Global inequalities in weather forecasts *

Manuel Linsenmeier† Jeffrey Shrader‡

December 7, 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 help reduce 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 numerical weather predictions of temperature and relate our findings to both existing economic inequalities and inequities in global weather monitoring infrastructure. We report three main results: First, temperature forecasts are currently substantially more accurate in high income countries than in low income countries. A seven-day-ahead forecast in a high-income country is on average more accurate than a one-day-ahead forecast in a low income country. Second, while forecast accuracy has improved steadily between 1985 and the present—with the largest increases in the 1990s—there is a persistent gap between high income and low income countries. Third, the infrastructure for weather observations is highly unequally distributed across countries, with fewer land-based weather stations and radiosondes in poorer countries. These inequalities grow even larger when lower reporting rates are taken into account. Remedying these differences in infrastructure would help close the forecast accuracy gap.

1 Introduction

Weather forecasts provide a wide variety of benefits to society, including protecting lives, aiding responses to extreme weather, and improving labor productivity (Shrader et al., 2023; Anand, 2022; Song, 2023). The value of these benefits is large, with recent estimates of the monetized economic benefits from accurate weather forecasts exceeding multiple times their production costs (Shrader et al., 2023). The value of weather forecasts critically depends on their accuracy (Leviäkangas, 2009). This accuracy is routinely assessed by international and national meteorological and hydrological services (NMHS), but the results are included

*Acknowledgments: We are grateful for comments from Tufa Dinku, Thomas Haiden, Dan Osgood, and Adam Sobel. This work received financial support from the Climate School and the National Science Foundation under Grant No. 2315555.

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

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

in technical reports (e.g. Haiden et al. (2021)) which are primarily intended for an expert audience and focus on technical improvements to forecasts rather than the socioeconomic dimension of forecast accuracy (Casati et al., 2008). As a result, outside a small community, little is known about how forecast accuracy is temporally and spatially distributed. This is potentially problematic, because inequalities in accuracy may have large socioeconomic implications. Here we provide the first comprehensive analysis of global inequalities in numerical weather predictions. We focus on short-range temperature forecasts that are highly relevant for economic decision making and for adaptation to climate change, especially for the reduction of temperature-related mortality (Carleton and Greenstone, 2022; Shrader et al., 2023).[1]

Overall, we find large inequalities. Poorer countries generally experience lower forecast accuracy across all forecasting horizons. Furthermore, we find that from 1985–2020, the gap between high-income and low-income countries has only marginally decreased. Mirroring—and potentially underlying—the pattern in forecast accuracy, we find that the density of weather observing infrastructure also exhibits substantial inequality. Land-based stations and radiosondes tend to be more common in richer countries. These inequalities in infrastructure are further exacerbated by lower reporting rates in poorer countries. Infrastructural inequities do not tell the whole story, however. Poorer countries also have lower capacity to translate global numerical weather predictions into local forecasts, suggesting that our results using numerical weather predictions underestimate the differences in quality in official forecasts. Finally, we report evidence from numerical experiments that show that the benefits of additional observations will be particularly large in regions where observations are currently relatively sparse, suggesting that reducing the gap in the global weather observation infrastructure can help close the gap in forecast accuracy.

For our analysis we combine several large datasets. They include a global dataset of daily forecasts of temperature and surface pressure from the digital archives of the ECMWF from 1985-2020 (ECMWF, 2023) and a global dataset on daily land-based station observations for the same time period from NOAA (Smith et al., 2011). We combine these meteorological data with economic data from the World Bank to examine the economic dimension of existing inequalities (World Bank, 2023). For the analysis of the observation infrastructure, we also use global datasets of land-based weather stations (Smith et al., 2011) and of radiosondes and pilot balloons (Durre et al., 2016), as well as a global dataset of drifting sea bouys (Lumpkin and Centurioni, 2010), all provided by NOAA. Together our data cover the most important components of the global in-situ weather observation infrastructure (Haiden et al., 2021).

[1]We also examine inequality in the forecast accuracy of air pressure—a variable that is more commonly used to evaluate numerical weather prediction skill—in supplementary results. See Figure S3.

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) which might not be as valuable for economic decisions (Millner and Heyen, 2021a), and generally focused primarily on atmospheric pressure (Bauer et al., 2015) and sometimes rainfall (Wheeler et al., 2017). The first part of our analysis is closest to 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 overlay forecast accuracy with any economic variable.

2 Results

We first examine the spatial distribution and temporal evolution of forecast accuracy. We focus on air temperature at two metres because of its relevance for many socioeconomic outcomes (Carleton and Hsiang, 2016), but in robustness tests we also analyse forecasts of surface pressure. Furthermore, we focus on 1-day-ahead forecasts because of their high relevance relative to forecasts with longer horizon (Millner and Heyen, 2021b), especially for avoiding mortality from temperature (Shrader et al., 2023), but we also show results for forecast horizons up to 7-days-ahead. We quantify forecast accuracy using the correlation of anomalies, a measure frequently used for forecast verification. Our main stylised fact about the spatial distribution of forecast accuracy are similar if we quantify accuracy using the correlation of levels, or if we use the root mean squared error relative to a measure of variability. We primarily combine forecasts from the ECMWF and a global dataset of surface stations focusing on daily forecasts and observations for 0 UTC and 12 UTC. In additional analysis we show robustness of our main finding to using forecasts from NOAA (GFS) and we also compare with a verification based on model analysis instead of station observations (see Methods for more details).

Our analysis reveals substantial inequality in forecast accuracy. Forecasts generally tend to be less accurate in tropical countries, but countries at high latitude in the Northern Hemisphere also exhibit low accuracy (Figure 1a). Inequality is also apparent on economic dimensions. Richer countries exhibit, on average, more accurate forecasts (Figure 1b), with median forecast accuracy rising monotonically across high versus low income country groups.

These inequalities exist at all forecast horizons from 1 to 7 days ahead with differences between income groups slightly decreasing at the longest forecast horizons (Figure 1c). The differences between income groups are large. For example, 1-day-ahead forecasts in low-income countries tend to be worse than 7-days-ahead forecasts in high-income countries (Figure 1c).

We conduct a large number of robustness tests to corroborate these stylised facts. We find that the overall pattern—higher forecast accuracy in richer countries—is robust to using a different way of aggregating results from individual stations to countries (SI Figure S2b; see Methods for details), restricting the station data to stations with continuous reporting between 1985 and 2020 (SI Figure S2c), using forecasts from the Global Forecast System (GFS) of NOAA instead of the ECMWF (SI Figure S3a), and using forecasts of surface pressure instead of air temperature (SI Figure S3b).

Numerical weather predictions have improved substantially between 1985 and 2020 for all income groups (Figure 2a); an important global success story for weather forecasts. But there has been—and remains—a persistent gap in forecast accuracy between rich and poor countries (Figure 2a,b). For example, the gap in anomaly correlation between low income and high income countries during the 1991–2000 period was 0.15. It decreased by 0.04 points between then and the 2011–2020 period, but is still around 0.11 points. Over the same time period, high income countries experienced an increase in accuracy of 0.15 points (Figure 2b). We find relatively small accuracy gaps between the Northern and Southern Hemispheres in our data (Figure 2c), though again the gap that does exist is persistent. Focus on hemisphere-level analysis, however, misses a large and persistent gap in forecast accuracy between the tropics and extratropics that partly explains the inequality results we find (Figure 2d).

Contrary to earlier work (Bauer et al., 2015) we do not find evidence for a substantial convergence in forecast accuracy between the Northern and Southern Hemispheres. A more detailed comparison suggests that the earlier evidence for convergence can partially be explained by the verification method (SI Figure S4), which evaluated forecasts using model analysis instead of station observations, with the perceived convergence between the Northern and Southern Hemisphere being potentially a consequence of weather forecasts and the associated analysis becoming more similar without proportional improvements in forecast accuracy relative to station observations. We conjecture such a pattern may be related to an increased and better assimilation of satellite data.

Having established these stylised facts about inequalities in forecast accuracy, we next examine inequalities in the infrastructure that contribute to them. We use the same global dataset of land-based stations but focus on stations that reported surface pressure, because

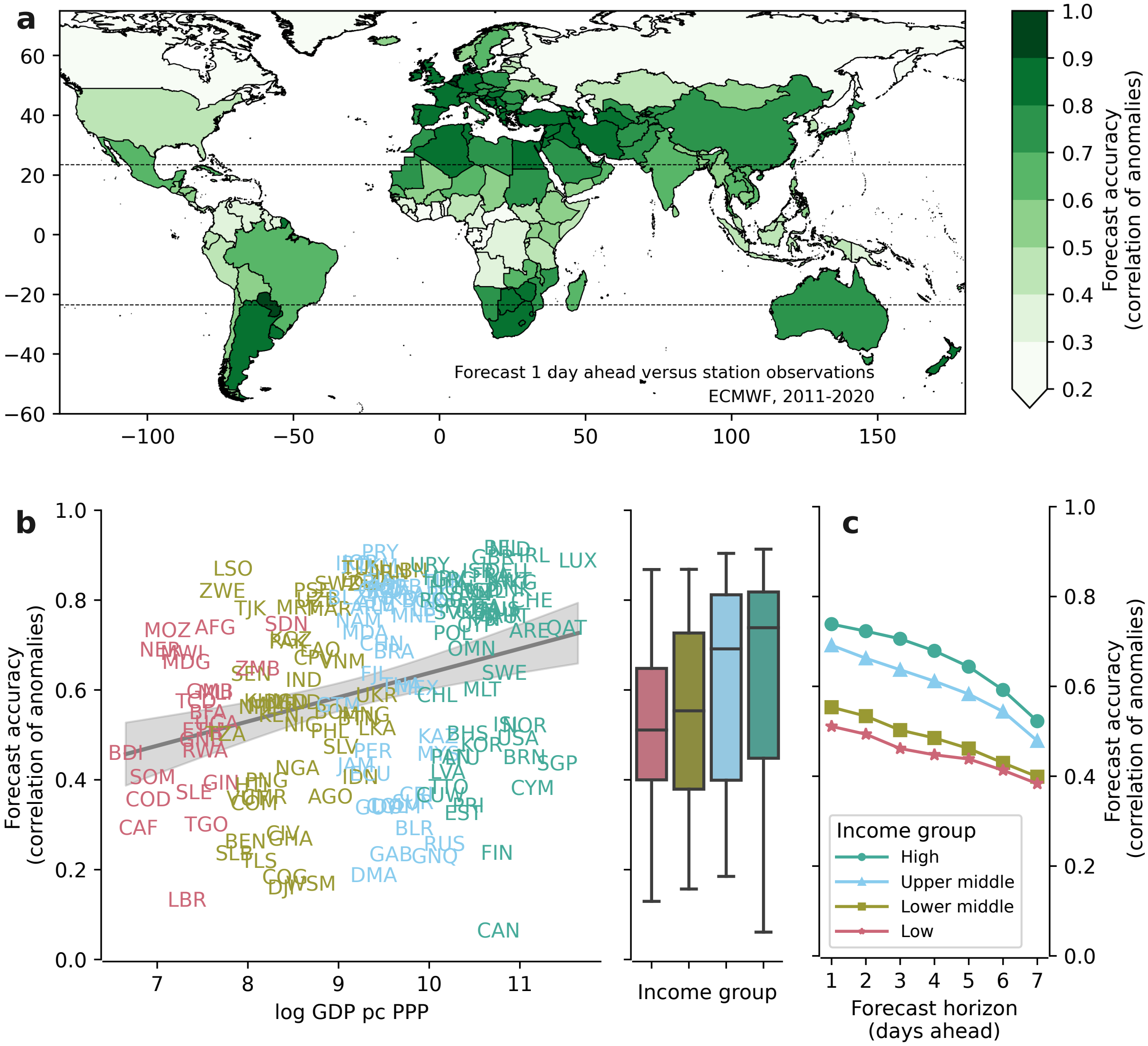

Figure 1. **The accuracy of weather forecasts is unequally distributed across countries; it tends to be higher in countries with higher GDP per capita**. All figures are based on a verification of daily forecasts from the ECMWF of air temperature in 2 metres above ground with land-based station observations. Forecast accuracy is quantified as correlation of anomalies. Forecast horizon in a and b is 1-day-ahead. 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 Gross Domestic Product (GDP) per capita in purchasing power parity (PPP). Colours indicate country income group of the World Bank in 2020. Boxplots show median values, interquartile ranges, and full ranges. **c.** Median forecast accuracy for different forecast windows. See SI Figure S1 for a map of country income groups.

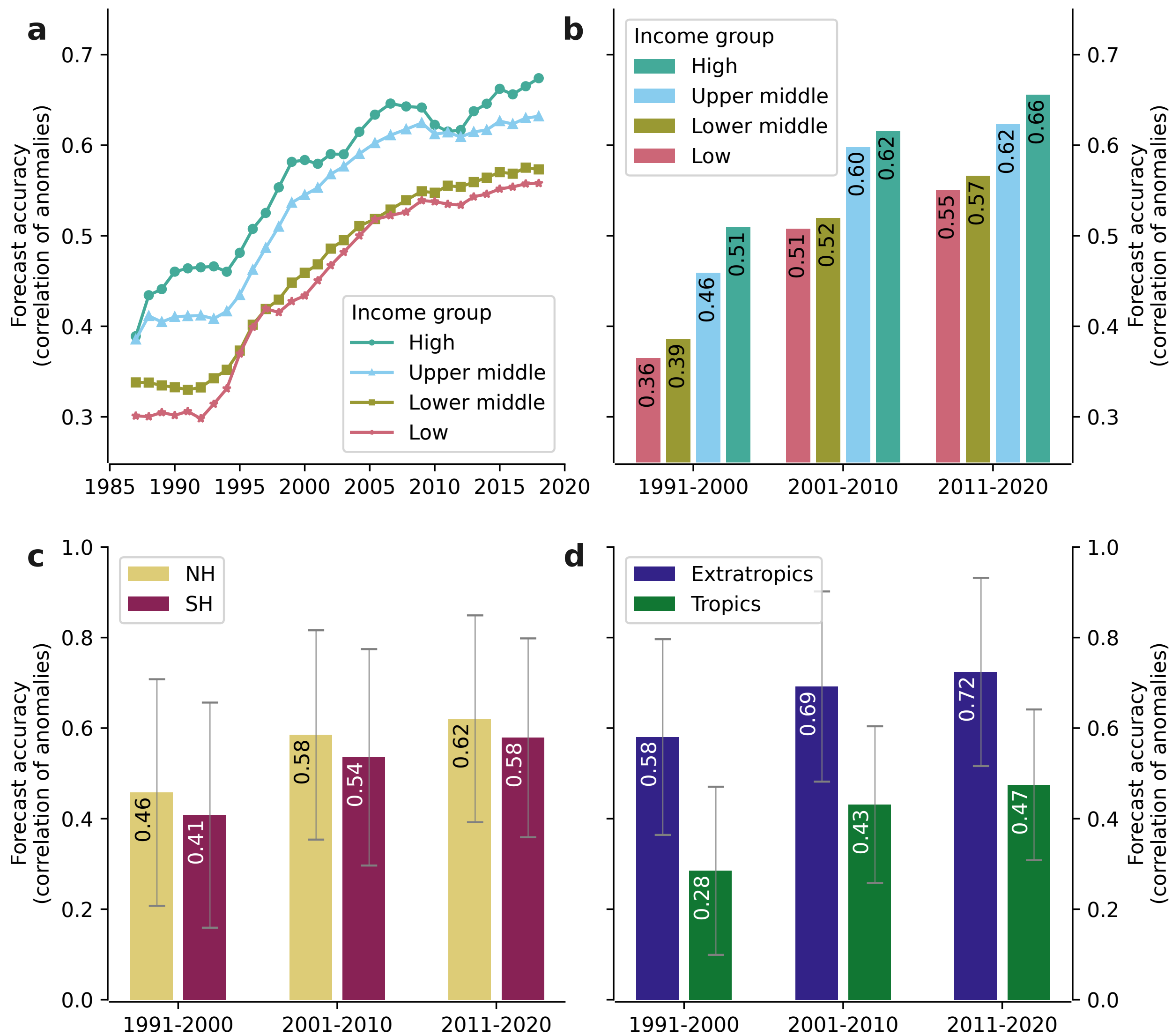

Figure 2. **Forecast accuracy has improved over time, but persistent gaps between richer and poorer countries remain**. Based on the same data as Figure 1. Forecast horizon is 1-day-ahead. **a.** Timeseries of forecast accuracy for different country income groups using a five-years moving average. **b.** Forecast accuracy by country income group and by time period. **c.** Forecast accuracy by hemisphere (NH = Northern Hemisphere, SH = Southern Hemisphere) and by time period. **d.** Forecast accuracy by world region (tropics, extratropics) and by time period.

this variable is more widely assimilated in global forecast models than air temperature at two metres (Haiden et al., 2018). Furthermore, we use a global dataset of radiosondes and a global dataset of drifting sea buoys (see Methods). We choose these three types of in-situ, surface-based meteorological observations because of their relative importance for forecast accuracy (Kull et al., 2021). From all three datasets, we extract the locations as well as the frequencies of observations between 2011 and 2020.

We find that the infrastructure for land-based stations and radiosondes is highly unequally distributed (Figure 3a,b). For both types of observations, richer countries tend to exhibit a larger density of observations per land area than poorer countries (SI Figure S8a,b). For drifting sea-buoys, there seems to be no significant statistical association between GDP per capita and the number of sea buoys within 1,000 km of coastal areas (SI Figure S8c).

These inequalities in observations become even larger when we also account for differences in reporting frequencies (Figure 4a,b). If we compare the medians of the average reporting frequency of countries in different income groups, we find that weather stations in high-income countries tend to report more than twice as frequently at 0 UTC and 12 UTC than stations in low-income countries (Figure 4a). For radiosondes, which typically ascend and report either 1 or 2 times per day—and in some places up to 4 times per day—we find a weaker statistical association between income and reporting rates, but high-income countries again exhibit the highest reporting frequency (Figure 4b).

More observations generally tend to increase forecast accuracy, but the effect of an additional observation is highly dependent on the existing density of observations in a location (Figure 4c). To illustrate this, we present the results of experiments by the UK MetOffice (Kull et al., 2021) which quantify the contribution of surface-based observations to global forecast accuracy. According to these experiments, the average surface observation has an impact of only 0.3 in Europe and 0.6 in North and Central America, but 1.6 in Asia, 2.3 in Africa, and 3.6 in South America (all numbers as absolute values in 10e-6 J per kg; Figure 4c). These numbers quantify the impact of observations on a global measure of forecast accuracy. For a similar local measure of accuracy, the discrepancies are likely even larger (Kull et al., 2021).

Differences in the accuracy of the predictions of global numerical weather models will propagate to official local forecasts issued by national agencies, where the inequalities in model results are potentially exacerbated by differences in institutional capacity between rich and poor countries. To examine inequalities in the issuance of official forecasts by national agencies as a proxy for institutional capacity, we download official national weather forecasts for countries' capitals that are reported to and disseminated by the WMO. The results again show large inequalities between rich and poor countries: about 80% of high-

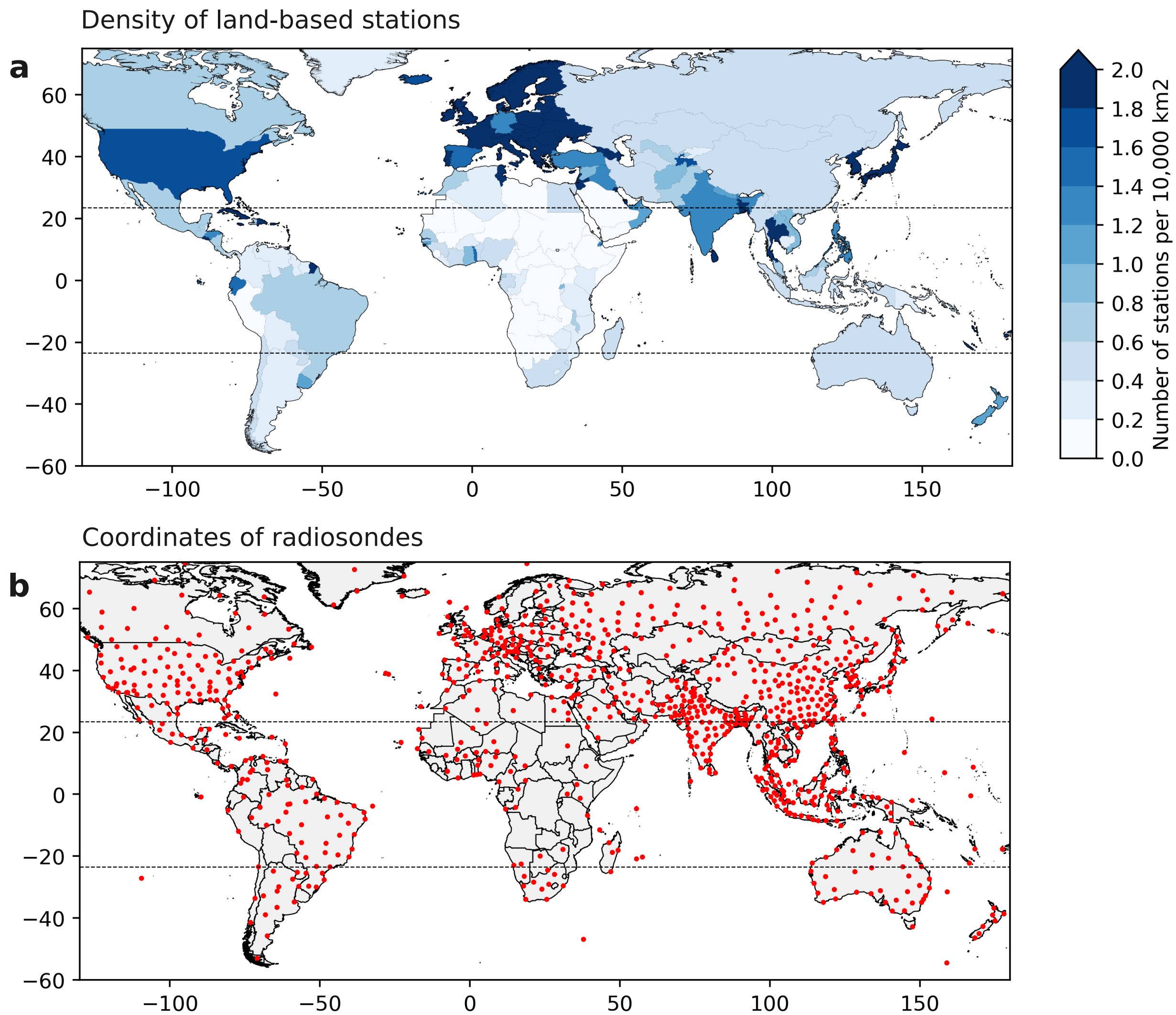

Figure 3. **The infrastructure for meteorological observations is unequally distributed across countries.**. Based on data on land-based weather stations as well as radiosondes from NOAA. All figures show locations that reported at some point 2011-2020. **a.** Map of the density of land-based weather stations that reported sea-level pressure. **b.** Map of locations of radiosondes. See SI Figure S8 for scatter plots of observation density and log GDP per capita of countries.

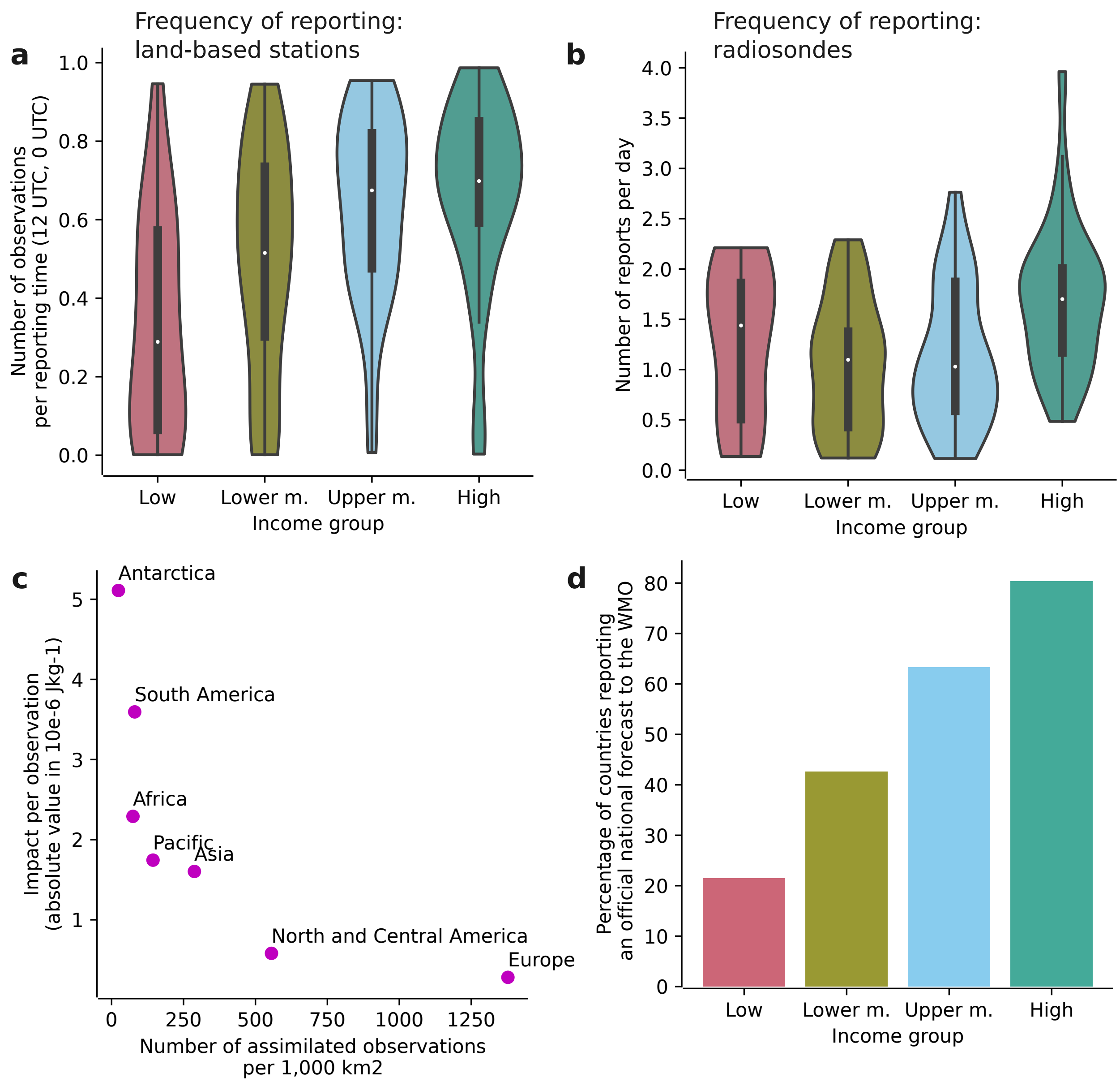

Figure 4. **Reporting rates of the observation infrastructure and the institutional capacities of national meteorological agencies exacerbate inequalities between countries; individual observations have a higher impact where there are fewer of them**. **a.** Reporting rates of land-based stations at 0 UTC and 12 UTC (mean of 2011-2020). Violin plots show distribution of countries, dots indicate medians, boxes interquartile ranges, and lines 95 percent of values. **b.** Reporting rates of radiosondes (mean 2011-2020). **c.** Scatter plot of results of simulations on the impact of individual observations on forecast accuracy as reported in Kull et al. (2021). **d.** Percentage of countries that report an official national forecast for the capital to the WMO.

income countries report an official forecast for their capital to the WMO, but only about 20% of low-income countries do so.

3 Discussion

In this study we examine the spatial variation in the accuracy of weather forecasts and in the physical infrastructure and reporting rates of meteorological observations. To illuminate the socioeconomic implications of existing inequalities, we combine numerical weather predictions, meteorological observations, and economic data and examine the relationship between these different variables. 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 weather observation infrastructure and reporting rates.

Our results reveal large and persistent inequalities in forecast accuracy between rich and poor countries. In additional analysis, we find similar patterns for the observation infrastructure and reporting rates, but we are not able to attribute differences in forecast accuracy to differences in observations. Specifically, large differences in forecast accuracy between the tropics and extratropics can explain a large share of the differences between rich and poor countries and are likely also due to fundamental differences in the predictability of weather (Goddard et al., 2001; Zhang et al., 2019). However, in some additional analysis we find that even within the tropics, poorer countries tend to exhibit worse forecasts. Furthermore, while we are not able to quantify the importance of observations in this paper, we report the results of prior experiments (Kull et al., 2021) (in Figure 4c) that show meteorological observations are particularly valuable in locations with lower observation density. Beyond that specific study, similar computationally demanding assessments, also referred to as data-denial experiments or observing system experiments (Kelly et al., 2007; Bormann et al., 2019), are routinely conducted by meteorological agencies. While the results of such experiments tend to agree that satellite images have become the most important source of information, they also suggest that the three types of observations examined here (land-based stations, radiosondes, and drifting buoys) are important determinants of forecast accuracy (Kull et al., 2021). Their importance is generally found to be largest for shorter forecast horizons and for locations with fewer observations.

An ethical evaluation of the large inequalities in forecast accuracy and observation infrastructure is beyond the scope of this paper. However, existing international initiatives suggest that these inequalities have been identified as problematic by some actors. Initiatives such as the Systematic Observations Financing Facility (SOFF) appear well designed to reduce some of the inequalities in physical infrastructure. Our results suggest that supporting more consistent reporting from existing infrastructure can also be a useful way to reduce inequality.

Closing the gap in forecast accuracy is especially important given evidence that poor countries are expected to experience relatively larger damage from increases in mean temperature (Burke et al., 2015; Carleton et al., 2022) and weather variability (Linsenmeier, 2023). Previous research in high-income countries has shown that short-horizon forecasts can be helpful for reducing the damage from temperature shocks (Shrader et al., 2023), and seasonal weather forecasts have shown promise for helping reduce risks to agriculture around the world (Meza et al., 2008).

Investments in better meteorological observations in developing countries by developed countries may of course not only be justified based on their local effects, but will also increase forecast accuracy globally. Encouragingly, our results suggest that some inequalities can already be reduced by increasing the frequency of data gathering and dissemination at sites with existing infrastructure (Ingleby, 2015; Dinku, 2019). Furthermore, the benefits of better observations extend beyond more accurate operational global weather forecasts and include more valuable disaster early warning systems (Tzachor et al., 2023), more accurate locally downscaled seasonal and decadal climate predictions (Bruno Soares et al., 2018), a higher quality of the data used to understand the impacts of weather on socioeconomic outcomes (Auffhammer et al., 2013), and more generally a better scientific understanding of the status quo and changes in weather and climate.

Methods

Data

Weather forecasts

We use daily weather forecasts from the European Center for Medium-Range Weather Forecasts (ECMWF) from 1985–2020 (ECMWF, 2023). 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(s)-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, for some years we also download forecasts with the validity time 0 UTC and find essentially identical results.

Land-based stations

We combine these forecasts with the world's largest dataset of historical meteorological 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 at 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 centroid.

Radiosondes and pilot balloons

For radiosondes and pilot balloons we use the Integrated Global Radiosonde Archive from NOAA (Durre et al., 2016). 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.

Drifting sea buoys

For drifting sea buoys, we use the archive of NOAA's Global Drifter Program (Lumpkin and Centurioni, 2010). 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 for every year calculate the mean number of buoys using all ocean hexagons within 1000 km based on centroid-to-centroid distances.

Economic data

We combine our meteorological data with economic data from the World Bank (World Bank, 2023). 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 its highest spatial resolution in version 4 for the year 2020.

Statistical methods

Forecast accuracy

For the verification of forecasts, for every station and every year we calculate 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 (Hersbach et al., 2020) using the closest grid cell of each station. We also use this climatology 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.

Aggregation from coordinates to countries

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 for every country calculate a weighted average of hexagons using the population residing inside the hexagon 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 essentially the same (SI Figure S3).

References

Anand, V. (2022). The value of forecast improvements: Evidence from advisory lead times and vehicle crashes. *Available at SSRN 4206910*.

Auffhammer, M., Hsiang, S. M., Schlenker, W., and Sobel, A. (2013). Using Weather Data and Climate Model Output in Economic Analyses of Climate Change. *Review of Environmental Economics and Policy*, 7(2):181–198.

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.

Bruno Soares, M., Daly, M., and Dessai, S. (2018). Assessing the value of seasonal climate forecasts for decision-making. *WIREs Climate Change*, 9(4):e523.

Burke, M., Hsiang, S. M., and Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239.

Carleton, T. and Greenstone, M. (2022). A guide to updating the us government's social cost of carbon. *Review of Environmental Economics and Policy*, 16(2):196–218.

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.

Carleton, T. A. and Hsiang, S. M. (2016). Social and economic impacts of climate. *Science*, 353(6304):1–15.

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.

Durre, I., Yin, X., Vose, R. S., Applequist, S., Arnfield, J., Korzeniewski, B., and Hundermark, B. (2016). Integrated Global Radiosonde Archive (IGRA), Version 2.

ECMWF (2023). *Meteorological Archival and Retrieval System (MARS)*. ECMWF, https://www.ecmwf.int/en/forecasts/access-forecasts/access-archive-datasets.

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., Dahoui, M., Ingleby, B., De Rosnay, P., Prates, C., Kuscu, E., Hewson, T., Isaksen, L., Richardson, D., Zuo, H., et al. (2018). *Use of in Situ Surface Observations at ECMWF*. European Centre for Medium Range Weather Forecasts.

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.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., De Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049.

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.

Kull, D., Riishojgaard, L. P., Eyre, J., and Varley, R. A. (2021). *The Value of Surface-Based Meteorological Observation Data*. World Bank, Washington, DC.

Leviäkangas, P. (2009). Valuing meteorological information. *Meteorological Applications*, 16(3):315–323.

Linsenmeier, M. (2023). Temperature variability and long-run economic development. *Journal of Environmental Economics and Management*, 121:102840.

Lumpkin, R. and Centurioni, L. (2010). NOAA Global Drifter Program quality-controlled 6-hour interpolated data from ocean surface drifting buoys.

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

Meza, F. J., Hansen, J. W., and Osgood, D. (2008). Economic value of seasonal climate forecasts for agriculture: review of ex-ante assessments and recommendations for future research. *Journal of Applied Meteorology and Climatology*, 47(5):1269–1286.

Millner, A. and Heyen, D. (2021a). Prediction: The long and the short of it. *American Economic Journal: Microeconomics*, 13(1):374–398.

Millner, A. and Heyen, D. (2021b). Prediction: The Long and the Short of It. *American Economic Journal: Microeconomics*, 13(1):374–398.

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.

World Bank (2023). *World Development Indicators.* World Bank, https://databank.worldbank.org/source/world-development-indicators.

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)

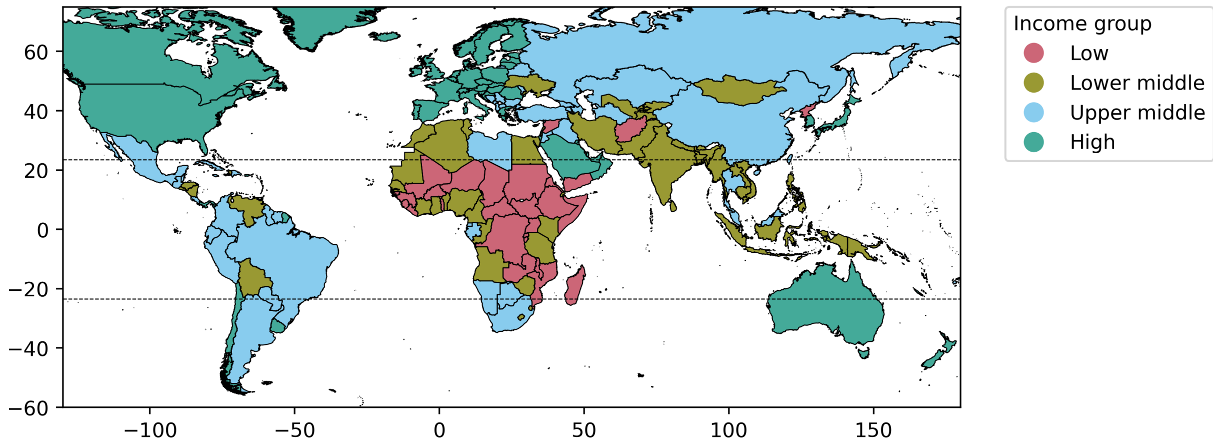

Figure S1. Map of countries and income groups. Income groups are shown in different colour. Income groups are from the World Bank classification in 2020.

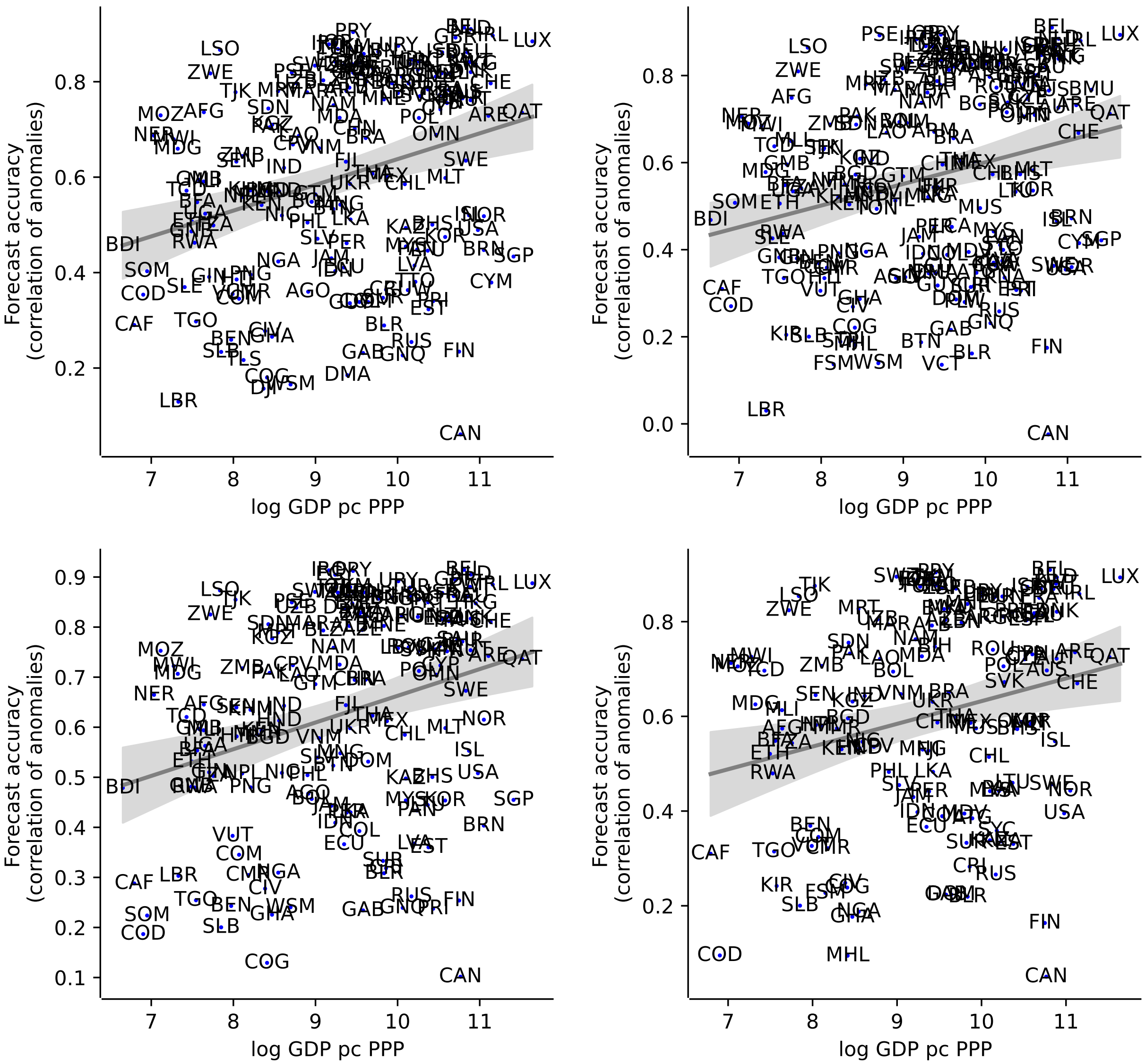

Figure S2. Scatterplot of forecast accuracy and log GDP per capita of countries for different ways of aggregation and for different samples of stations. Figure a is identical to Figure 1b. Figure b shows results aggregated from stations without the use of population weights. Figure c shows results with the population weights, but only based on subset of stations with continuous reporting between 1985-2020. Figure d shows results without population weights and only for the subset of stations in c.

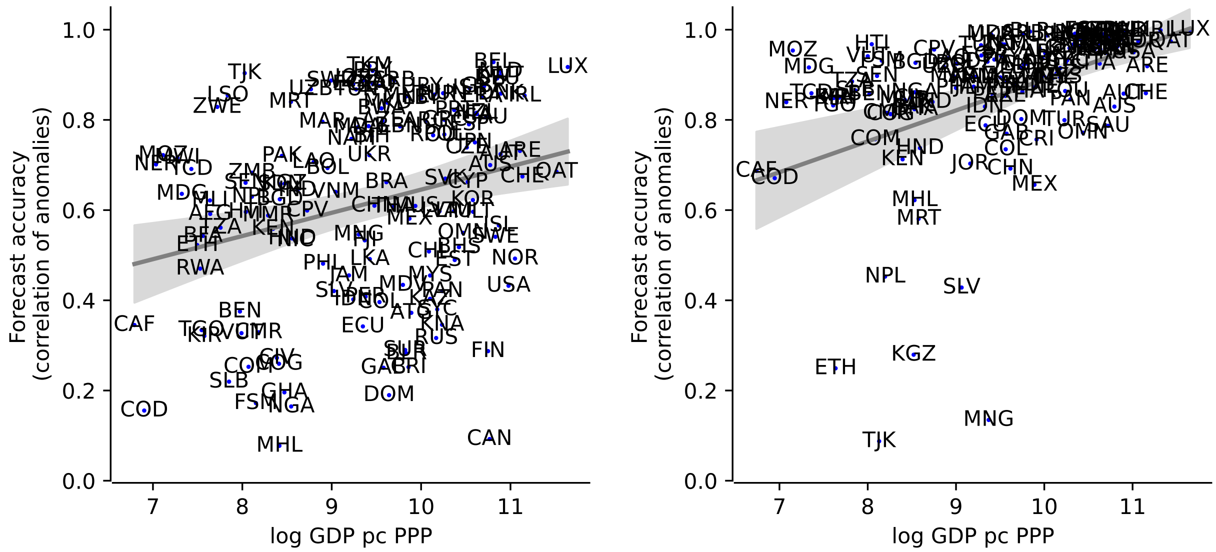

Figure S3. Scatterplot of forecast accuracy and log GDP per capita of countries for different weather forecast models and for different meteorological variables. The reference is Figure 1b, which shows the same as SI Figure S2a. **a.** Forecast from GFS instead of ECMWF. **b.** Forecast of surface pressure instead of air temperature in 2 metres. Shaded areas show 95% confidence intervals obtained from heteroscedasticity-robust standard errors.

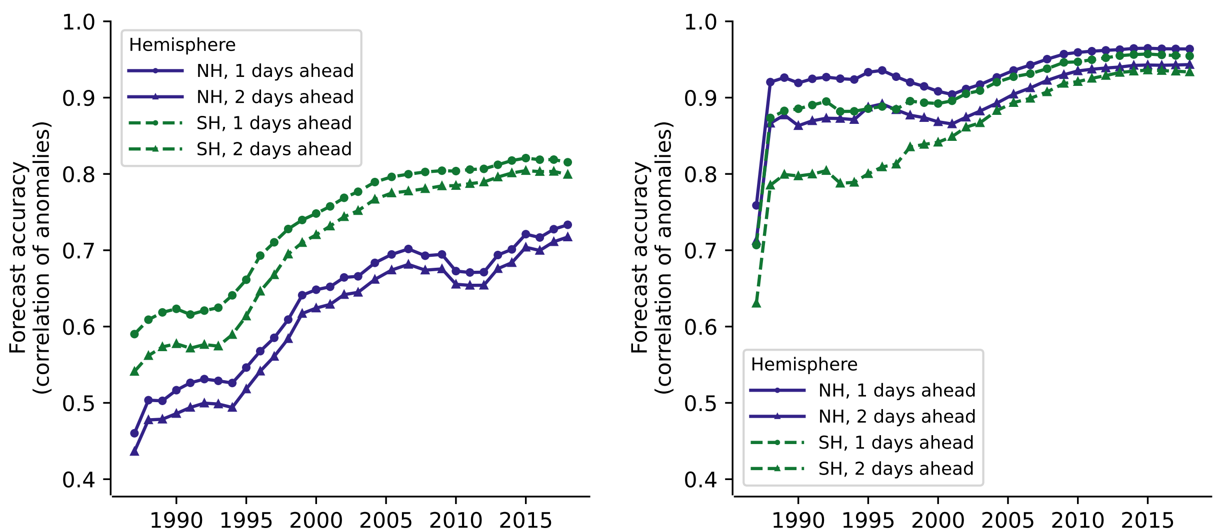

Figure S4. Timeseries of forecast accuracy by hemisphere and by forecast horizon for two alternative verification methods. Verification based on: **a.** Station observations, **b.** Model analysis. Both figures show five-years moving averages. NH = Northern Hemisphere, SH = Southern Hemisphere.

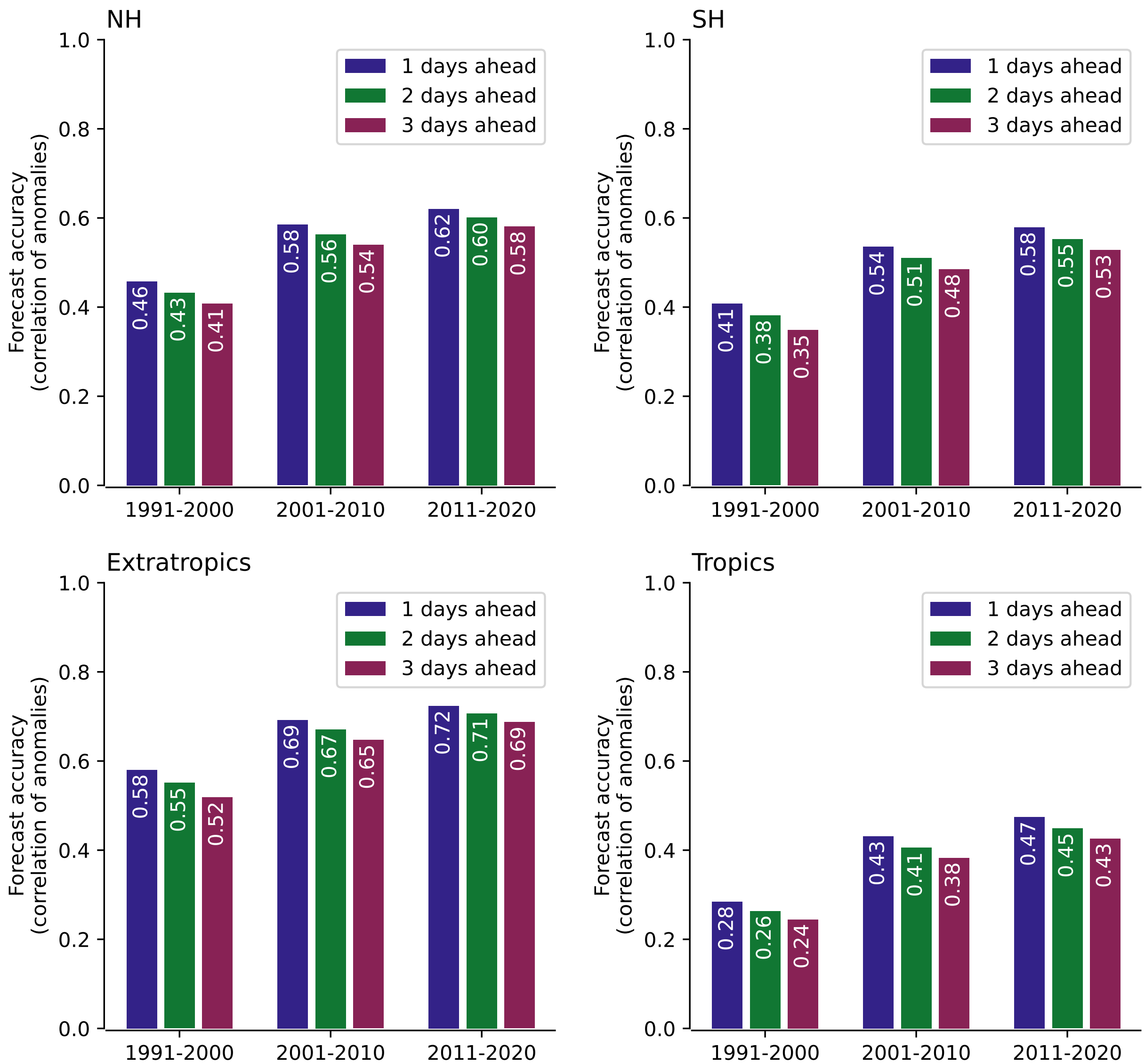

Figure S5. Improvements in forecast accuracy over time by forecast horizon. **a.** and **b.** Improvements by hemisphere (NH = Northern Hemisphere, SH = Southern Hemisphere). **c.** and **d.** Improvements by region (tropics, extratropics). See also Figures 2c and 2d.

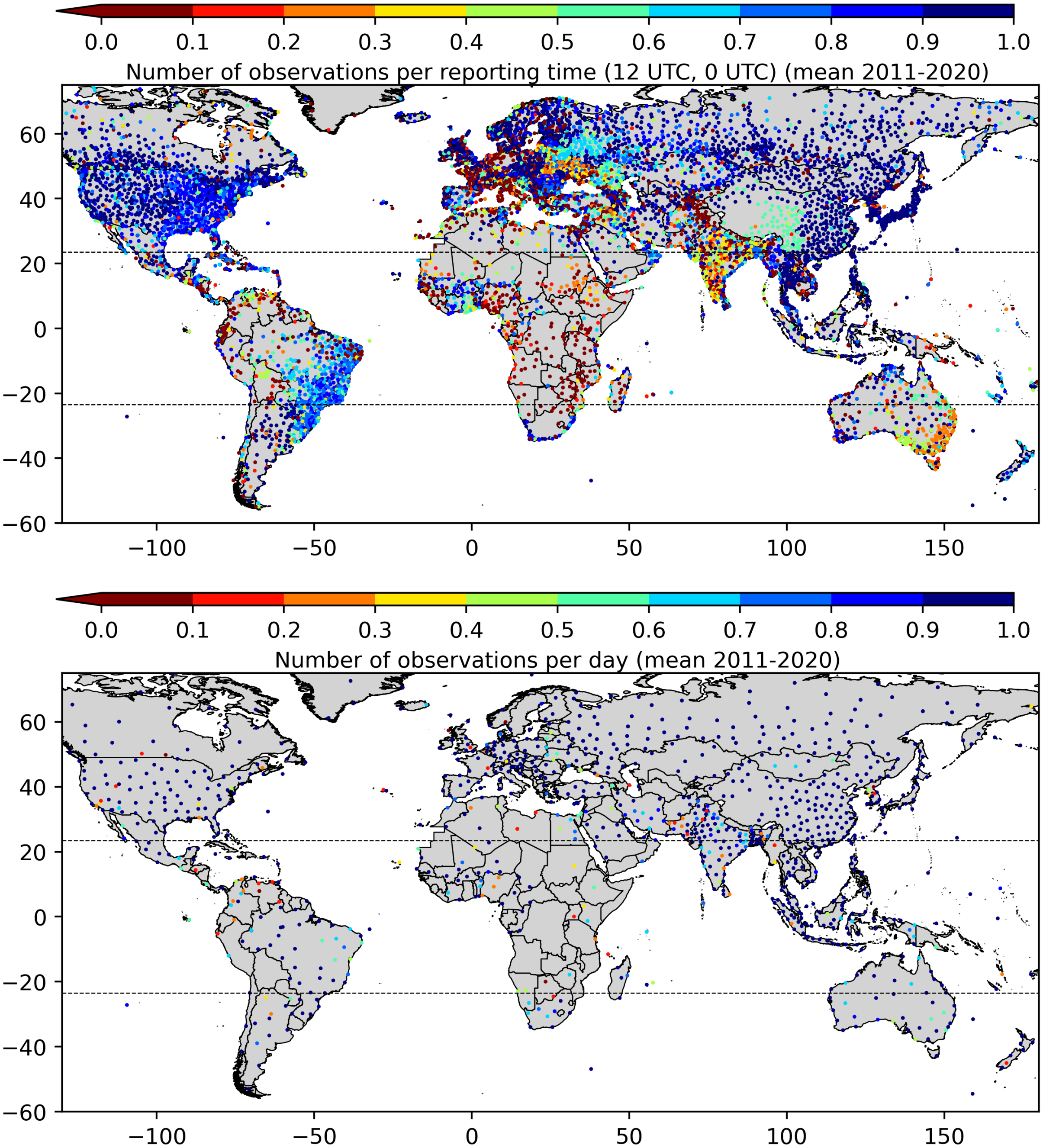

Figure S6. Maps of the observation infrastructure. **a.** Location and reporting frequency of land-based weather stations. **b.** Location and reporting frequency for radiosondes. All maps are generated from data from NOAA over the years 2011-2020.

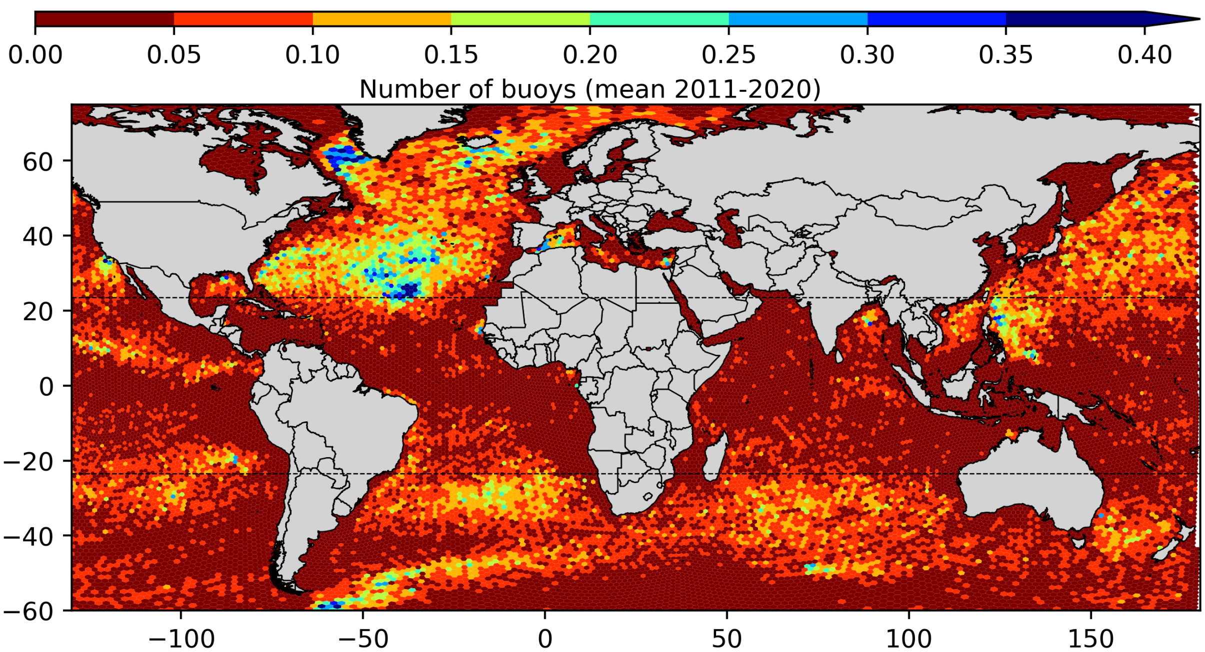

Figure S7. Map of the drifting sea buoys. The figure shows the average number of buoys per hexagon based on data from NOAA over the years 2011-2020.

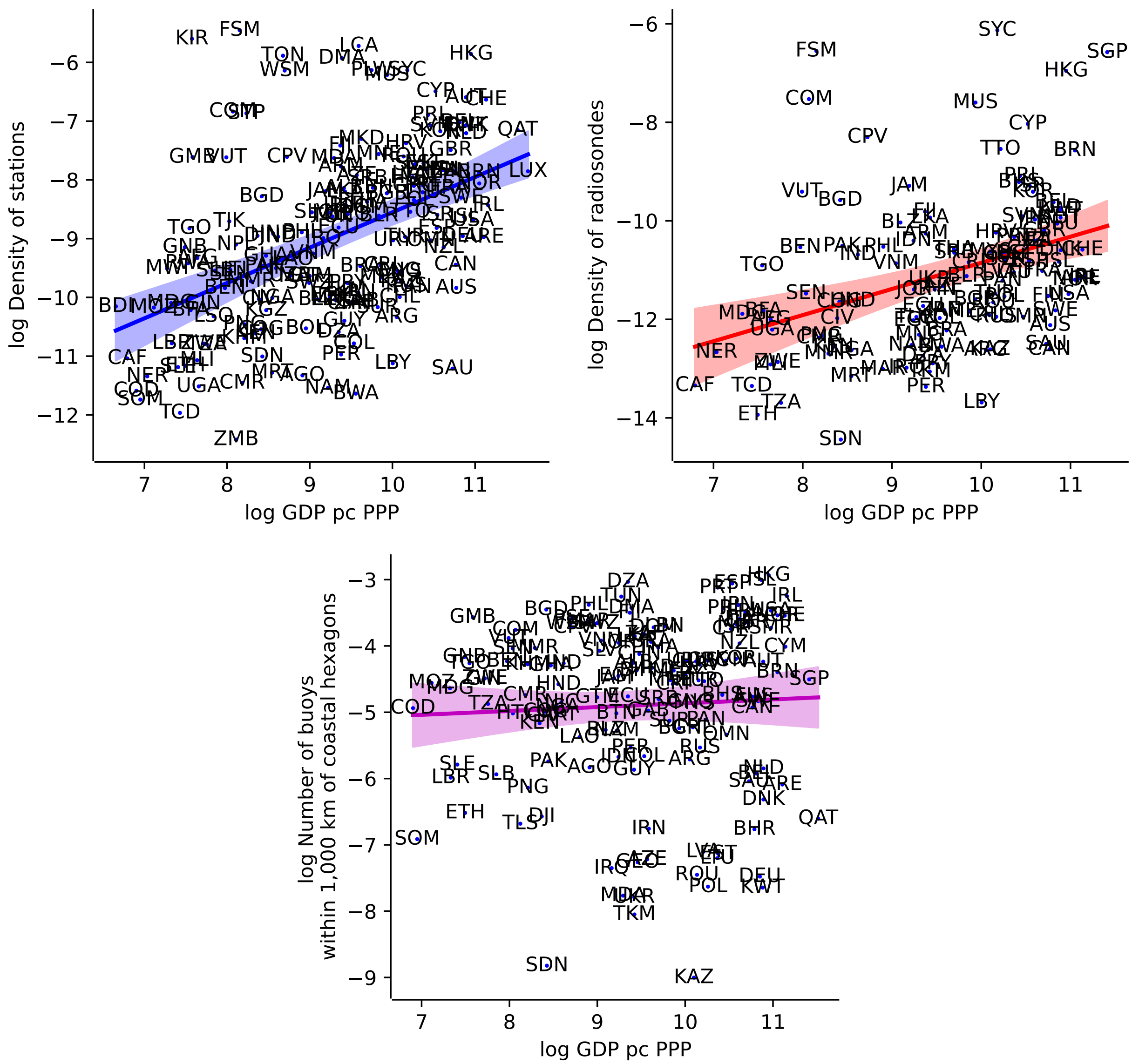

Figure S8. **a.** Scatter plot of the log density of land-based weather stations and log GDP per capita of countries. **b.** Scatter plot of the log density of radiosondes and log GDP per capita of countries. **c.** Scatter plot of the log number of drifting sea buoys within 1,000 km of coastal hexagons and log GDP per capita of countries. Shaded areas show 95% confidence intervals obtained from heteroscedasticity-robust standard errors.