The unequal global distribution of weather forecast accuracy and the value of ground-based observations *

Manuel Linsenmeier† Jeffrey Shrader‡

March 17th, 2023

Abstract

Global weather forecasts are of great economic value for society, but regional 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 find 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 more than 25% higher than the average accuracy in low-income countries. Second, after a period of converging forecast quality during the late 1980s and 1990s, forecast accuracy has strongly diverged across countries in different income groups for the last two decades. This effect is largely driven by a rapid decline in forecast accuracy in low-income countries between the late 1990s and mid 2000s. Decline in quality and divergence across regions stands in sharp contrast to the average improvements in global forecast quality that have been witnessed in recent decades (Bauer et al., 2015). Third, the difference in forecast quality is strongly correlated with differences in the density of weather stations—a piece of public infrastructure that exhibits high inequality across countries and which is an important input into forecasts.

1 Introduction

Weather forecasts can help to protect lives and they provide multiple other benefits to society. 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 depends on their accuracy but while this accuracy is routinely assessed by international and national meterorological and hydrological services (NMHS), outside this small community little is known about its temporal and spatial distribution. This is

*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

potentially problematic because the methodologies and metrics used for the verification of weather forecasts by NMHS are not necessarily the most relevant ones for all beneficiaries of weather forecasts (Casati et al., 2008). For example, most routine verifications focus on atmospheric pressure in the middle of the atmosphere, while extreme temperature warning systems require accurate forecasts of temperature close to the surface (Ebi et al., 2004). Furthermore, routine verifications often focus on only a part of the world, such as the extratropics, or provide results with coarse spatial aggregation, such as the Northern and the Southern hemispheres. To identify and address inequalities that are associated with certain uses of forecasts, more specific evidence on the spatial distribution of forecast accuracy is needed.

Given the prospects of future climate change (IPCC, 2022) and the economic importance of temperature-related mortality among all projected climate impacts, forecasts of near surface temperature are of particular interest. Most protection against extreme temperature events does not require much anticipation, which means that forecasts one or more days ahead are most relevant (Shrader et al., 2023). Furthermore, given the global variation in exposure to temperature extremes (Tuholske et al., 2021), the spatial dimension of forecast accuracy deserves great attention. These demands contrast most of the prior work on forecast accuracy that focused on trends over time instead of spatial differences (Magnusson and Källén, 2013), used coarse spatial aggregation or excluded part 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). A notable exception 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 spatial inequalities.

Global variation in forecast accuracy arises from differences in the physically 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). While such exercises provide useful information on the overall importance of different types of information for short periods of time (typically a few months), they do not say much about the marginal effect of small changes to the existing measurement infrastructure.

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 discuss the economic implications of existing inequalities. The analysis consists of four main parts. In the first part, we examine the cross-section of forecasts between 2011-2020. In the second part, we examine trends in forecast accuracy between 1985 and 2020. In the third part, we combine our data on forecast accuracy with data on land-based weather stations and estimate to what extent the entering and exiting of stations can explain variation in forecast accuracy from year to year. In the last part, we use our empirical estimates for some counterfactual calculations that illustrate how much additional weather stations in different parts of the world can reduce inequalities in forecast accuracy.

2 Methods

We combine daily weather forecasts from the European Center for Medium-Range Weather Forecasts (ECMWF) with the world's largest dataset of historical weather observations from land-based measurement stations (Smith et al., 2011) and subnational economic data (DOSE).[1] For robustness checks, we replicate some of our analysis with global forecasts from NOAA/NCEP and the UK MetOffice. We focus on air temperature in 2 meters. In robustness checks, we examine forecast accuracy for surface pressure and the geopotential height of the 500 hPa isobar surface. Forecast accuracy is measured using the correlation between the anomaly of the forecast and the anomaly of the corresponding analysis. The anomalies are obtained by subtracting the climatology from 1991-2020 (the climatology is obtained from ERA5 reanalysis).

To examine the role of ground-based observations we conduct an econometric analysis of forecast accuracy using year-to-year variation in the existence and reporting of weather stations for identification. The estimating equation is given by

$$a_{it} = f(s_{it}; \beta) + \rho_i + \gamma_t + \varepsilon_{it} \tag{1}$$

where a_{ct} is the forecast accuracy (anomaly correlation) in grid cell i and year t. The main right-hand-side variable of interest is the density of weather stations, s. The equation includes grid cell and year fixed effects, ρ_i and γ_t. These fixed effects absorb time-invariant

[1] The ECMWF forecast is typically considered to be the most accurate global, numerical weather forecast.

differences between grid cells and globally uniform trends over time, respectively. Including these fixed effects means that the effect of changes in station density on forecast accuracy is identified based on variation within a grid cell over time and relative to the global trend.

3 Results

We first examine currently existing inequalities in weather forecast accuracy using the average anomaly correlation of short-term forecasts of air temperature in 2 meters over the period 2011-2020. We calculate this measure of forecast accuracy based on forecasts and the corresponding analysis at the level of grid cells and then aggregate to countries and country income groups of the World Bank. We find large variation in forecast accuracy across countries (Figure 1a). Weather forecasts tend to be most accurate in high-income countries and least accurate in low-income countries (Figure 1b). On average, one-day-ahead forecasts in low-income countries tend to be substantially worse than three-days-ahead forecasts in high-income countries (Figure 2a). Middle-income countries exhibit a large variation in forecast accuracy with higher accuracy in richer countries (Figure 1b).

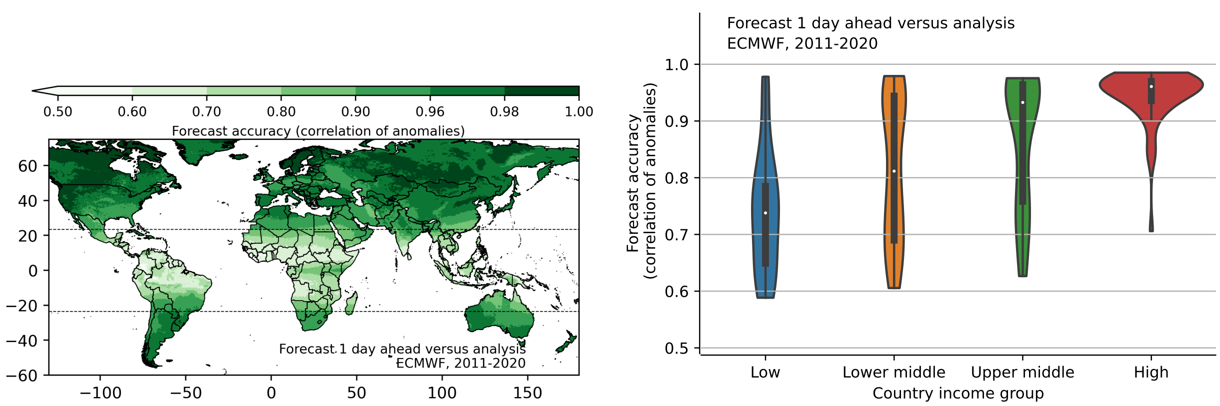

Figure 1. **Inequities in forecast quality across regions and income groups are substantial**. The left panel shows 1-day-ahead temperature forecast accuracy (measured by the correlation between forecast and temperature realization anomalies) for each $0.5 \times 0.5°$ global, land surface grid cell. The right panel shows the distribution of forecast quality within four different country income groups, based on the World Bank's classification. The point in the middle of each area shows the median, the black bars show the interquartile range, and the colored areas show the frequency distribution.

For robustness checks, we conduct the same analysis using forecasts and analysis from two other NMHI, NOAA/NCEP in the USA and the MetOffice of the UK. We find similar statistical associations between income and mean forecast accuracy as for the forecast from the ECMWF (SI Figures 1a and 1b). We interpret this as evidence that the spatial inequalities are not specific to the data assimilation algorithms or numerical model of the ECMWF, but instead related to physics, model capabilities, or weather observations that are shared by currently state-of-the-art weather forecasts. We next examine whether this spatial pattern is specific to near-surface temperature. To do so, we repeat again the same analysis but with surface pressure and the geopotential height of the 500 hPa isobar surface. We find again a similar pattern, with low-income countries featuring the lowest and high-income countries the highest average forecast accuracy (SI Figures 2a and 2b).

We next examine to what extent the association between income and forecast accuracy can be explained by differences in weather and physically constrained predictability between the tropics and the extratropics. To do so, we stratify our sample by bands of latitude, distinguishing between the tropics (-23 to 23 degrees latitude, mid latitudes (-66 to -23 and 23 to 66 degrees), and high latitudes (-90 to -66 and 66-0- degrees). As expected from visual inspection of Figure 1a, we find that latitude can explain a large share of the differences in forecast accuracy between low-income and high-income countries (SI Figure 3a). However, we also find that income and forecast accuracy are systematically positively correlated inside the tropics and outside the tropics, suggesting that countries with lower income tend to experience lower forecast accuracy within all three bands of latitude (SI Figure 3b).

We next examine trends in forecast accuracy from 1985-2020. We distinguish between the period before and after the year 2000, approximately the time when new data assimilation techniques were introduced that substantially improved the ability to incorporate satellite data into global weather models. Previous research in meteorology has shown that starting around the year 2000, global weather forecasts became substantially better, particularly for the extra-tropical Southern Hemisphere (Bauer et al., 2015). At the ECMWF, the year 2000 roughly coincides with the introduction of 3D-Var and 4D-Var assimmilation techniques in 1996 and 1997, respectivey (Magnusson and Källén, 2013). We find large improvements in forecast accuracy in high-income countries in the periods before and after 2000 (Figure 2a). For example, three-days-ahead forecasts in 2020 were approximately as accurate as one-day-ahead forecasts in 2000. However, for lower-middle-income and low-income countries we find a stagnation and even decline of the accuracy of one-day-ahead forecasts since 2000, respectively. Forecasts with longer forecast horizons, two-days ahead and three-days ahead, improved everywhere, but less so in poorer countries, leading to a global divergence in forecast accuracy since 2000 (Figure 2b). Between 2000 and 2020, 1-day-ahead forecasts became 5% more accurate in high-income countries. For the poorest countries, however, the became almost 10% worse.

In the last part of the analysis, we relate differences in weather forecast accuracy to

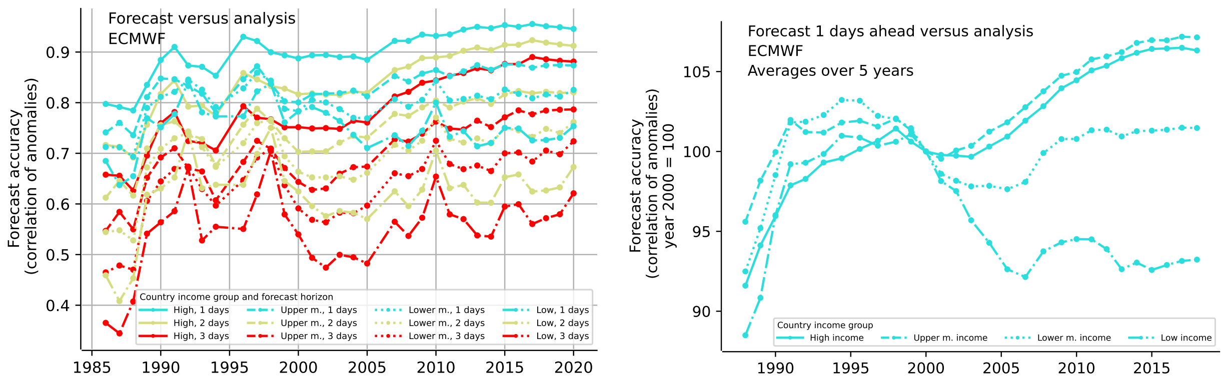

Figure 2. **Rapid progress in forecast quality reversed after 2000 in low-income countries**. The left panel shows forecast accuracy (based on correlation of forecast and temperature realization anomalies) for 1, 2, and 3-day-ahead forecasts and for low, lower, upper, and high-income countries for each year from 1986 to 2020. The right panel shows just the 1-day-ahead forecast accuracy, smoothed using a rolling 5-year average. Accuracy is normalized to 100 in the year 2000 to assess relative trends since that year.

changes in the density of weather stations over time. To do so, we estimate regression models (Equation 1) that relate changes in forecast accuracy over time for a given grid cell after subtraction of global trends to changes in the number of stations per area that reported 2 m temperature in a given year. We find that between 1985 and 2020, higher station density significantly improved forecast accuracy ((Figure 3a). The effect appears to be non-linear with a positive effect of higher density up to a density of about 8 stations per 100,000 km2 (Figure 3a). This value corresponds to approximately a fourth of the average density in high-income countries but it is larger than the average density in low-income countries (Figure 1b). When we split the time period in pre and post 2000, we find that the positive effect of higher density slightly decreased over time outside and increased inside the tropics (Figure 3a).

The overall differences in forecast quality across countries and over time are strongly associated with differences in weather station density. The countries with the lowest average forecast accuracy (Figure 1a) are also those with the lowest average station density over the last decade (Figure 4a). And over time, high and upper-middle income countries have substantially increased their station density, while low income countries have experienced stagnation and, during periods in the mid-2000s, declines in station density (Figure 4b).

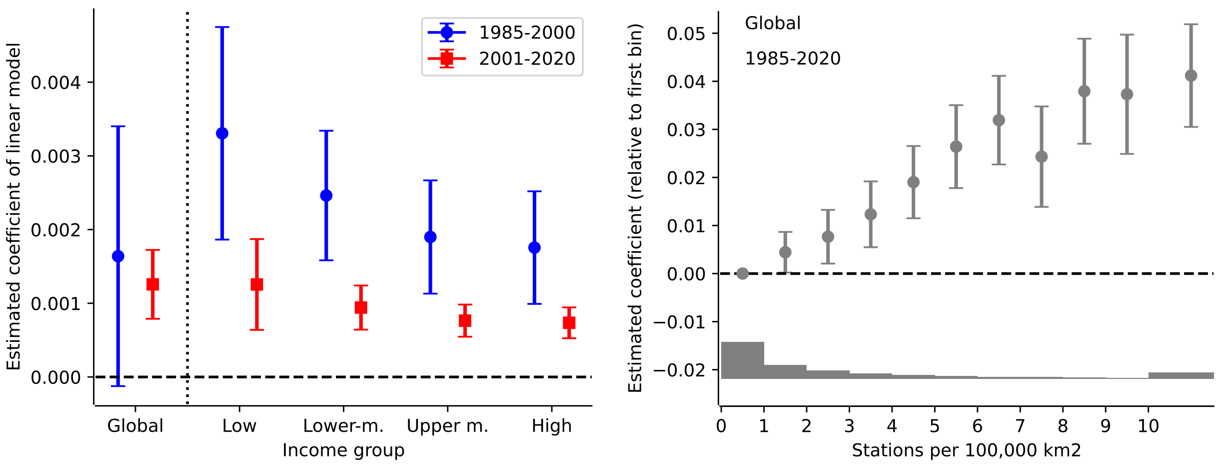

Figure 3. **Weather stations are an important input into accurate forecasts**. The left panel shows the linear correlation between weather station density and 1-day-ahead forecast accuracy across the globe and broken down into extratropical and tropical regions obtained from the estimation of Equation 1. The right panel uses a global analysis to assess nonlinearities in the effect of increased weather station density on forecast accuracy.

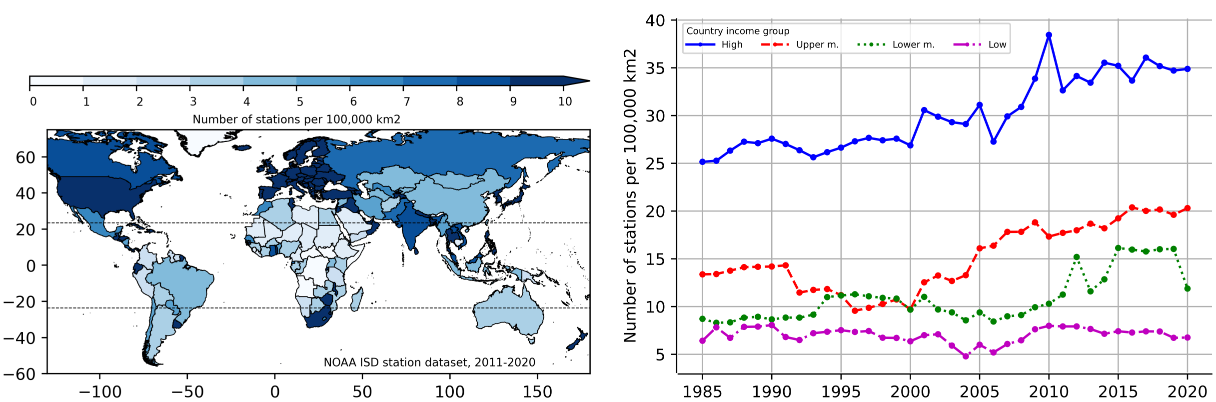

Figure 4. **Weather station inequality reflects forecast accuracy inequality**. The left panel shows the average number of stations per 100,000 km^2 for each country in the world between 2011 and 2020. The right panel shows trends in station density over time for the four income groups.

4 Discussion

We examine currently existing global inequalities in weather forecast accuracy, discuss their economic implications, and relate our results to the globally unequal infrastructure of weather stations. 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 primarily with its explicit considerations of an economic dimension of forecast accuracy and it novel analysis of the influence of station observations using the entering and exiting of stations as "natural" experiments. This empirical approach is similar in spirit to recent work that studied the importance of aircraft observations using the reduction of transatlantic flights during the early phase of the COVID-19 pandemic (Ingleby et al., 2021).

Our results suggest that low-income countries are systematically disadvantaged in terms of the accuracy of global weather forecasts. This pattern can be partially attributed to their systematically lower density of weather stations. Back of the envelope calculations suggest that a targeted increase in weather stations in low-income countries from approximately 300 internationally reporting stations to 900 can close up to a fifth of the average gap in forecast accuracy relative to high-income countries. This corresponds roughly to the improvement of one-day-ahead forecasts in high-income countries between 2000 and 2020. Given that better observations improve weather forecasts also elsewhere (Magnusson, 2017), the benefits of additional weather stations in currently relatively data-sparse regions are likely even higher. Similar benefits can likely be reaped from improvements in the quality and consistency of reports from existing stations (Ingleby, 2015; Dinku, 2019).

References

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

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

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

Carleton, T., A. Jina, M. Delgado, M. Greenstone, T. Houser, S. Hsiang, A. Hultgren, R. E. Kopp, K. E. McCusker, I. Nath, 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., L. J. Wilson, D. B. Stephenson, P. Nurmi, A. Ghelli, M. Pocernich, U. Damrath, E. E. Ebert, B. G. Brown, and S. Mason (2008, March). Forecast verification: Current status and future directions. *Meteorological Applications 15*(1), 3–18.

de Perez, E. C., M. van Aalst, K. Bischiniotis, S. Mason, H. Nissan, F. Pappenberger, E. Stephens, E. Zsoter, and B. van den Hurk (2018, May). 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*, pp. 71–80. Elsevier.

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

Ebi, K. L., T. J. Teisberg, L. S. Kalkstein, L. Robinson, and R. F. Weiher (2004, August). Heat Watch/Warning Systems Save Lives: Estimated Costs and Benefits for Philadelphia 1995–98. *Bulletin of the American Meteorological Society 85*(8), 1067–1074.

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

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

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

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

Ingleby, B., B. Candy, J. Eyre, T. Haiden, C. Hill, L. Isaksen, D. Kleist, F. Smith, P. Steinle, S. Taylor, W. Tennant, and C. Tingwell (2021, February). The Impact of COVID-19 on Weather Forecasts: A Balanced View. *Geophysical Research Letters 48*(4).

IPCC (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability.* Summary for Policymakers. Cambridge, UK and New York, USA: Cambridge University Press.

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

Kelly, G., J.-N. Thépaut, R. Buizza, and C. Cardinali (2007, October). 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, July). Diagnostic methods for understanding the origin of forecast errors. *Quarterly Journal of the Royal Meteorological Society 143*(706), 2129–2142.

Magnusson, L., J.-H. Chen, S.-J. Lin, L. Zhou, and X. Chen (2019, July). 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 E. Källén (2013, September). Factors Influencing Skill Improvements in the ECMWF Forecasting System. *Monthly Weather Review 141*(9), 3142–3153.

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

Smith, A., N. Lott, and R. Vose (2011, June). 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.

Tuholske, C., K. Caylor, C. Funk, A. Verdin, S. Sweeney, K. Grace, P. Peterson, and T. Evans (2021, October). Global urban population exposure to extreme heat. *Proceedings of the National Academy of Sciences 118*(41), e2024792118.

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

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

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

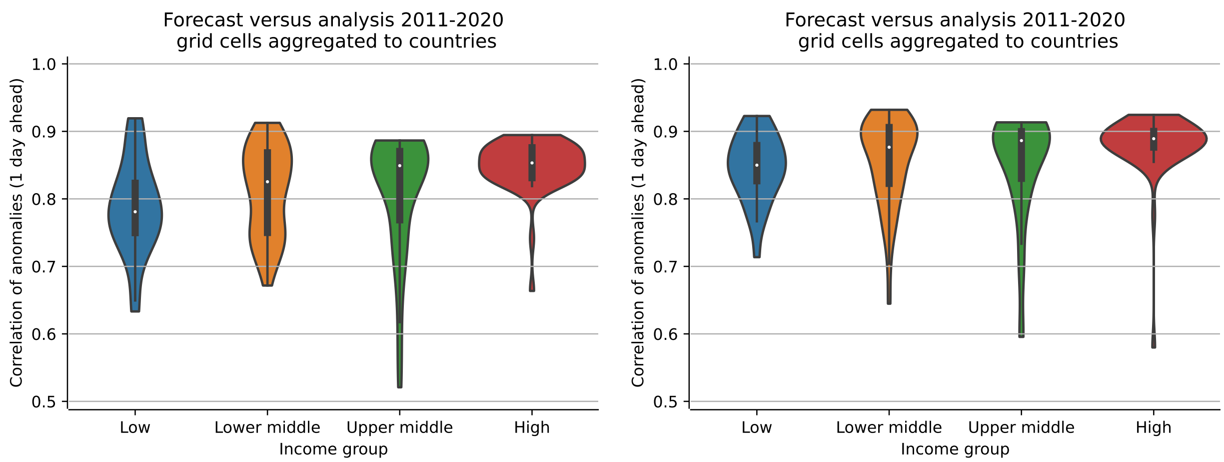

Figure 1. **Inequities across regions and income groups are similar for forecasts from different institutions**. The left panel shows the accuracy of the global forecast of the UK MetOffice. The right panel shows the same for forecasts with GFS of NOAA/NCEP. Both panels show the mean accuracy for the period 2011-2020.

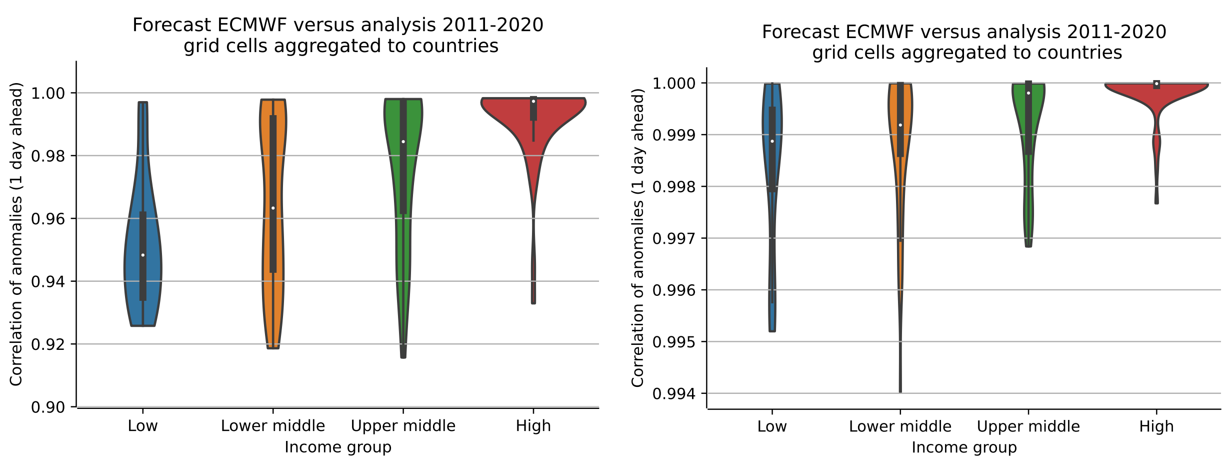

Figure 2. **Inequities across regions and income groups are similar for different variables.** The left panel shows forecast accuracy of surface pressure. The right panel shows the same for the geopotential height of the 500 hPa pressure surface.

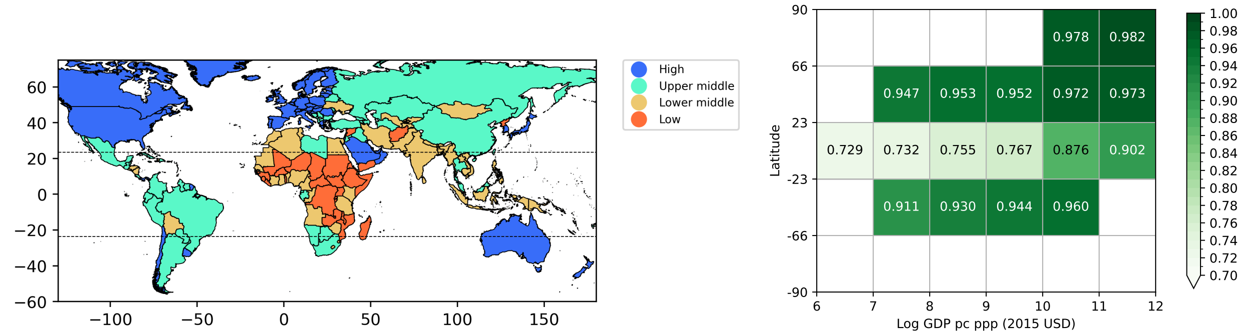

Figure 3. **Inequities in forecast accuracy can be observed inside and outside the tropics.** The left panel shows the distribution of countries of the four income groups. The right panel shows the forecast accuracy for t2m for countries with different income and bands of latitude. Note that forecast accuracy increases with both income and latitude.

References