours, etc.). The forecast initialised at the time of validity is also referred to as analysis. We use this analysis as an alternative to station observations for verification of the forecast. As a robustness test, we also download for some years forecasts with the validity time 0 UTC and find essentially identical results.

We combine these forecasts with the world's largest dataset of historical weather observations from land-based measurement stations from NOAA (Smith et al., 2011). We use the station data for the verification of the forecast and for an analysis of inequalities in the observation infrastructure. For the verification, we use all stations that reported temperature in 2 metres. For the analysis of observation infrastructure, we use all stations that reported sea-level pressure because in contrast to air temperature this variable has always been assimilated in the ECMWF forecasts.

For the examination of mean differences between countries over the period 2011-2020 we use all available stations. For the examination of trends, we select only those stations that reported at least once every year between 1985 and 2020. To match forecasts with station observations, we identify for every year and for every station the grid cell of the gridded forecast data that is closest in space based on the coordinates of the station and the coordinates of grid cell centroids.

For the verification of forecasts, we calculate for every station and for every year the correlation of the anomaly of the forecast with the anomaly of the observation. We calculate these anomalies by subtracting the climatological mean value of the period 1991-2020. The use of anomalies avoids that locations with large seasonality tend to have larger correlations simply because of that seasonality. The climatologies are obtained from ERA5 reanalysis using the closest grid cell of each station. We use this climatology also to calculate for every station a climatological between-year standard deviation of the period 1991-2020. We use this standard deviation to identify possibly erroneous station measurements. Specifically, we ignore all station observations that deviate by more than five standard deviations from the forecast.

For radiosondes and pilot balloons we use the Integrated Global Radiosonde Archive from NOAA. We extract all historical locations from the metadata file and process the raw data with actual observations to identify the number of ascensions for every location for every year. For drifting buoys, we use the archive of NOAA's Global Drifter Program. For every buoy, we calculate its monthly mean coordinates and then aggregate all buoys to ocean hexagons by counting the mean number of buoys in every hexagon in every year. We then select all coastal land hexagons and calculate for every year the mean number of buoys using all ocean hexagons within 1000 km based on centroid-to-centroid distances.

We combine our meteorological data with economic data from the World Bank. For national income, we use data on GDP per capita in purchasing power parity in constant 2011 international USD. For country income groups, we use the official World Bank classification from 2020. For population, we use the Gridded Population of the World dataset in version 4 for the year 2020.

We use two different ways of aggregating forecast accuracy from individual stations to countries. In our main specification, we first average all stations in the same hexagon and then calculate for every country a weighted average of hexagons using population of hexagons as weights. The use of population as weight is primarily motivated by the existence of few countries with large landmasses where large areas are very sparsely populated, such as Russia and Canada. For robustness, we do not weigh by population and instead assign all hexagons within a country equal weight. The results are qualitatively the same.

## References

Barnston, A. G., Li, S., Mason, S. J., DeWitt, D. G., Goddard, L., and Gong, X. (2010). Verification of the First 11 Years of IRI's Seasonal Climate Forecasts. *Journal of Applied Meteorology and Climatology*, 49(3):493–520.

Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55.

Bormann, N., Lawrence, H., and Farnan, J. (2019). Global observing system experiments in the ECMWF assimilation system.

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Kopp, R. E., McCusker, K. E., Nath, I., et al. (2022). Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits. *The Quarterly Journal of Economics*, 137(4):2037–2105.

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., and Mason, S. (2008). Forecast verification: Current status and future directions. *Meteorological Applications*, 15(1):3–18.

de Perez, E. C., van Aalst, M., Bischiniotis, K., Mason, S., Nissan, H., Pappenberger, F., Stephens, E., Zsoter, E., and van den Hurk, B. (2018). Global predictability of temperature extremes. *Environmental Research Letters*, 13(5):054017.

Dinku, T. (2019). Challenges with availability and quality of climate data in Africa. In *Extreme Hydrology and Climate Variability*, pages 71–80. Elsevier.

Downey, M., Lind, N., and Shrader, J. G. (2023). Adjusting to rain before it falls. *Management Science*.

Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J., Tobias, A., Tong, S., Rocklöv, J., Forsberg, B., et al. (2015). Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386(9991):369–375.

Goddard, L., Mason, S., Zebiak, S., Ropelewski, C., Basher, R., and Cane, M. (2001). Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology*, 21(9):1111–1152.

Haiden, T., Janousek, M., Bouallegue, Z. B., Ferranti, L., Prates, F., and Richardson, D. (2021). Evaluation of ECMWF forecasts, including the 2021 update. *ECMWF Technical Memoranda*, 884.

Ingleby, B. (2015). Global assimilation of air temperature, humidity, wind and pressure from surface stations. *Quarterly Journal of the Royal Meteorological Society*, 141(687):504–517.

Katz, R. W. and Lazo, J. K. (2011). Economic value of weather and climate forecasts. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*. Oxford University Press, Oxford.

Kelly, G., Thépaut, J.-N., Buizza, R., and Cardinali, C. (2007). The value of observations. I: Data denial experiments for the Atlantic and the Pacific: THE VALUE OF OBSERVATIONS. I: DATA DENIAL. *Quarterly Journal of the Royal Meteorological Society*, 133(628):1803–1815.

Magnusson, L. (2017). Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society*, 143(706):2129–2142.

Magnusson, L., Chen, J.-H., Lin, S.-J., Zhou, L., and Chen, X. (2019). Dependence on initial conditions versus model formulations for medium-range forecast error variations. *Quarterly Journal of the Royal Meteorological Society*, 145(722):2085–2100.

Magnusson, L. and Källén, E. (2013). Factors Influencing Skill Improvements in the ECMWF Forecasting System. *Monthly Weather Review*, 141(9):3142–3153.

Shrader, J., Bakkensen, L., and Lemoine, D. (2023). Fatal errors: The mortality value of accurate weather forecasts.

Smith, A., Lott, N., and Vose, R. (2011). The Integrated Surface Database: Recent Developments and Partnerships. *Bulletin of the American Meteorological Society*, 92(6):704–708.

Song, Y. (2023). The value of weather forecasts: Labor responses to accurate and inaccurate temperature forecasts in china.

Tzachor, A., Richards, C. E., Gudoshava, M., Nying'uro, P., Misiani, H., Ongoma, J. G., Yair, Y., Mulugetta, Y., and Gaye, A. T. (2023). How to reduce Africa's undue exposure to climate risks. *Nature*, 620(7974):488–491.

Wheeler, M. C., Zhu, H., Sobel, A. H., Hudson, D., and Vitart, F. (2017). Seamless precipitation prediction skill comparison between two global models. *Quarterly Journal of the Royal Meteorological Society*, 143(702):374–383.

Žagar, N. (2017). A global perspective of the limits of prediction skill of NWP models. *Tellus A: Dynamic Meteorology and Oceanography*, 69(1):1317573.

Zhang, F., Sun, Y. Q., Magnusson, L., Buizza, R., Lin, S.-J., Chen, J.-H., and Emanuel, K. (2019). What Is the Predictability Limit of Midlatitude Weather? *Journal of the Atmospheric Sciences*, 76(4):1077–1091.

# Supplementary Information (SI)

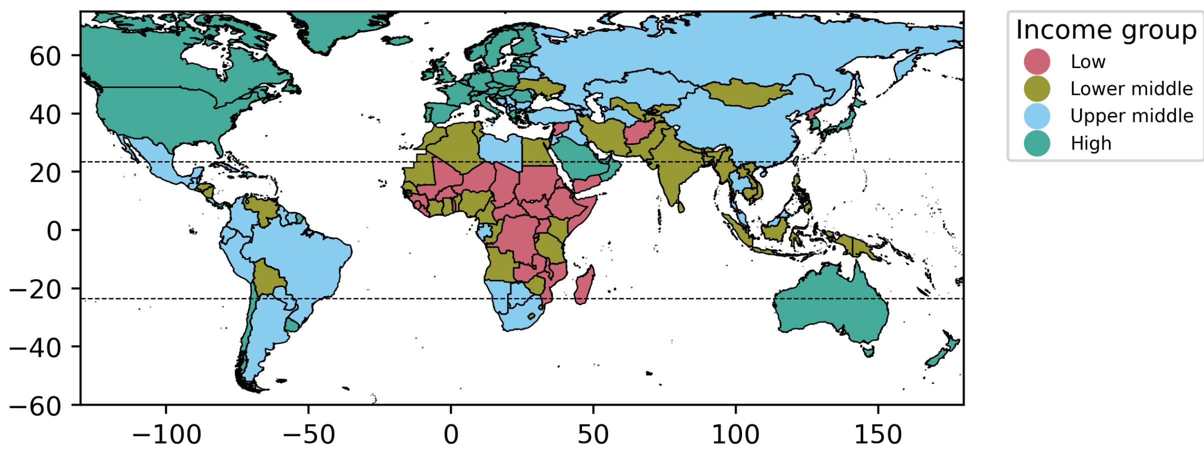


Figure S1. Map of income group of countries. Based on World Bank classification in 2020.

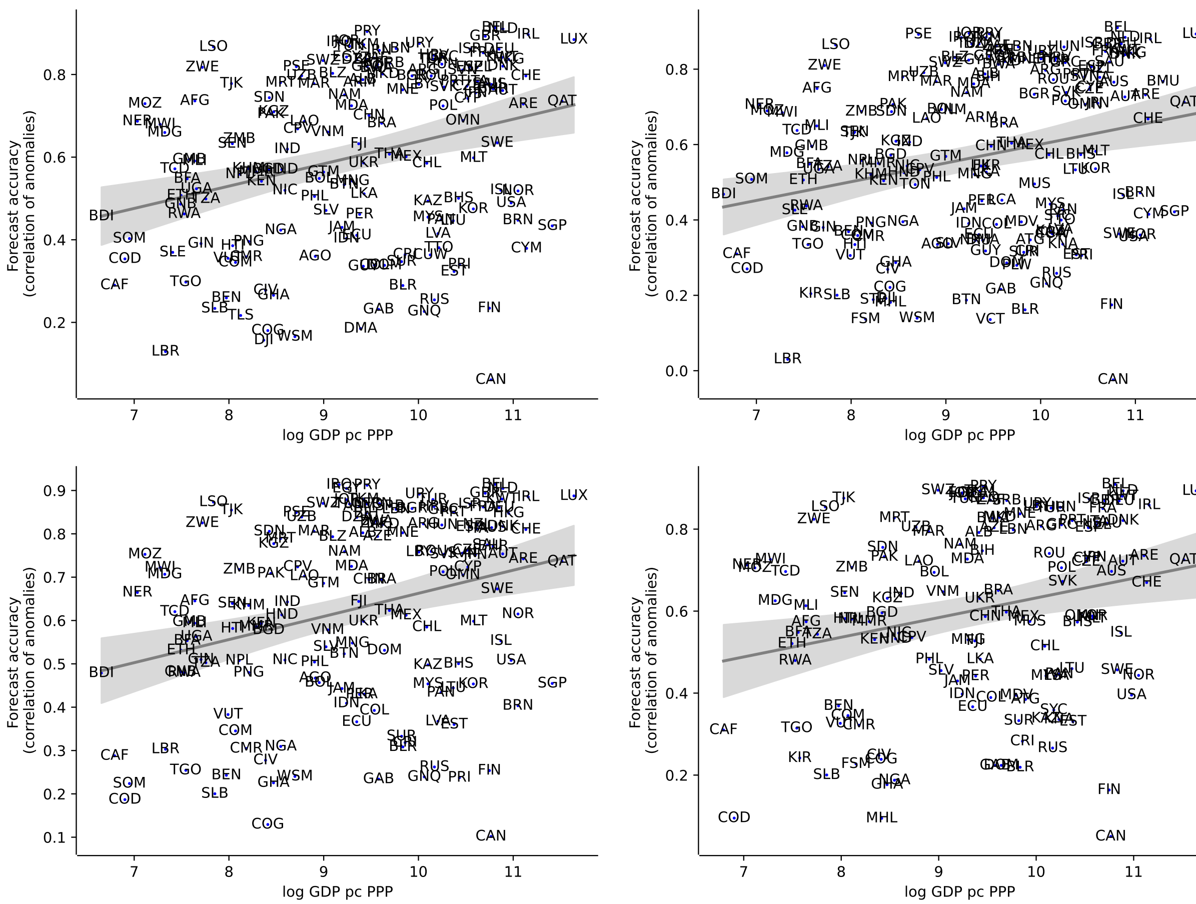


Figure S2. Scatterplot of forecast accuracy and countries GDP pc for different ways of aggregation and for different samples of stations. Figure a is identical to Figure 1b. Figure b shows results aggregated from stations without population weights. Figure c shows results with population weight, but only based on subset of stations with continuous reporting. Figure d shows results without population weights and only for subset of stations.

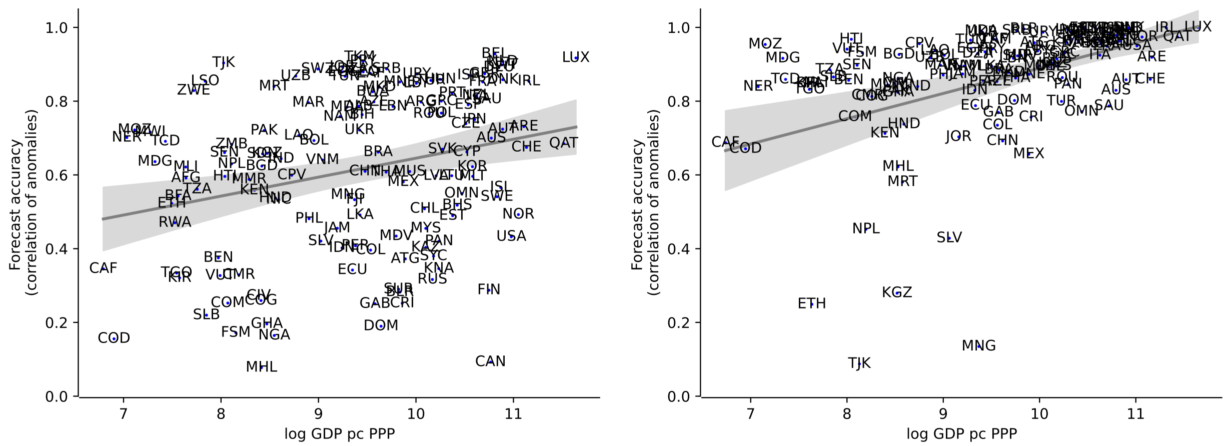


Figure S3. Scatterplot of forecast accuracy and countries GDP pc for different weather forecast model and for different meteorological variable. a. Forecast from GFS instead of ECMWF. b. Forecast of surface pressure instead of air temperature in 2 metres.

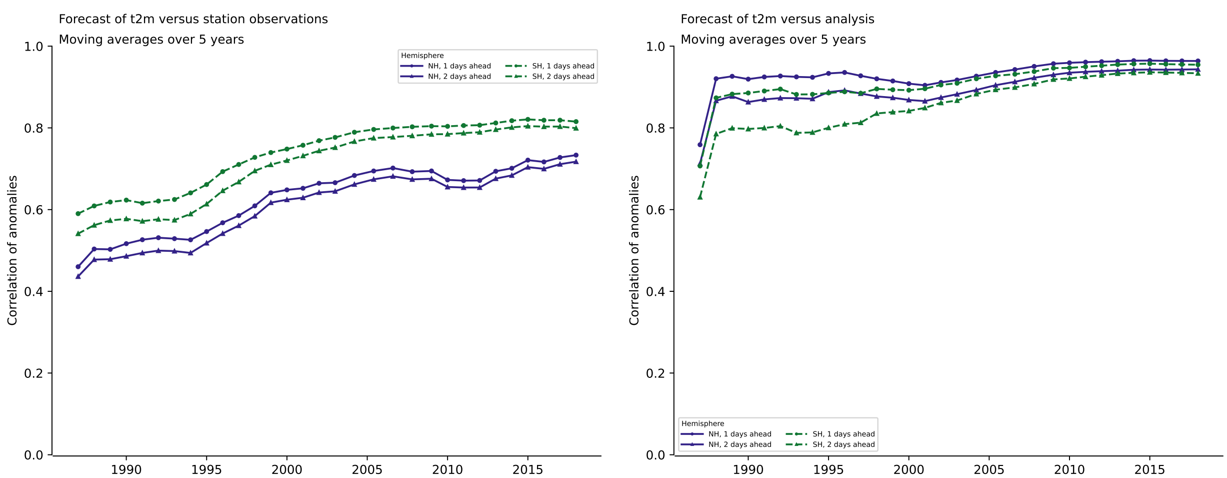


Figure S4. Timeseries of forecast accuracy by hemisphere and by forecast horizon for two alternative verification methods. Verification based on: a. Station observations, b. Model analysis. Both figures show five-years moving averages. NH = Northern hemisphere, SH = Southern hemisphere.

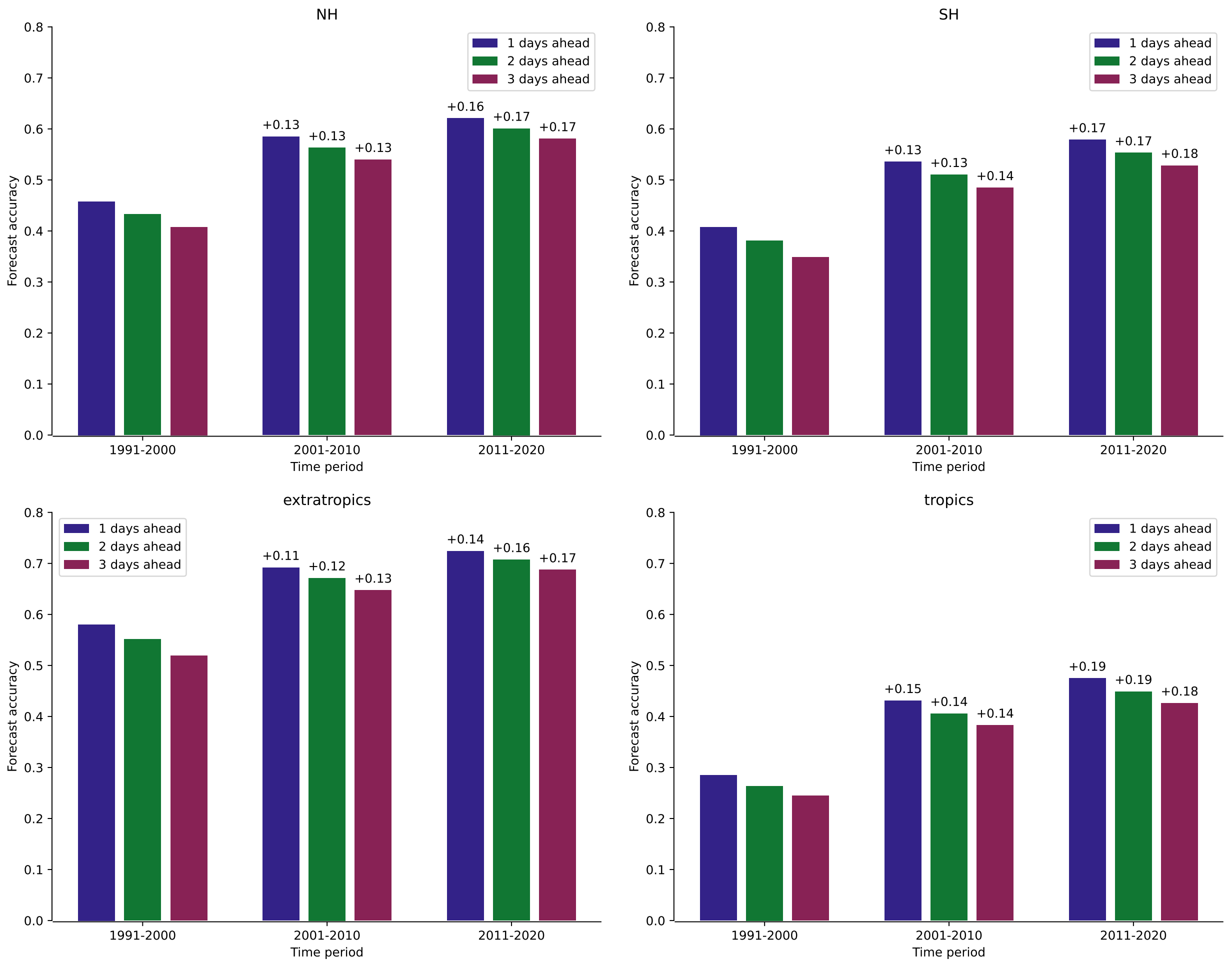


Figure S5. Improvements in forecast accuracy over time by forecast horizon. a. and b. Improvements by hemisphere (NH = Northern hemisphere, SH = Southern hemisphere). c and d. Improvements by region (tropics, extratropics). See also Figures 2c and d.

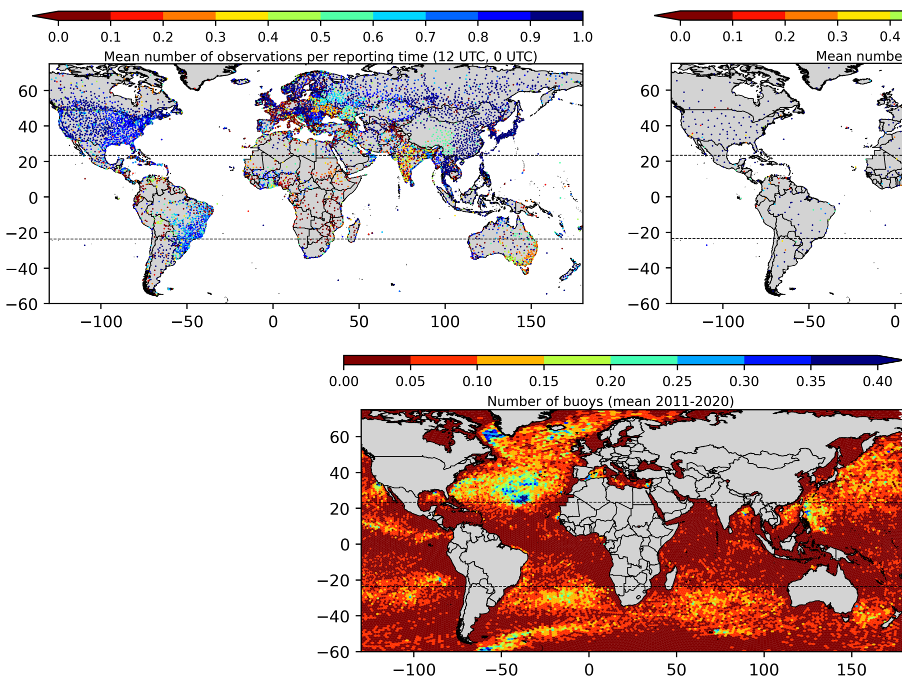


Figure S6. Maps of the observation infrastructure. Figure a shows the location and reporting frequency of land-based weather stations. Figure b shows the location and reporting frequency for radiosondes and pilot ballons. Figure c shows the average number of buoys per hexagon. All maps are generated from data from NOAA covering the years 2011-2020.