Global and regional long-term climate forecasts: A heterogeneous future *

María Dolores Gadea Rivas †
University of Zaragoza

Jesús Gonzalo ‡
U. Carlos III de Madrid

January 31, 2025

Abstract

Climate is a long-term issue, and as such, climate forecasts should be designed with a long-term perspective. These forecasts are critical for crafting mitigation policies aimed at achieving one of the primary objectives of the Paris Climate Agreement (PCA) and for designing adaptation strategies to alleviate the adverse effects of climate change. Furthermore, they serve as indispensable tools for assessing climate risks and guiding the green transition effectively. This paper introduces a straightforward method for generating long-term temperature density forecasts using observational data, leveraging the realized quantile methodology developed by Gadea and Gonzalo (JoE, 2020). This methodology transforms unconditional quantiles into time series objects. The resulting forecasts complement those produced by physical climate models, which primarily focus on average temperature values. By contrast, our density forecasts capture broader distributional characteristics, including spatial disparities that are often obscured in mean-based projections. The proposed approach involves conducting an out-of-sample forecast model competition and integrating the forecasts from the resulting Pareto-superior models. This method reduces dependency on any single forecast model, enhancing the robustness of the results. Additionally, recognizing climate change as a non-uniform phenomenon, our approach emphasizes the importance of analyzing climate data from a regional perspective, providing differentiated predictions to address the complexities of a heterogeneous future. This regional focus underscores the necessity of accounting for spatial disparities to better assess risks and develop effective policies for mitigation, adaptation, and compensation. Finally, this paper advocates that future climate agreements and policymakers should prioritize analyzing the entire temperature distribution rather than focusing solely on average values.

JEL classification: C31, C32, Q54
Keywords: Climate change; long-run climate forecast; density forecast; realized quantiles; trends; forecast combination; global warming; heterogeneous climate change.

*We would like to thank participants of the Fourth conference on Econometric Models of Climate Change (2019, Milan) and the CFE meeting (2019, London). We are grateful for Financial support from the grants PID2019-104960GB-100, PID2023-147593NB-I00, PID2020-114646RB-C44, PID2023-150095NB-C44, RED2018-102563-T, RED2022-134122-T, CEX2021-001181-M and TED2021-129784B-I00 funded by MCIN/AEI/ 10.13039/501100011033.

† Department of Applied Economics, University of Zaragoza. Gran Vía, 4, 50005 Zaragoza (Spain). Tel: +34 9767 61842, fax: +34 976 761840 and e-mail: lgadea@unizar.es

‡ Department of Economics, University Carlos III, Madrid 126 28903 Getafe (Spain). Tel: +34 91 6249853, fax: +34 91 6249329 and e-mail: jesus.gonzalo@uc3m.es

1 Introduction

Climate change has emerged as a critical challenge for humanity, drawing extensive academic attention and significantly influencing political decision-making. Governments, policymakers, stakeholders, international institutions, central banks, and other key actors in political and economic spheres are deeply engaged in analyzing its impacts, assessing associated risks, and understanding its far-reaching consequences. According to multiple reports by the World Economic Forum (2024, 2025), climate change underlies 5 of the top 10 long-term global risks. It plays a pivotal role in shaping globalization trends, reshaping geopolitical dynamics, influencing international agreements, driving large-scale involuntary migration, and integrating into financial risk analyses. In this context, accurate climate forecasting becomes imperative for designing more effective mitigation, adaptation, and compensation policies. Given the long-term nature of climate change, the development of long-term climate forecasts is a crucial objective. This paper focuses on this issue.

This paper aims to produce long-term forecasts for various characteristics of temperature distributions—such as quantiles and volatility—moving beyond the traditional emphasis on the mean, which has been the primary focus of most climate projections (see Chapters 11 and 12 of the IPCC-AR5, Collins et al., 2013 and Chapters 10–12 and the Atlas of IPCC-AR6, 2021). While average temperature predictions are useful, they often fail to capture critical details required for assessing phenomena such as ice melting, sea-level rise, and risk evaluation. A more comprehensive understanding of temperature distribution, through density forecasts, is vital.[1] To achieve this, the paper employs the realized quantile methodology developed by Gadea and Gonzalo (2020) (GG2020) to produce long-term temperature density forecasts based on historical observational data. This approach is applied globally, using data from climate stations worldwide, and regionally, by narrowing the analysis to specific geographic areas. In this methodology, quantiles and other distributional characteristics are transformed into time series objects, allowing them to be forecast similarly to mean temperature time series.

The existing literature on long-term forecasting in economics is relatively sparse, likely due to the substantial uncertainty inherent in long-term predictions and the limitations of stationary process models, where long-run forecasts often converge

[1] Note that the density forecast of the temperature is different from the density forecast of the average temperature.

to the unconditional mean.[2] An exception is Müeller and Watson (2016), who developed a method to quantify uncertainty in long-term economic predictions by analyzing low-frequency components of time series data. Their approach focuses on long-term averages and incorporates both Bayesian and frequentist perspectives.[3] Clemens and Galvão (2024) provide a survey of the most recent forecasting methods in macroeconomics that include the risks of climate change. More specifically, Castle et al. (2024) offer an overview of econometric contributions to climate change and its macroeconomic risks.

Projections in climate change literature primarily relies on complex climate models that simulate scenarios based on various assumptions. While these models are highly explanatory, they are less suited for prediction. As Shmueli (2010) notes, conflating explanation and prediction is a common issue. Climate models provide valuable insights into the Earth's future climate but often lack rigorous evaluation against real-world data. When such evaluations are conducted (e.g., Hausfather et al., 2020), inconsistencies emerge, with some models predicting excessive warming and others underestimating it.[4]

These findings highlight the need for complementary methods to improve prediction, which is crucial for effective climate policy design. Econometric approaches, such as the one proposed in this paper, offer advantages like flexibility, ease of application, out-of-sample evaluation, and the ability to analyze the full temperature distribution, enriching the predictive analysis. An additional strength of this work lies in its attention to the heterogeneity of climate change. By acknowledging that climate change is not a uniform phenomenon across time, space, and distribution, this study provides valuable insights for the design of more effective mitigation and adaptation policies. It highlights the importance of incorporating local dimensions into global climate agreements, thereby addressing the challenges of "free rider" behavior. While it is important to recognize that forecasts do not imply causation, their analysis offers critical guidance for action and contributes to advancing causality-attribution methodologies. Note that in this paper we work with a "business as usual" (BAU) scenario and only with the temperature series without

[2]For a discussion on the challenges of long-horizon forecasts, see Kemp (1999), Stock (1996, 1997), Phillips (1998), Pesavento and Rossi (2006), and Granger and Jeon (2007).

[3]Other notable works include Rossi (2005), Stock (1996, 1997), Lee (2011), Pastor and Stambaugh (2012), and Raftery et al. (2012).

[4]Several climate models coordinate their updates around the IPCC schedule. These coordinated efforts are part of the Coupled Model Intercomparison Projects (CMIP). IPCC-AR5 presented CMIP5 climate models, while IPCC-AR6 presents new state-of-the-art CMIP6 models.

including explanatory variables that may implicitly be in its trend.

Before proposing models to forecast is important to consider some facts about temperature data. According to GG2020 and Gadea and Gonzalo (2025) is clear that most of the characteristics of the temperature distribution (mean, quantiles, etc.) contain a trend as a consequence of the warming process. It is not clear which particular trend. This is the reason for considering a set of different trend models jointly with two standard non-trend models: unconditional mean (a constant) and an AR process.

Forecasts are model dependent. Therefore our final forecasts will be a combination of forecast. How to combine forecast may be an important issue. We can use an ad-hock trimming combination forecast or as we propose in this paper combine only the models that satisfy certain properties. In our case we propose to combine only the Pareto superior model for a give horizon "h". The weights of the combination are those used in the model average literature, an inverse of the Bayesian Information Criterion (BIC) value for a given model (Claeskens and Hjort, 2012).

This study employs two forecasting methods: a standard approach focused on a specific characteristic of the temperature distribution, and a density forecast that captures the entire distribution. For explanatory purposes, let us focus on a given distributional characteristic C_t (mean, quantiles, etc.). The standard forecast methodology for any distributional characteristic C_t involves the following steps:

1. Model selection: estimate a set of models $(m_1, \ldots, m_M)$ for the time series C_t ensuring that the models capture key distributional trends (see results in GG2020).

2. Forecast construction: generate forecasts of C_t for horizons $h = (1, 10, 25, 50, 100, ..)$ by using the direct forecasting method, which regress C_t at time "t+h" on model information about the regressors at time "t".

3. Model evaluation: Select the best model using BIC.

4. Model combination: combine forecasts and remove poorly performing models (poisoning models).

5. Forecast competition: Conduct an out-of-sample forecasting competition for each horizon "h", selecting Pareto-superior models based on forecast evaluation tests.

6. Final forecast: Combine forecasts from Pareto-superior models to minimize model forecast dependency.

For long-term density forecasting, the procedure is extended to include predictions for various quantiles of the temperature distribution:

1. Select a set of quantiles that represent the temperature distribution. This can be done for any quantile of the temperature distribution producing a wider-angle picture of the future than if we only focus on the mean. The forecast of all the quantiles is equivalent to a density forecast.

2. Select Pareto-superior models for all quantiles or, at their second best, for the largest number of quantiles.

3. Combine the winning models using the BIC criterion.

The proposed methodology has been applied to the Globe over an extended period, 1880-2023, and at a regional level, dividing the globe into eight geographical zones with the continents as the primary reference. The main forecast results obtained in this paper relate to the pre-industrial period (1850-1900) mentioned in the Paris climate agreement and other baseline periods used in the IPPCC-AR5 and AR6, 1986-2005 and 1995-2014, respectively.

Key findings include:

- Global Trends: Mean temperature is predicted to increase by over 2°C by 2050 and approximately 3°C by 2123. Lower quantiles (e.g., *q05*) indicate more substantial increases (4–6°C), while higher quantiles (*q95*) show smaller increases (1.5–2.2°C).[5] These are conservative predictions when the whole sample is considered and the acceleration processes suffered since 1960 is not taken into account.

- Variance Trends: Long-term forecasts suggest a decline in temperature variance, consistent with findings by Gadea and Gonzalo (2020) and Diebold and Rudebusch (2019).

[5]For context, the Earth's atmosphere during the last Ice Age was about 4°C cooler than pre-industrial times (Lynas, 2007).

- Regional Disparities: Predictions underscore substantial regional heterogeneity in warming patterns. The Arctic is expected to experience the most pronounced warming, followed by Europe and Asia. In Europe, both the lower and upper quantiles, particularly in Southern Europe, are projected to increase by more than 4°C. North America and Africa exhibit distinct warming trends. In North America, lower temperature quantiles are anticipated to rise more significantly than upper quantiles. Conversely, Africa is expected to experience greater increases in upper temperature quantiles, potentially exceeding 2°C. Australia is forecasted to undergo comparatively milder warming, while South America is projected to see even less pronounced temperature increases. Notably, Antarctica shows minimal evidence of significant warming based on current projections.

Compared to physical climate models (see Chapters 11 and 12 of the IPCC AR5 and AR6 reports), our mean temperature forecasts align closely with the intermediate scenario RCP4.5 or SSP5-4.5. These scenarios assume emissions peak around 2040, then decline by approximately 2045, reaching roughly half of 2025 levels by 2100. This alignment can be attributed to the conservative nature of our long-term forecasts, which are constructed using models based on the full historical sample (1880–2023). As a result, these models are less sensitive to the acceleration in warming observed during the late 20th century.

Furthermore, the use of the Atlas tool, as presented in the latest IPCC report, enables a more comprehensive comparison of climate model projections across various scenarios with observation-based predictions, moving beyond purely narrative analyses. Specifically, our forecasts of future temperature distributions across different horizons offer detailed insights into the nature and severity of warming experienced by different regions of the globe.

Two additional analyses enhance the methodology:

- Synthetic Control Experiment: This counterfactual predicts temperature trends using data up to 1960, before the exponential rise in greenhouse gas emissions. Results confirm the influence of emissions and highlight attribution effects.

- Pseudo-Real-Time Exercise: An out-of-sample prediction for 2000–2024 demonstrates the superiority of model combinations and highlights the smoothing effects on predictions amidst high short-term temperature variability.

In summary, this paper introduces a novel methodology for long-term temperature density forecasts, offering a complementary perspective to traditional physical climate models. By emphasizing distributional characteristics and accounting for regional heterogeneity, this approach provides deeper insights into climate dynamics, which are essential for comprehensive risk assessment and the formulation of effective policies. Importantly, the proposed methodology captures future heterogeneity both within and across regions, offering a more nuanced understanding than simple average temperature projections. Future research should aim to integrate these econometric approaches with climate model projections to further improve predictive accuracy and policy relevance.

The rest of the paper is organized as follows: Section 2 summarizes the theoretical framework and forecasting methodology. Section 3 presents the data and preliminary analysis. Section 4 discusses the empirical results, including the application of the standard forecasting methodology and the density forecast for global data over an extended period, as well as for regional data over a shorter timeframe. Finally, Section 5 outlines the conclusions.Tables and figures supporting the analysis are provided in Section 6. An Appendix includes the findings from the synthetic control experiment and the real-time exercise.

2 Econometrics Methodology

In this section, we first briefly summarize the econometric methodology introduced in GG2020 that will be used to analyze Global and Local warming processes. Second, we introduce the methodology to carry out the long-term forecast.

2.1 GG2020 approach

Following GG2020, Warming is defined as an increasing trend in certain characteristics of the temperature distribution. More precisely:

Definition 1. *(Warming)*: *Warming is defined as the existence of an increasing trend in some of the characteristics measuring the central tendency or position (quantiles) of the temperature distribution.*

An example is a deterministic trend with a polynomial function for certain values of the β parameters $C_t = \beta_0 + \beta_1 t + \beta_2 t^2 + ... + \beta_k t^k$, as well as, an integrated process of order one $(I(1))$[6].

In GG2020 temperature is viewed as a functional stochastic process $X = (X_t(\omega), t \in T)$, where T is an interval in $\mathbb{R}$, defined in a probability space $(\Omega, \Im, P)$. A convenient example of an infinite-dimensional discrete-time process consists of associating $\xi = (\xi_n, n \in \mathbb{R}_+)$ with a sequence of random variables whose values are in an appropriate function space. This may be obtained by setting

$$X_t(n) = \xi_{tN+n},\ 0 \leq n \leq N,\ t = 0, 1, 2, ..., T \tag{1}$$

so $X = (X_t, t = 0, 1, 2, ..., T)$. If the sample paths of ξ are continuous, then we have a sequence $X_0, X_1, ...$ of random variables in the space $C[0, N]$. The choice of the period or segment t will depend on the situation in hand. In our case, t will be the period of a year, and N represents cross-sectional units or higher-frequency time series.

We may be interested in modeling the whole sequence of $\mathbf{G}$ functions, for instance the sequence of state densities $(f_1(\omega), f_2(\omega), ..., f_T(\omega)\ )$ as in Chang et al. (2015, 2016) or only certain characteristics $(C_t(w))$ of these $\mathbf{G}$ functions, for instance, the state mean, the state variance, the state quantile, etc. These characteristics can be considered time series objects and, therefore, all the econometric tools already developed in the time series literature can be applied to $C_t(w)$. With this characteristic approach we go from Ω to $\mathbb{R}^T$, as in a standard stochastic process, passing through a $\mathbf{G}$ functional space:

$$\underset{(w)}{\Omega} \xrightarrow{X} \underset{X_t(w)}{\mathbf{G}} \xrightarrow{C} \underset{C_t(w)}{\mathbb{R}}$$

Going back to the convenient example and abusing notation ($X_t(n) = X_{tn}$ and n a natural number), the stochastic structure can be summarized in the following array:

[6] An $I(1)$ process is the accumulation of an $I(0)$ process. Our definition of an I(0) process follows Johansen (1995). A stochastic process Y_t that satisfies $Y_t - E(Y_t) = \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i}$ is called I(0) if $\sum_{i=1}^{\infty} \Psi_i z^i$ converges for $|z| < 1 + \delta$, for some $\delta > 0$ and $\sum_{i=1}^{\infty} \Psi_i \neq 0$, where the condition $\varepsilon_t \sim$ iid(0,σ^2) with $\sigma^2 > 0$ is understood.

$X_{00}(w) = \xi_0(w)$	$X_{01}(w) = \xi_1(w)$	...	$X_{0N}(w) = \xi_N(w)$	$C_0(w)$
$X_{10}(w) = \xi_{N+1}(w)$	$X_{11}(w) = \xi_{N+2}(w)$	...	$X_{1N}(w) = \xi_{2N}(w)$	$C_1(w)$
$X_{20}(w) = \xi_{2N+1}(w)$	$X_{21}(w) = \xi_{2N+2}(w)$	...	$X_{3N}(w) = \xi_{3N}(w)$	$C_2(w)$
.	.	...	.	.
.	.	...	.	.
.	.	...	.	.
$X_{T0}(w) = \xi_{(T-1)N+1}(w)$	$X_{T1}(w) = \xi_{(T-1)N+2}(w)$	...	$X_{TN}(w) = \xi_{TN}(w)$	$C_T(w)$

(2)

One of the objectives of this section is to provide a simple test to detect the existence of a general unknown trend component in a given characteristic C_t of the temperature process X_t. To do this, we need to convert Definition 1 into a more practical definition.

Definition 2. *(Practical trend definition)*: *A characteristic C_t of a functional stochastic process X_t contains a trend if in the LS regression,*

$$C_t = \alpha + \beta t + u_t, \ t = 1, ..., T, \tag{3}$$

$\beta = 0$ *is rejected.*

GG2020 shows that a simple $t - test(\beta = 0)$ is able to detect most of the existing deterministic trends(polynomial, logarithmic, exponential, etc.) and also the trends generated by any of the three standard persistent processes considered in the literature (see Müeller and Watson, 2008): (i) fractional or long-memory models $(1/2 < d < 3/2)$; (ii) near-unit-root AR models; and (iii) local-level models.

Several remarks are relevant with respect to this definition: (i) regression (3) has to be understood as the linear LS approximation of an unknown trend function $h(t)$ (see White, 1980); (ii) the parameter β is the *plim* of $\widehat{\beta}_{ols}$; (iii) if the regression (3) is the true data-generating process, with $u_t \sim I(0)$, then the OLS $\widehat{\beta}$ estimator is asymptotically equivalent to the GLS estimator (see Grenander and Rosenblatt, 1957) and the $t - test(\beta = 0)$ is *N(0,1)*; (iv) in practice, in order to test $\beta = 0$, it is recommended to use a robust HAC version of $t_{\beta=0}$ (see Busetti and Harvey, 2008); and (v) this test only detects the existence of a trend but not the type of trend. Notice also that in (3) we could be totally agnostic about u_t being *I(0)* or *I(1)*. In this case following Perron and Yabu (2009a) we can estimate the model by Feasible Generalized Least Squares and construct a similar *t-stat* of $\beta = 0$ that sitll will follow a *N(0,1)*. This method depends on a tuning parameter (how closs is u_t of being *I(1)*). To avoid that, in this paper, we follow an alternative approach. We pre-test the temperature data for unit roots, once they are rejected we proceed as in 3 u_t is *I(0)*.

For all these reasons, in the empirical applications we implement Definition 2 by estimating regression (3) using OLS and constructing a HAC version of $t_{\beta=0}$ (Newey and West, 1987). In the definition of C_t we can consider any distributional characteristics as time series objects. Note that this set comprises both the quantiles that make up the distribution and other characteristics that may be of interest, allowing for a much richer analysis of temperature than is possible with the mean alone.

2.2 Long-term forecast strategy

Predictive ability is one of the most valued qualities of any econometric model and serves as a highly useful tool for model selection.[7] Applying this principle to our observational climate models, we begin by proposing a set of candidate models to identify the most effective ones for forecasting. It is important to emphasize that the test proposed by GG2020, despite its simplicity, demonstrates strong power in detecting a wide variety of trends, which do not necessarily need to be linear. As indicated in GG2020's results, most of these models incorporate a trend component that captures the presence of global warming. The remaining models, which lack a trend component, can be viewed as a control group representing scenarios without global warming. Nevertheless, even models without a trend component exhibit a degree of persistence in their predictions.

After selecting the models to be part of the competition, we adopted two forecasting approaches: standard average forecast and density forecast.

2.2.1 Competing models

To run our model competition, 13 models have been selected, reflecting the main trends noted in the literature.

1. Unconditional mean model (*mean*)

2. Linear trend model (*linear-trend*)

[7] Prediction has always been a central challenge in econometrics, resulting in an extensive body of literature. Comprehensive references include Elliott and Timmermann (2006) and Clarke and Clark (2018). However, when focusing on long-term prediction, the number of references diminishes significantly. Among the most notable works are Müeller and Watson (2008).

3. Polynomial trend model (*pol-trend*)
4. Polynomial trend model average slope (*pol-trend-av-sl*)
5. Logarithmic polynomial trend model (*pol-trend-log*)
6. Structural breaks model (*struct-breaks*)
7. Polynomial trend autoregressive model (*pol-trend-arp*)
8. Polynomial trend autoregressive model average slope (*pol-trend-arp-av-sl*)
9. Autoregressive model (*arp*)
10. Random walk model (*rw*)
11. Random walk model with drift (r*wd*)
12. Ima model (*ima*)
13. Fractional model (*arfima*)
14. Large autoregressive model (*arp20*)

A detailed description of these models is given in Table 3. Notice that although most models consider different types of trend, for the sake of completeness we have included two additional models without trend: *mean* and *arp*. *arp20* captures possible long-memory different from the one described by a fractional model.

Other potential candidates were evaluated but ultimately dismissed due to their inadequacy as long-term predictors. This is exemplified by spline functions. The selection of appropriate nodes and boundary conditions poses significant challenges, and their application exhibits a strong dependence on the sample's endpoints.

Let's denote by C_t any distributional c*h*aracteristic (mean, max, min, std, iqr, rank, kur, skw, *q05*, *q10*, *q20*, *q30*, *q40*, *q50*, *q60*, *q70*, *q80*, *q90*, *q95*) as a time series object. The general *pol-trend-arp* model we propose for C_t is:

$$C_t = \sum_{i=0}^{k} \beta_i t^i + \sum_{j=1}^{p} \phi_j C_{t-j} + u_t \tag{4}$$

where k and p denote the order of the polynomial trend and the autoregressive part, respectively.

The *average* $-$ *slope* models are time varying linear trend models derived from their corresponding polynomial trend models in the following way:

$$C_t = \beta_0 + \tau_T t + u_t \tag{5}$$

where

$$\tau_T = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial t}C_t. \tag{6}$$

An example: $C_t = \beta_0 + \beta_1 t + \beta_2 t^2 + U_t$, $\tau_T = \beta_1 + 2\beta_2 \sum_{t=1}^{T} t/T$.

2.2.2 Standard average long-term forecast

In the standard long-term forecast we work individually with a given characteristic, for instance the average. We apply the following roadmap that comprises three steps. In the first, we select one of the trend models in-sample (via BIC) and forecast out-sample (h=1, 10, 25, 50, 100, ..., etc., depending of the sample size and objectives) by using the direct method.[8] The direct forecasting method involves producing a specific "h-steps-ahead" forecast without calculating intermediate steps. For a given model, it uses the information the model has at time "t" about the horizon "t+h". Two examples:

1. AR(1)

$$\widehat{C}_{t+h} = \widehat{\alpha}_{o,h} + \widehat{\phi}_{1,h} C_t$$

2. linear trend

$$\widehat{C}_{t+h} = \widehat{\alpha}_{o,h} + \widehat{\phi}_{1,h}(t+h)$$

Secondly, following a common practice in the forecasting literature because forecast are model dependent, we combine the considered models.[9] Some in-sample non-linear trend models do "strange" things out-of-sample in the long-run. For this, we need a combination forecast method, combining all the models, trimming the extreme models, or selecting only the "Pareto" forecasting superior models, etc. At this point, it is necessary to reflect on what type of information we obtain when we produce point forecasts of the mean (*q05* or other characteristic)

[8] Note that the prediction horizon will depend on the sample size and in some models will be constrained by the order of the autoregressive.

[9] The combination of models for forecasting was initially proposed by Bates and Granger (1969) and has been discussed in the literature on numerous occasions. Recent surveys on this technique can be found in Atiya (2020) and Wang et al. (2022).

time series object and its confidence intervals, because we are not saying much about the future temperature distribution. For instance, when we use the mean time series, we are forecasting the population mean distributional characteristic and nothing is said about the lower or upper quantiles of the whole temperature distribution. Following the model averaging literature (Claeskens and Hjort, 2012), forecasts model are combined using weighted average, where weights calculated as follows: $w_m = \frac{e^{-1/2BIC_m}}{\sum_{m=1}^{M} e^{-1/2BIC_m}}$ where BIC_m is the BIC criterion of each model from $m = 1...M$ in sample.

Thirdly, for each predictive horizon, "h", we select the "Pareto" forecasting superior models via an out-of-sample forecasting competition, using the Giacomini and White (2006) rolling window test (GW test). A model is defined as *Pareto-superior when it beats at least one other model and is not beaten by any other model in the competition.* We use the specification selected for each type of model and re-estimate the parameters in each iteration. The GW test is used to arrange the conditional predictive ability of the different models. This test, unlike that of Diebold and Mariano (1995), has the advantages of capturing the effect of estimation uncertainty on relative forecast performance and handling forecasts based on both nested and non-nested models.

The null hypothesis proposed by GW to compare the accuracy of two competing forecasts $f_t(\beta_{1t})$ and $g_t(\beta_{2t})$ for the h-steps-ahead variable Y_{t+h}, using the loss function $L_{t+h}(.)$ is

$$H_0 : E[L_{t+h}(Y_{t+h}+, f_t(\widehat{\beta}_{1t})) - L_{t+h}(Y_{t+h}+, g_t(\widehat{\beta}_{2t}))|G_t] = 0 \tag{7}$$

This proposal differs from Diabold and Mariano's in that the loss function depends on the estimates $\widehat{\beta}_{1t}$, $\widehat{\beta}_{2t}$ rather than their probability bounds and, in addition, the experiment is conditional on some information set G_t.

Confidence intervals are constructed by standard methods for individual forecasts and following this procedure for combinations:

$$\begin{aligned}
\widetilde{c}_{t+h} &= \sum_{m=1}^{M} \varpi_m f_{t,m}(h) \\
CI_{0.05} &= \sum_{m=1}^{M} \varpi_m f_{t,m}(h) + 1.96\sqrt{var(\sum_{m=1}^{M} \widehat{\epsilon}_{t+h,m})} \\
var(\sum_{m=1}^{m} \epsilon_{t+h,m}) &= \sum_{m=1}^{M} \varpi_m \widehat{\sigma}_m^2 + 2\sum_{m \neq j} \varpi_m \varpi_j \widehat{\sigma}_{m,j}
\end{aligned}$$

where $\widehat{\epsilon_{t+h,m}}$ are the estimated errors and $\widehat{\sigma}_m^2$ their estimated variance.

Summarizing, the strategy of the standard forecast comprises the following steps:

(1) Forecast with BIC selected models

(2) Combine models

   a) Combine all models, detect and remove poisonous models

   b) Apply a Pareto-superior model strategy with a rolling competition by using GW test; in this case, in addition to combining the winning models with the BIC weights, we have also used the simple average and the method proposed by Bates and Granger (1969), obtaining very similar results.

2.2.3 Long-term density forecast

Using the method described in the previous section for the average, it is possible to generate predictions for any characteristic of interest within the temperature distribution. More importantly, our approach enables predictions for the entire distribution, providing a proxy for the density forecast of temperature.[10] This methodology allows for the selection of quantiles that are most representative of the temperature distribution, such as a subset *q05*, *q50*, *q95*, or a complete set of quantiles, *q05*, *q10*, *q20*, *q30*, *q40*, *q50*, *q60*, *q70*, *q80*, *q90*, *q95*. This strategy can be applied either individually to each quantile or characteristic, identifying the Pareto-superior models for each one, or jointly by selecting models that are superior across all characteristics

[10]The literature on density forecasting is extensive. Notable contributions include Granger and Timmermann (2006), Corradi and Swanson (2006), Hall and Mitchell (2007), Rossi (2014), Rossi and Sekhposyan (2014), Kapetanios et al. (2015), and Ganics (2017), among others.

of the analyzed subset. In this paper we use the latter approach. A Pareto-superior model is defined as one that outperforms at least one other model while being outperformed by none, according to the competition based on the GW test.[11] It is important to highlight that this competition is conducted for various window sizes and forecasting horizons.

Specifically, the following strategy has been implemented in the empirical analysis. Once the ability of each model *m=1,...,M* is analyzed for each characteristic, we look for those that are Pareto superior for all the characteristics ($q_1 ... q_j ... q_Q$). If this is not possible and we obtain an empty set, we select the model(s) that are Pareto superior for the largest number of characteristics.

Q/mod	mod_1	...	mod_i	...	mod_M
q_1	I_{11}	...	I_{1i}	...	I_{1M}
...	...	...	...	...	...
q_j	I_{j1}	...	I_{ji}	...	I_{jM}
...	...	...	...	...	...
q_Q	I_{Q1}	...	I_{Qi}	...	I_{QM}
	$IP_1 = \sum_{j=1}^{Q} I_{j1}$		$IP_i = \sum_{j=1}^{Q} I_{ji}$		$IP_M = \sum_{j=1}^{Q} I_{jM}$

This matrix represents the decision. For each model m, the indicator $I_{jm} = 1$ if that model is Pareto-superior in quantile j. Each model m will be chosen as first-best if $\sum_{j=1}^{Q} I_{jm} = Q$; the second-best will be those that maximize the previous expression. A heatmap is a good option representing this decision matrix. The winning models can be combined in 3 possible ways: with the BIC weights, with the simple average or using the method in Bates and Granger (1969).

Two alternative methods were considered:

- method 0: The Pareto-superior models are chosen for each quantile.
- method 1: We select those models that are Pareto superior for all the characteristics of the set of interest. The loss functions of the different quantiles are aggregated over quantiles and the GW test is applied, $L_{t+h}^{m} = \sum_{j=1}^{Q} L_{t+h}^{j}$ where $m = 1, ..., M$ is the model and Q the number of quantiles. It is also possible

[11]This approach shares certain similarities with Hansen et al. (2011).

in a second step to find the optimum of the chosen set of models as the one that beats the largest number of models, although this would imply reducing the number of models selected.

Method 0 and method 1 are not being included for reasons of space.

At this point some considerations should be taken into account. Firstly, our approach differs from others such as Adrian et al. (2019) who obtain by applying a quantile regression the density of the mean temperature, not the density of the temperature. By means of a simple exercise we will show in the empirical part both results. It can be show that our approach has a quantile regression (qr) equivalent. It is equivalent to performing the following qr-regression of Y=temperature on X=trend where N is the number of units (stations):

years	Y	X
$t=1$	$temp_{11}$	1
...	...	...
$t=1$	$temp_{1N}$	1
$t=2$	$temp_{21}$	2
...		...
$t=2$	$temp_{2N}$	2
...	...	...
$t=T$	$temp_{T1}$	T
...		...
$t=T$	$temp_{TN}$	T

However, our approach has important advantages. It is more parsimonious and allows us to obtain not only the trend coefficients but explicitly the series of all the desired characteristics, which greatly expands the analytical capacity.

3 The data

In this paper we adopt a cross sectional perspective and work with data from the CRU. The CRU offers monthly and yearly data of land and sea temperatures in both hemispheres from 1850 to the present, collected from different stations around the world. HadCRUT5 is a global temperature data set, providing gridded temperature anomalies across the world, as well as averages for the hemispheres and for the globe as a whole. CRUTEM5 and HadSST4 are the land and ocean components of this overall data set, respectively.[12] This database (in particular, the annual temperature of the Northern Hemisphere) has become one of the most widely used to illustrate GW from records of thermometer readings. These records form the blade of the well-known 'hockey stick" graph, frequently used by academics and other institutions, such as, the IPCC. In this paper, we prefer to base our analysis on raw station data.[13] These data show high variability at the beginning of the period, probably due to the small number of stations in this early stage of the project, as noted by Jones et al. (2012). Following these authors, our study period begins in 1880 and ends in 2023.

Although there are over 10,000 stations on record in the last update, the effective number fluctuates each year during the period 1850-2023. To guarantee the stability of the characteristics over the whole sample, we select only those month-stations units with data for all years in the sample period, which forces us to reduce the sample size. We have also removed stations that present problems of inhomegeneities (Jones et al., 2012). Applying this procedure to the sample period 1880-2023, we have N=1230 month-stations units belonging to 135 stations. These characteristics are constructed for each year using monthly temperature records. Note that a benefit of using stable raw station data is that we always have perfect knowledge of every observation, and can easily detect the origin of any extreme observations or outliers.[14] In summary, we analyze raw global data (stations instead of grids) for the period 1880 to 2023. However, for reasons of homogeneity and stability, we use only data from stations that are represented in the whole sample period for each month and year.

In addition to the globe analysis, we adopt a regional perspective to make more accurate predictions at the regional level. To do so, we divide the globe into eight geographical regions: the Arctic Polar Circle, Europe, North America, South America, Asia, Africa, Australia and Antarctic. The partitioning of the sample into smaller geographical areas significantly reduces the number of stable observations, especially

[12]These data sets were developed by the Climatic Research Unit (University of East Anglia) in conjunction with the Hadley Centre (UK Met Office), with the exception of the sea surface temperature (SST) data set, which was developed solely by the Hadley Centre. We use CRUTEM version 5.0.1.0, which can be downloaded from (https://crudata.uea.ac.uk/cru/data/temperature/).

[13]Gadea and Gonzalo (2024b) discuss the advantages and disadvantages of using stations versus grids.

[14]A more detailed description of the construction process of the series can be found in Gadea and Gonzalo (2020).

in some areas. Therefore, for the regional analysis, the sample has been reduced to the period 1960-2023, which guarantees a reasonable number of observations. Furthermore, this period is the subject of the majority of climatological studies due to the intensity with which the phenomenon of climate change is beginning to manifest itself. Applying the strategy described in the previous paragraph, the unit month-season number for the globe is 10,897, 469 for the Arctic, 3,208 for Europe, 2,382 for North America, 344 for South America, 4,590 for Asia, 249 for Africa, 417 for Australia and, finally, 63 for the Antarctic. It should be noted that this is not an exact political classification but a geographical one, so the stations are selected according to a rectangle formed by the latitude and longitude that circumscribes each continent. In this way, the latitude variable is given priority over other classifications that strictly consider the political variable. Figure 1 shows the distribution of the selected stations according to the unit month-station criterion for the whole globe for the periods 1880-2023 and 1960-2023, and Figure 2 the distribution for the different geographical regions into which we have divided the globe for the second period.

3.1 Preliminary analysis: A first look of the data and their trends

To illustrate the application of our approach, it is helpful to visualize the evolution of temperature data distributions over time. Figure 3 displays the density of raw temperature data for the entire globe during the 1880–2023 period. From this, the distributional characteristics of interest can be computed and transformed into time series objects. Figure 4 shows the temporal evolution of these characteristics (*mean*, *max*, *min*, *std*, *iqr*, *rank*, *kur*, *skw*, *q05*, *q10*, *q20*, *q30*, *q40*, *q50*, *q60*, *q70*, *q80*, *q90*, *q95*). Finally, Figure 5 shows the geographical location of some quantiles of interest.

Before conducting the forecasting exercise, we test the presence of unit roots in the time series of distributional characteristics and test for the existence of trends.[15] The results are presented in Tables 1 and 2. Most characteristics show no evidence of unit roots; the null hypothesis cannot be rejected in only 6 out of 190 analyzed series. Nevertheless, there is a trend in all the characteristics. This finding supports the stationarity of temperature characteristics around a deterministic trend, a key assumption for the rest of the analysis.[16]

[15]The null hypothesis of a unit root is rejected in 96.8% of cases (characteristics and regions), a value lower than the size of the test.

[16]This result challenges the common belief that global temperature exhibits non-stationary be-

Regarding global warming, the results can be summarized as follows:

1. Trends in central and position distributional characteristics: Most measures of central (mean, median) and position (quantiles) exhibit an increasing trend which confirms the existence of a clear warming process.

2. Lower Quantile Intensification: The increasing trend is more pronounced in the lower quantiles than in the mean, median, or upper quantiles.

3. Dispersion Trends: Features measuring dispersion indicate a decreasing trend, meaning lower temperatures are converging toward the median faster than higher temperatures.

4. Acceleration Phenomenon: The magnitude of these trends has accelerated significantly in the 1960–2023 period compared to 1880–2023. For instance, the coefficients for the mean, *q05*, *q50* and *q95* were 0.0117, 0.0182, 0.0101 and 0.0078, respectively, during 1880–2023. These values increased to 0.0304, 0.0468, 0.0251 and 0.0216, respectively, in the later period.

When the analysis is extended to the regional level, the findings confirm the heterogeneity of climate change. Key observations include:

- Arctic Circle: The Arctic Circle demonstrates the most significant warming process, with higher intensity in lower quantiles and decreasing dispersion.[17]

- Europe: Europe exhibits the second-highest degree of warming, though the trend magnitude across quantiles, lower and upper, is relatively uniform. Decreasing trends in dispersion are not significant.

- Asia and North America: These regions follow Europe in warming intensity, with patterns resembling the Arctic Circle—greater warming in lower quantiles and significant decreases in dispersion.

havior, which often arises from the use of grid-analyzed mean temperature series with a heavy concentration in the Northern Hemisphere. Our hypothesis, supported by simulation exercises, is that spatial and temporal aggregation effects artificially increase the persistence of such series, Gadea et al. (2024a).

[17] Due to its critical role in global climatological processes, the Arctic Circle has been analyzed specifically by Gadea and Gonzalo (2024b), who identify clear acceleration and amplification phenomena.

- Australia and South America: These areas show lower levels of warming without discernible differences across quantiles or significant trends in dispersion measures.

- Africa: While Africa's average warming is comparable to that of North America, the warming is more pronounced in the higher quantiles than the lower quantiles, accompanied by an increase in dispersion.

- Antarctica: Antarctica shows no significant evidence of warming, but this may result from data limitations. The available stable sample comprises stations concentrated in East Antarctica, while climatologists have identified warming trends in West Antarctica and regions closer to South America. The IPCC's Fourth Assessment Report concluded that Antarctica was the only continent without detected anthropogenic warming, likely due to sparse data collection.[18]

4 Results

In the empirical application, we detail all the steps of our approach using global observations from 1880 to 2023, focusing initially on the average temperature. Subsequently, we compute the density forecast for a comprehensive set of quantiles, Q=$\{q05,q10,q20,q30,q40,q50,q60,q70,q80,q90,q95\}$ employing the methodology introduced in the previous section. For the regional data, we directly compute forecasts for our selected set of quantiles and analyze the differences between predictions at varying horizons and the mean across the two reference periods typically considered by the IPCC. This approach enables us to relate the projections of global warming across different geographical regions generated by climate models with our predictions.

4.1 The Globe, 1880-2023

Focusing on the global long-term dataset, we conducted a rolling analysis using a window of w=100 years and several prediction horizons, $h = 1, 10, 25, 50, 100$ years.[19]

[18]Based on this and related analyses, Gadea and Gonzalo (2025) propose a typology of climate change and introduce the concept of "Warming Dominance", which enables comparisons between regions. The former characterizes heterogeneity inside regions and the latter between regions.

[19]It is important to note that these prediction horizons are used to compare model performance across different time intervals and are conditioned by factors such as sample size, window length,

The three comparison periods correspond to the pre-industrial era (1880–1900), the period commonly used in previous IPCC-AR5 reports (1986–2005), and the recent-past period referenced in IPCC-AR6. Our forecast horizons align with the so-called near-term and mid-term projections. Additionally, forecasts were generated for a specific long-term horizon, 2100 ($h = 77$), with temperature increases calculated for each comparison period. This enables direct comparisons with the projections provided in IPCC.

4.1.1 Standard forecast for the average temperature

In this section, we adopt a standard forecasting approach for the average temperature treated as a time series object. This methodology will be applied to this characteristic using the 1880–2023 sample, which includes all stations with continuous data throughout this period.[20] Initially, we rely on the BIC to guide the selection process without incorporating the results from the inter-model competition. In the subsequent step, we present the results based on the models identified as Pareto-optimal during the competition. Because forecast are clear model dependent, we propose forecast combination in both cases.

Table 4 presents the BIC values and their corresponding weights. The preferred model in sample is the *pol-trend*; consequently, Table 5 includes this model, its variant *pol-trend-av-sl*, and the *linear trend* model as a benchmark.[21] The results predict a temperature increase of over two degrees with the *linear trend* model and over three degrees with the *pol-trend-av-sl* model over 100 years relative to the baseline period.

Since the predictions depend on the model, Table 6 presents several model combinations that progressively eliminate those with undesirable behavior. As shown, "combined3" produces results very similar to the *pol-trend-av-sl* model.

Instead of combining forecasts in an ad-hoc manner, our approach provides an alternative by identifying Pareto-superior models. Within this framework, the aver-

and the specific characteristics of certain models, particularly those with autoregressive components. The selected horizons can be applied to short-term forecasts or extended to very long-term predictions.

[20] Although the mean has been chosen because it is the most studied characteristic in the climate literature, this procedure can be applied to any other characteristic. Results are available upon request from the authors.

[21] It is important to note that the best in-sample model does not necessarily produce the best out-of-sample predictions. For instance, the *pol-trend* model generates predictions that lack meaningful interpretation. To address this issue, we introduce its counterpart, the *pol-trend-av-sl*.

age temperature is projected to increase by slightly more than two degrees over the next 25 years, relative to the baseline model established by the Paris Agreement.

Two points merit attention. First, as shown in Table 7, the Pareto-superior model selection avoids extreme-behavior models, unlike earlier approaches that did not filter out such models. Second, although rolling competitiveness imposes constraints where the prediction horizon cannot exceed w-p (where w is the rolling window size and p is the autoregressive order), it remains feasible to select a model for this horizon and extend predictions to longer horizons.

4.1.2 Temperature density forecast

Until now, we have focused on generating forecasts for the average. However, these forecasts provide an incomplete picture. Our methodology allows us to perform similar forecasting exercises for any quantile and approximate the forecast density from a set of selected quantiles using the method described in the methodological section. If more than one Pareto-superior model is identified, these are combined using rescaled BIC weights, and confidence intervals are calculated as outlined earlier.

The results obtained using a window size (w) of 100 and prediction horizons "h" of 1, 10, and 25 years are presented in Table 8 and Figure 6, illustrating the decision matrix at each horizon. This approach identifies models that are Pareto-superior across all quantiles, or, alternatively, second-best for the largest number of cases. The findings indicate that warming forecast generally increase with "h". However, for most quantiles, particularly the upper quantiles, the projections tend to stabilize in the medium term. It is important to note that selecting models specific to horizon "h" enhances prediction accuracy but may introduce temporal inconsistencies.[22] A similar issue arises in climate projections when the number of models or simulations varies across periods or scenarios.

To analyze the temperature increases at different horizons relative to selected reference periods, such as the pre-industrial period (1880–1900), the baseline period used in IPCC reports (1986–2005) in AR5, and the "recent past" (1995–2014) from AR6, we have included a long-term horizon (h=77) corresponding to 2100. The numerical results are summarized in Table 9 when confidence intervals are included.

[22]One approach for future research to addressing these temporal inconsistencies is to incorporate the prediction horizon "h" into the Pareto superior selection process (in the spirit of Martinez, 2017). This refinement ensures the selection of models that are Pareto superior across all quantiles and prediction horizons "h", thereby enhancing consistency and robustness in the results.

Figure 7 depicts the forecasted temperature increases of the different quantiles in 2100 relative to each reference period.

Key Conclusions:

- Global Warming Intensity: The intensity of warming varies across quantiles showing a clear future heterogeneity. Temperature increases in lower quantiles could reach 4°C above the pre-industrial period and over 2°C relative to the two most recent periods.
- Acceleration: The method demonstrates a clear acceleration in warming over time, as evidenced by the evolution of temperature increases across periods.
- Asymmetry: The warming process is asymmetric, with the largest increases occurring in the lowest quantiles. This indicates that the acceleration of warming is accompanied by a narrowing of the distribution (shrinkage effect) at the global level. Using our unconditional quantile methodology (see Figure 5), we can identify zones corresponding to *q70* and *q80* that exhibit larger temperature increases compared to the surrounding quantiles. These zones include locations such as Kansas, Madrid, and Marseille."

Comparison with IPCC AR6 Projections[23]:

- Direct Comparability with Scenario-Based Projections[24]: Our mean temperature forecasts closely align with the intermediate scenario RCP4.5 or SSP5-4.5. These scenarios assume emissions peak around 2040, decline by approximately 2045, and reach roughly half of 2025 levels by 2100. This alignment is likely due to the conservative nature of our long-term forecasts, which are based on models trained on the full historical sample (1880–2023). Consequently, these models are less sensitive to the accelerated warming observed in the late 20th century.

[23]The IPCC utilizes CMIP5 and CMIP6 as reference climate models in their respective assessment reports, AR5 and AR6. In the latter, the scenarios proposed are the Shared Socioeconomic Pathways (SSPs), which represent the most complex and comprehensive frameworks developed to date. These scenarios range from highly ambitious mitigation efforts (SSP1) to continued emissions growth (SSP5). Each SSP defines a distinct trajectory for future greenhouse gas emissions and land-use changes under specific baseline assumptions. The SSPs are an enhanced evolution of the Representative Concentration Pathways (RCPs)—RCP2.6, RCP4.5, RCP6.0, and RCP8.5—used in CMIP5 and AR5, allowing for a rough comparison between the two frameworks.

[24]The comparison of absolute values for both mean and density forecasts is affected by discrepancies between the historical series provided by the CMIP6 model and CRU observations.

- Insights from Density Forecasts: While the results of the density forecast are not directly comparable to IPCC projections, they offer valuable insights into the quantiles driving the average temperature increase. The most significant increases occur in the lower quantiles, which are typically associated with high latitudes (regions near the poles), as well as in quantiles *q70* and *q80*. This heterogeneity, as revealed by the density forecast, cannot be captured by the IPCC's average projections.

Unlike IPCC projections, our methodology does not rely on predefined scenarios. Instead, it follows a BAU approach. Some researchers equate this scenario with RCP8.5/SSP5-8.5, while others associate it with RCP4.5, which represents a more plausible trajectory if modest mitigation efforts or technological advances occur alongside aggressive policy actions. Our results align more closely with the latter (see Figure 8).

4.2 Regional approach

For various geographical regions, we present the results of the long-term density forecast for the globe and individual regions over the period 1960–2023, encompassing all quantiles and prediction horizons (h=1,10,25). To accommodate the smaller sample size, the rolling window was reduced to 25, allowing all models to compete, even as the prediction horizon extends to 2100 (h=77). Detailed predictions are provided in Table 10 for global forecasts, and in Tables 11-18 for the Arctic, Europe, North America, South America, Asia, Africa, Australia, and the Antarctic.

These results facilitate an analysis of temperature increases across regions and quantiles for the reference periods 1986–2005 (end of the 20th century) and 1995–2014 (beginning of the 21st century). Although primarily descriptive, Table 19 offers valuable insights into regional temperature level differences by presenting mean quantiles over the period and their most recent values. For instance, significant differences are observed, such as the Antarctic's -58.55°C compared to Africa's 10.35°C for quantile *q05*, and -0.90°C versus 31.15°C for *q95* within the same regions. Table 20 provides detailed temperature increases by quantiles and reference periods for each region for 2100, including confidence intervals, while Figures 9-12 offer a more illustrative depiction of temperature increases by quantiles and horizons over the two reference periods.

A key observation from these graphs is the pronounced heterogeneity in warming patterns across quantiles and regions. By regions, and focusing on the predictions for 2100 (Table 20 and Figure 12):

- Arctic: The Arctic stands out as the region experiencing the most significant warming, particularly in the lower quantiles. Temperature increases in these quantiles range from 5°C to 8°C, depending on the reference period. This extreme warming in the lower quantiles is a key driver of ice melt and future sea level rise.
- Europe: Europe, alongside Asia, exhibits substantial temperature increases, undergoing a dangerously rapid warming process. In Europe, temperature increases exceed 4°C in the lower quantiles and are slightly lower in the upper quantiles, resulting in a U-shaped distribution. This pattern is primarily driven by the intense warming observed in Southern Europe, particularly in Spain, Italy, and Greece.
- Asia: Asia follows a warming pattern similar to that of the Arctic, with temperature increases decreasing across quantiles and peaking at approximately 5°C.
- North America: While warming is evident, it is less pronounced compared to the Arctic, Europe, and Asia. The temperature increase is more prominent in the lower quantiles, whereas the upper quantiles exhibit only minimal warming. The U-shaped pattern observed in Europe is barely noticeable in North America.
- South America and Australia: These regions exhibit minimal or negligible warming, suggesting a less pronounced impact of temperature changes.
- Africa: Africa experiences significant warming at the upper end of the temperature distribution, in contrast to northern regions. Warming peaks at approximately 3°C in the upper quantiles, which is a major contributor to severe droughts across the continent.
- Antarctica: Antarctica shows near-zero or even negative temperature changes during the second reference period, indicating minimal or no warming.

Some methodological Observations: The forecasts highlight the critical importance of considering both quantile-based temperature distributions and the pronounced regional heterogeneity of future warming. These insights are essential for comprehensively understanding and effectively addressing the localized impacts of climate change. For example, the warming trend observed in the upper quantiles in Europe has been identified as a key factor contributing to heat-related mortality across the continent (see Masselot et al., 2025).

4.2.1 Comparing regional predictions with regional projections of Atlas

Since AR5, the IPCC has emphasized the critical importance of understanding the local dimensions of climate change while highlighting the inherent complexity of regional analyses. However, the methodology for regional climate assessments remains non-standardized, and the process of synthesizing regional climate information from multiple lines of evidence varies significantly across studies. While comparisons between climate model projections and observations are often made narratively, the practice of directly comparing observation-based predictions with climate model projections is less common. This section addresses this gap. Specifically, we utilize the Atlas introduced in AR6 (see Gutierrez et al., 2021), a tool that generates regional average projections based on various climate models, enabling deeper numerical insights into regional climate dynamics.

The Atlas project evaluates average climate conditions at regional scales, focusing primarily on temperature and precipitation over land areas.[25] The division of the globe into 14 regions, as depicted in Figure 13, facilitates comparisons with the regions defined in our study. Differences include the subdivision of Central and South America (CAM, SAM) and the segmentation of Asia into Central Asia (CAS), East Asia (EAS), South Asia (WAS), and Southeast Asia (SEA). CAS was excluded due to a limited number of simulations, as were the Mediterranean (MED) and North Africa (MNA) regions. Consequently, comparisons were made using EAS, which encompasses the majority of the continent.

The regions with available data and reproducible results are: Europe (EUR), Africa (AFR), South America (SAM), North America (NAM), East Asia (EAS), Australasia (AUS), Arctic (ARC), and Antarctic (ANT). The Atlas utilizes data from the CORDEX (Coordinated Regional Climate Downscaling Experiment) project

[25]The interactive Atlas is accessible at https://interactive-atlas.ipcc.ch/. Supplementary files, data sources, R scripts, and materials required for reproducing the results are available at https://github.com/IPCC-WG1/Atlas. A concise summary of these resources can be found at https://digital.csic.es/handle/10261/280324 (Iturbide et al., 2022). We extend our sincere gratitude to Professor Iturbide (Instituto de Física de Cantabria) for assistance in accessing these resources.

at the regional level, offering simulations based on multiple climate models. These models reproduce historical data from 1970–2005 and generate projections for 2006–2100 under RCP2.6, RCP4.5, and RCP8.5 scenarios, reflecting varying levels of emissions intensity.

The comparison of regional density forecast results with Atlas projections is presented in Figure 14. The conclusions are clear: in all cases, the projections for the three scenarios fall within the density distributions for both the 2050 and 2100 predictions. The relative position of these projections with respect to the median provides insight into whether our forecasts anticipate greater or lesser warming. If a scenario lies below the median, our forecasts indicate greater warming than the projections; conversely, if it lies above, our forecasts suggest less warming. Moreover, a greater distance between a projection and the median implies a stronger alignment of our predictions with that particular scenario.

Beyond the standard advantage of density distributions providing more information than mean values, our approach offers an additional benefit: in our case, quantiles correspond to specific zones within a given region. This added value underscores the strength of our unconditional quantile methodology.

By region, and focusing on the long-term (2100), the following observations emerge:

- Arctic: For both horizons, AR6 projections under RCP4.5 and RCP8.5 lie below the median of our density forecast, indicating that we predict greater warming with high probability. Alternatively, this suggests that our forecasts identify numerous zones within the Arctic experiencing more intense warming than projected by the mean.

- Europe: Europe exhibits a similar pattern, with the median of our forecast distribution located above the high-emissions scenario, suggesting greater warming than IPCC projections.

- North America: North America follows a comparable trend, with all three scenarios situated within the left tail of our density forecast distribution, implying a higher likelihood of greater warming.

- Asia: In contrast, our density forecast for Asia indicates less warming with high probability, or identifies numerous zones with lower warming compared to the projections. However, it should be noted that the EAS region excludes North Asia.

- South America: In the near term, the median of our density forecast aligns with the low-emissions scenario. For the long-term (h=77), our predictions suggest less warming than RCP8.5 but more than RCP2.6, with closer alignment to RCP4.5.

- Central America: The differing geographical delineation of Central America may affect results observed in North and South America.

- Africa: In Africa, projections lie within the right tail of our distribution. However, similar to South America, in the long-term, the median aligns with the most optimistic projection.

- Australia: Australia exhibits a pattern similar to South America, with the median lying between the projections for different scenarios, depending on the horizon.

- Antarctic: In the Antarctic, our forecasts indicate greater warming than the projections. However, these results should be interpreted cautiously due to the limited observational data and simulations available for this region.

Summary: The forecasts underscore the need to account for both quantile-based temperature distributions and the significant regional heterogeneity in future warming. This approach provides critical insights into localized climate dynamics and emphasizes the importance of integrating density-based predictions with scenario-based projections for a more comprehensive understanding of regional climate change.

However, a stylized fact emerges: projections below our predicted median are associated with colder regions, characterized by higher warming intensity in the lower quantiles.[26] In contrast, projections that exceed our predicted median are observed in warmer regions, which exhibit lower warming intensity across quantiles. The analysis of these results underscores the usefulness of our approach, as it provides richer predictive insights by considering the entire distribution rather than just

[26]A possible explanation for this result, particularly in the case of Europe, is that climate model projections may be biased by changes occurring in the lower quantiles. Unlike density estimates, these models may not fully account for the warming observed in the higher quantiles, particularly in the Mediterranean region.

the mean. The same average projection could indicate less or more future warming. Furthermore, our methodology achieves this with minimal implementation and updating costs.

5 Conclusions

Climate is inherently a long-term phenomenon, and designing effective mitigation and adaptation policies for climate change requires reliable long-term forecasts. However, long-term prediction poses significant challenges for both econometric and geophysical models. Econometric models, while requiring less data and offering greater flexibility by relying on observational information, are often surrounded by substantial uncertainty. Geophysical models, on the other hand, possess greater causal explanatory power but depend heavily on theoretical assumptions and predefined greenhouse gas concentration pathways. These models are primarily focused on mean temperature projections, limiting their capacity to explore broader distributional dynamics.

This paper introduces a simple novel methodology for generating long-term forecasts of temperature distributional characteristics beyond the commonly studied mean temperature. The approach builds on Gadea and Gonzalo (2020), treating temperature as a functional stochastic process from which various distributional characteristics (e.g., mean, quantiles, volatility) can be extracted as time series objects. Recognizing that these distributional characteristics often exhibit trend components, our out-of-sample forecasting competition incorporates 13 models designed to accommodate various types of trend elements. Because forecasts are model dependent and to address the uncertainty inherent in model selection, we propose long-term forecast combinations derived from Pareto-superior models, as determined by an accuracy evaluation test.

Our methodology is versatile, applicable to global cross-sectional data (monthly temperatures from diverse global stations) and regional analyses, dividing the globe into continents and other regions. Focusing on a global analysis for the period 1880–2023, our findings are clear: our mean temperature forecasts align closely with the intermediate scenario RCP4.5 or SSP5-4.5. These scenarios assume emissions peak around 2040, then decline by approximately 2045, reaching roughly half of 2025 levels by 2100. This alignment can be attributed to the conservative nature of our long-term forecasts, which are constructed using models based on the full historical sample (1880–2023). As a result, these models are less sensitive to the acceleration in warming observed during the late 20th century.

Importantly, our analysis of the Globe reveals that, behind the average temperature increases are the predicted increase in the lower quantiles. They are predicted to increase more than the upper ones. This asymmetry has more severe implications than the standard focus on mean temperature increases, emphasizing the need for a broader perspective on climate impacts. Furthermore, our methodology not only enables the estimation of the probability that the mean temperature will exceed or fall below a given value in the future but also identifies the specific quantiles—corresponding to latitudes—where the most significant increases are expected to occur.

A regional analysis for the period 1960–2023 highlights the substantial heterogeneity of climate change, which our density forecast methodology effectively captures. The results enable the establishment of a regional ranking based on the intensity and characteristics of future warming processes. The Arctic ranks first, exhibiting the highest numerical increases in temperature. Europe, in second place, displays a concerning U-shaped pattern due to the intense warming in the Southern Europe. Africa ranks third, marked by significant increases in the upper quantiles. Asia and North America, occupying the fourth position, share some similarities with Europe but demonstrate more moderate temperature increases and behavioral patterns. Australia and South America follow in fifth place, showing the least concerning warming trends. Lastly, Antarctica's future situation remains challenging to assess due to the scarcity of observational data."

These findings align with the projections provided in the recent IPCC AR6 report (e.g., the Atlas tool), while offering a complementary perspective based on observational data. This approach enhances flexibility and simplifies analysis, making it a valuable addition to traditional climate modeling frameworks.

Given these conclusions, we advocate for climate agreements to move beyond the sole focus on mean temperature. Greater attention must be paid to the full temperature distribution, particularly lower and upper quantiles, to better capture the breadth of climate impacts. The average temperature is not enough, we need the whole distribution. Furthermore, the design of mitigation, adaptation, and compensation policies should incorporate detailed regional analyses to address the uneven distribution of climate change effects. Future research should merge forecasting and atribution-causal models.

References

[1] Adrian, T., Boyarchenko, N., Giannone, D. 2019. Vulnerable growth. *American Economic Review* 109(4), 1263-89.

[2] Martinez, A. B. 2017. Testing for Differences in Path Forecast Accuracy: Forecast-Error Dynamics Matter. Working Paper 17-17, Federal Reserve Bank of Cleveland.

[3] Atiya, A.F. 2020. Why does forecast combination work so well? *International Journal of Forecasting*g 36, 197-200.

[4] Bennedsen, M., Hillebrand, E., and Lykke, J.Z. 2022. Global temperature projections from a statistical energy balance model using multiple sources of historical data.

[5] Botosaru, I. Giacomini, R., and Weidner, M. 2023. Forecasted Treatment Effects. arXiv:2309.05639v2 [econ.EM].

[6] Busetti, F. and Harvey, A. 2008. Testing for trend. *Econometric theory* 24, 72-87.

[7] Castle J.L., Hendry, D.F., Miller, J.I. 2024. Econometric Forecasting of Climate Change. In Clements, M.P., Galvão, A.B., (eds.) Handbook of Research Methods and Applications in Macroeconomic Forecasting Edition:1, Edward Elgar Publishing.

[8] Chen, L., Dolado, J.J., Gonzalo, J., Ramos, A. 2023. Heterogeneous predictive association of CO2 with global warming. Economica 90(360), 1397-1421.

[9] Claeskens, G. Hjort, N.L. 2012. Model Selection and Model Averaging. Cambridge University Press.

[10] Clarke, B.S. and Clarke, J.L. 2018. *Predictive Statistics. Analysis and Inference Beyond Models.* Cambridge University Press.

[11] Clements, M.P., Galvão, A.B. 2024. (eds.) Handbook of Research Methods and Applications in Macroeconomic Forecasting Edition:1, Edward Elgar Publishing.

[12] Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, 2013. Long-term Climate Change: Projections, Commitments and Irreversibility. In: *Climate Change* 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Chapter 12 of IPCC-AR5.

[13] Corradi, V. and Swanson, N. R. 2006. Chapter 5 Predictive Density Evaluation. In Elliott, G., Granger, C. W. J., and Timmermann, A., editors, Handbook of Economic Forecasting, volume 1, 197–284. Elsevier.

[14] Diebold, F. X. and Mariano, R.S. 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13, 253–263.

[15] Diebold, F.X., Rudebusch, G.D. 2019. for the US. On the Evolution of U.S. Temperature Dynamics. PIER Working Paper No. 19-012.

[16] Diffenbaugh1,, N.S., Field, Ch.B. 2013. Changes in Ecologically Critical Terrestrial Climate Conditions *Science* Volumne 341, Issue 6145, pp. 486-492 DOI: 10.1126/science.1237123.

[17] Elliot, G. 2006. Forecasting with Trending Data, in Elliott, C.W. J. G. G. and Timmerman A. (eds) *Handbook of Economic Forecasting*, Vol. 1, (Amsterdam: North-Holland), 555–604.

[18] Eyring, V., Cox, P.M., Flato, G.M., Gleckler, P.J., Abramowitz, G., Caldwell, P., Collins, W.D., Gier1, B.K., Hall, A.D., Hoffman, F.M., Hurtt1, G.C., Jahn, A., Jones, Ch. D., Klein, S.A., Krasting, J.P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl1, G.A., Pendergrass, A.G., Pincus, R., Ruane, A.C., Russell, J.L., Sanderson, B.M., Santer, B.D., Sherwood, S.C., Simpson, I.R., Stouffer, R.J., Williamson, M.S. 2019. Taking climate model evaluation to the next level. *Nature Climate Change* 9(2), 102-110.

[19] Gadea, M.D., Gonzalo, J. 2020. Trends in distributional characteristics: Existence of global warming. *Journal of Econometrics* 214, 153-174.

[20] Gadea, M.D., Gonzalo, J., Ramos, A. 2024a. Trends in Temperature Data: Microfundations of their Nature. Economic Letters 244, 111992.

[21] Gadea, M.D., Gonzalo, J. 2024b. Polar warming. mimeo.

[22] Gadea, M.D., Gonzalo, J. 2024c. Regional heterogeneity and warming dominance in the United States. Working Paper UC3M, 2415.

[23] Gadea, M.D., Gonzalo, J. 2025. Climate change heterogeneity: A new quantitative approach, *PLOS ONE* (forthcoming).

[24] Ganics, G.A. 2017. Optimal Density Forecast Combinations. Documentos de Trabajo 1751, Banco de España.

[25] Giacomini, R., White, H. 2006. Test of Conditional Predictive Ability. *Econometrica* 74, 1545-1578.

[26] Global Carbon Budget 2024. – with major processing by Our World in Data. "Annual CO2 emissions – GCB" [dataset]. Global Carbon Project, "Global Carbon Budget" [original data].

[27] Global Economic Forum 2024. The Global Risks Report 2024. 19the Edition. Insight Report.

[28] Global Economic Forum 2025. The Global Risks Report 2025. 20the Edition. Insight Report.

[29] Granger, C.W. and Jeon, Y. (2007), Long-Term Forecasting and Evaluation. *International Journal of Forecasting* 23, 539–551.

[30] Granger, C.W. and & Timmermann, A. (Eds.), *Handbook of economic forecasting*, vol. 1 (pp. 197–284). North Holland: Elsevier.

[31] Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. New York: Wiley.

[32] Gutiérrez, J.M., R.G. Jones, G.T. Narisma, L.M. Alves, M. Amjad, I.V. Gorodetskaya, M. Grose, N.A.B. Klutse, S. Krakovska, J. Li, D. Martínez-Castro, L.O. Mearns, S.H. Mernild, T. Ngo-Duc, B. van den Hurk, and J.-H. Yoon, 2021: Atlas. In Climate Change 2021: The Physical Science Basis.

Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1927–2058, doi:10.1017/9781009157896.021.

[33] Hall, S.G., Mitchell, J. 2007. Combining density forecast. International Journal of Forecasting 23, 1-13.

[34] Hansen, P.R., Lunde, A., Nason, J.M. 2011. The Model Confidence Set. *Econometrica* 79(2), 453–497

[35] Hausfather, Z., Drake, H. F., Abbott, T., & Schmidt, G. A. 2000. Evaluating the performance of past climate model projections. *Geophysical Research Letters*, 47, e2019GL085378.

[36] IPCC, 2021: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391.

[37] Instituto de Física de Cantabria (IFCA). 2022. IPCC-WGI AR6 Interactive Atlas Dataset; Version 2 final; https://digital.csic.es/handle/10261/280324.

[38] Iturbide, M., Fernández, J., Gutiérrez, J.M. et al. Implementation of FAIR principles in the IPCC: the WGI AR6 Atlas repository. Scientific Data 9, 629 (2022). https://doi.org/10.1038/s41597-022-01739-y

[39] Jones, P.D., Lister, D.H., Osborn, T.J., Harpham, C., Salmon, M., Morice, C.P., 2012. Hemispheric and large-scale land surface air temperature variations: an extensive revision and an update to 2010. Journal of Geophysical Research 117, 1-29.

[40] Kapetanios, G., Mitchell, J., Price, S., Fawcett, N. 2015. Generalized density forecast combinations. *Journal of Econometrics* 188, 150-165.

[41] Kemp, G. C. R. 1999. The Behavior of Forecast Errors from a Nearly Integrated AR(1) Model as Both Sample Size and Forecast Horizon Become Large. *Econometric Theory* 15, 238–256.

[42] Kirtman, B., S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore, M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J. Wang, 2013. Near-term Climate Change: Projections and Predictability. In: *Climate Change* 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. Chapter 11 of IPCC-AR5.

[43] Lee, R. 2011. The Outlook for Population Growth *Science*, 333, 569–573.

[44] Lynas, M. 2007. *Six Degrees: Our Future on a Hotter Planet*. Fourth Estate.

[45] Masselot, P., Mistry, M.N., Rao, S. et al. 2025. Estimating future heat-related and cold-related mortality under climate change, demographic and adaptation scenarios in 854 European cities. Nat Med. https://doi.org/10.1038/s41591-024-03452-2.

[46] Müeller, U.K., Watson, M., 2008. Testing models of low-frequency variability. *Econometrica* 76, 979-1016.

[47] Müeller, U.K., Watson, M., 2016. Measuring Uncertainty about Long-Run Predictions. Review of Economic Studies 83, 1711-1740.

[48] Newey, W.K., West, K.D., 1987. A Simple positive semi-definite, heterekodasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

[49] Parker, D.E., T.P., Folland, Legg, C.K. 1992. A new daily Central England Temperature Series, 1772-1991. Int. J. Clim., Vol 12, 317-342

[50] Pastor, L. and Stambaugh, R. F. 2012. Are Stocks Really Less Volatile in the Long Run? *Journal of Finance* LXVII, 431–477.

[51] Perron, P. and Yabu, T. 2009a. Estimating deterministic trends with an integrated or stationary noise component. *Journal of Econometrics* 151(1), 56-69.

[52] Perron, P. and Yabu, T. 2009b. Testing for Shifts in Trend With an Integrated or Stationary Noise Component. Journal of Business and Economic Statistics 27(3), 369-396.

[53] Pesavento, E. and Rossi, B. 2006., Small-Sample Confidence Intervals for Multivariate Impulse Response Functions at Long Horizons *Journal of Applied Econometrics* 21, 1135–1155.

[54] Phillips, P. C. B. 1998. Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VARs. *Journal of Econometrics* 83, 21–56.

[55] Raftery, A. E., LI, N., Sevcikova, H. 2012. Bayesian Probabilistic Population Projections for All Countries. USA, 109, 13915–13921.

[56] Rossi, B. 2005. Testing long-horizon predictive ability with high persistence, and the Meese-Rogoff puzzle. *International Economic Review* 46(1), 61-92

[57] Rossi, B. 2014. Density forecasts in economics and policymaking. *CREI.*

[58] Rossi, B. and Sekhposyan, T. 2014. Evaluating predictive densities of US output growth and inflation in a large macroeconomic data set, *International Journal of Forecasting* 30(3), 662-682.

[59] Shmueli, G. 2010. To Explain or to Predict?. *Statist. Sci.* 25(3), 289-310.

[60] Stock, J. H. 1996. VAR, Error Correction and Pretest Forecasts at Long Horizons. *Oxford Bulletin of Economics and Statistics* 58, 685–701.

[61] Stock, J. H. 1997. Cointegration, Long-Run Comovements, and Long-Horizon Forecasting, in Kreps D. and Wallis, K. *Advances in Econometrics: Proceedings of the Seventh World Congress of the Econometric Society* (Cambridge: Cambridge University Press) 34–60.

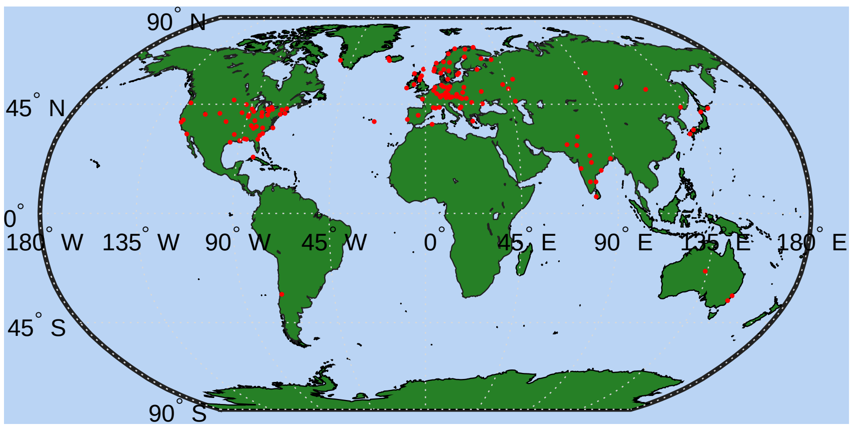
[62] Wang, X., Hyndman, R.J., Fi, F., and Kang, Y. 2022. Forecast combinations: an over 50-year review. *International Journal of Forecasting* (https://doi.org/10.1016/j.ijforecast.2022.11.005).

[63] White, H., Granger, C.W.J., 2011. Consideration of trends in time series. Journal of Time Series Econometrics 3, 1-4.

6 Tables and Figures

Figure 1
Selected stations in the Globe

(a) Panel A. 1880-2023



(b) Panel B. 1960-2023

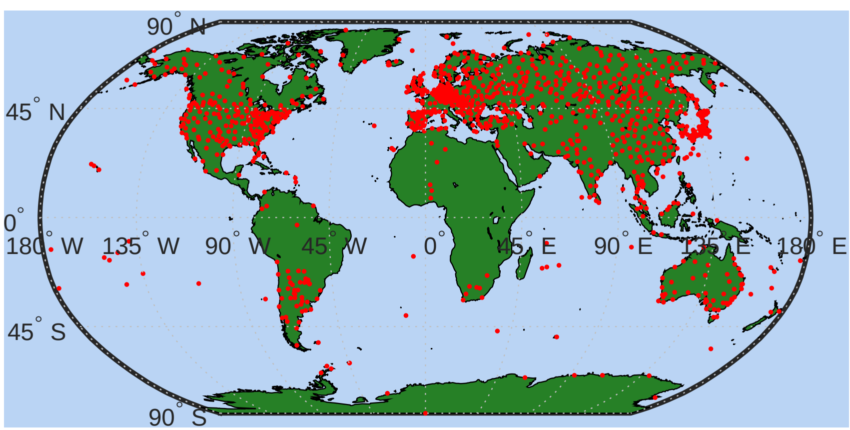
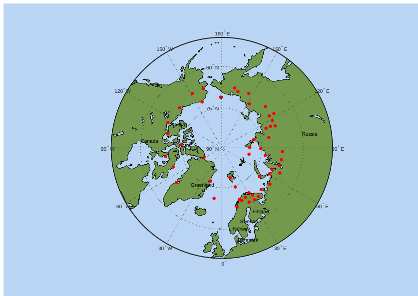
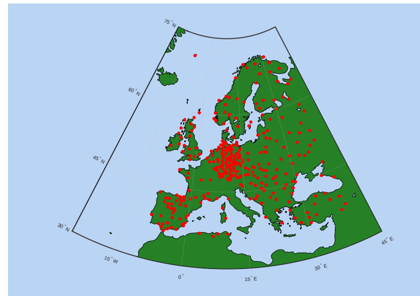


Figure 2
Selected stations by regions

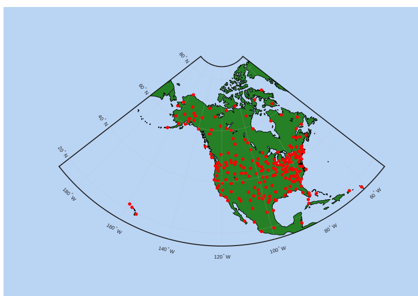
(a) Arctic Polar Circle



(b) Europe

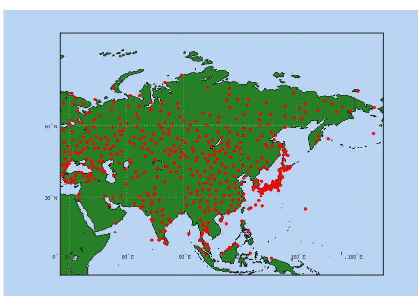


(c) North America

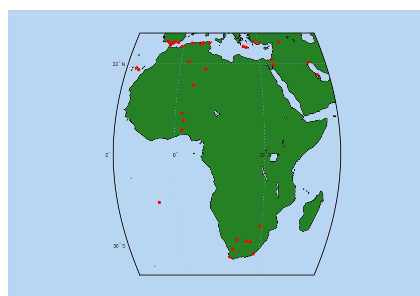


(d) South America

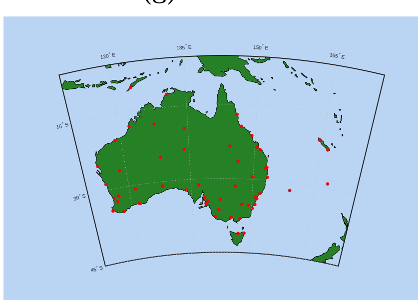
(e) Asia



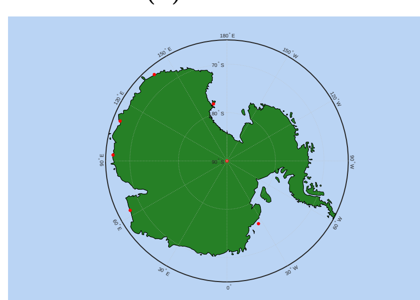
(f) Africa



(g) Australia



(h) Antarctic



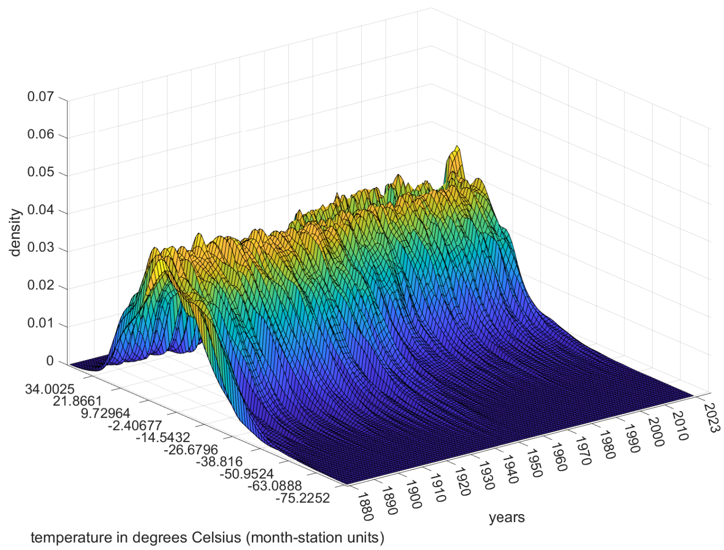


Figure 3
Globe temperature(CRU data 1880-2023)

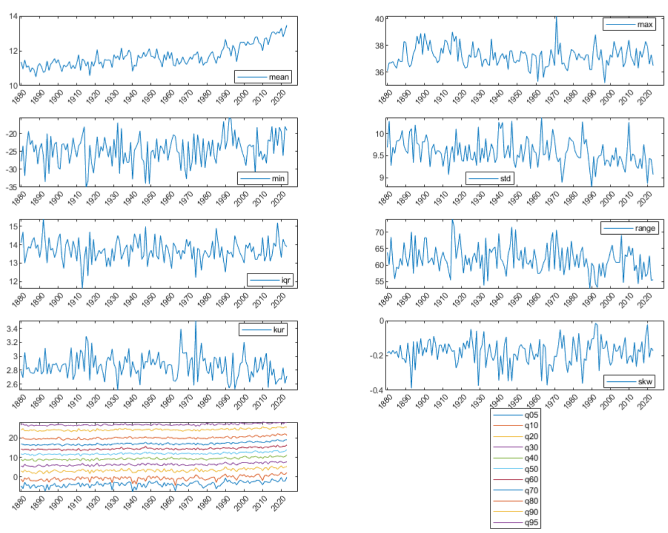


Figure 4
Distributional characteristics of the Globe temperature as time series objects (CRU data 1880-2023)

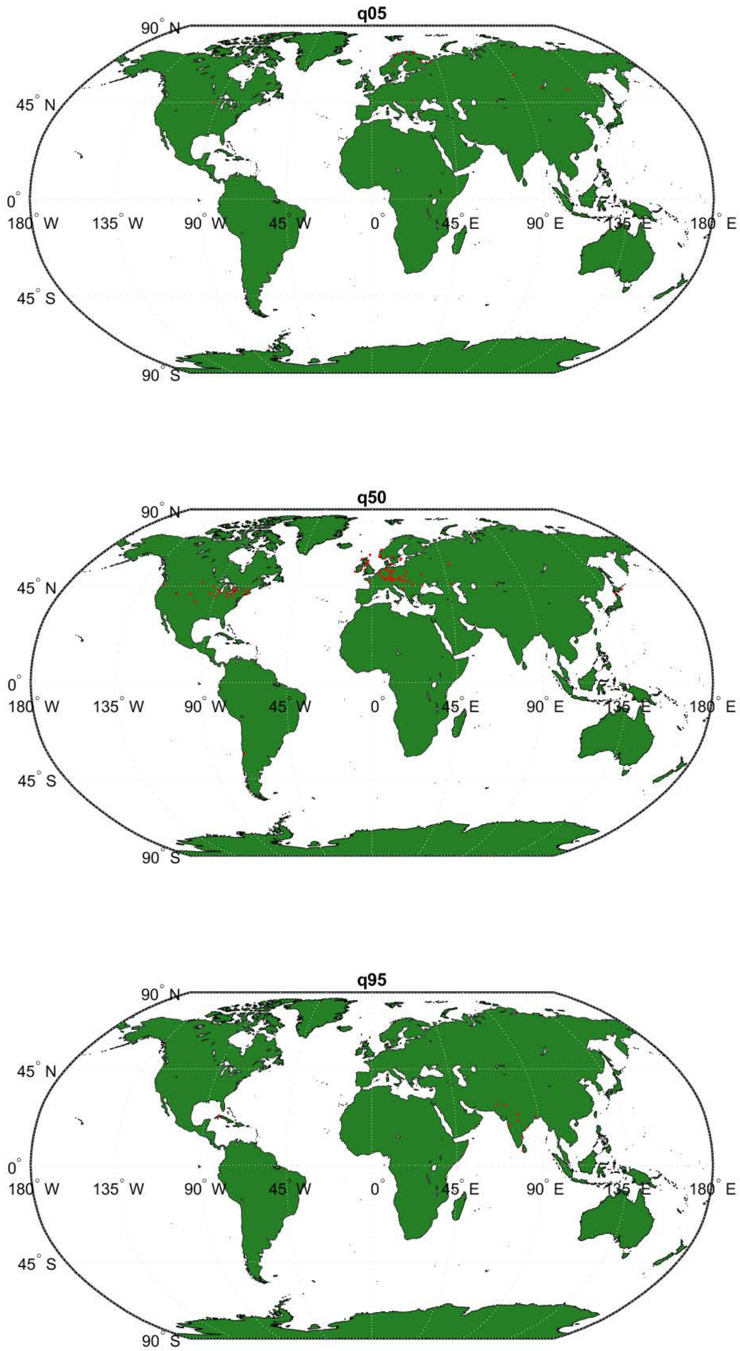


Figure 5
Geographic location of quantiles (CRU data 1880-2023)

Table 1
UR tests (Globe, 1880-2023, Globe regions,1960-2023)

Characteristic	Globe (1880-2023)	Globe (1960-2023)	Arctic Polar Circle	Europe	North-America	South-America	Asia	Africa	Australia	Antarctica
mean	-4.1870	-5.0774	-6.7657	-5.7861	-7.9723	-6.8985	-6.2625	-4.7278	-8.3806	-6.8522
	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)
max	-9.3101	-7.8904	-9.3069	-7.7026	-5.9964	-3.5653	-7.8904	-4.4724	-6.2391	-6.5458
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.043)	(0.000)	(0.004)	(0.000)	(0.000)
min	-13.1042	-6.7483	-7.9836	-6.9283	-7.7393	-10.5299	-6.3115	-5.2667	-7.5143	-6.7483
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
std	-11.2885	-7.5630	-9.1176	-6.5618	-6.9401	-7.2441	-7.6369	-7.4243	-6.7136	-7.4692
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
iqr	-12.6301	-6.9232	-3.0585	-8.0302	-6.9431	-6.1364	-7.8286	-7.6917	-7.1301	-8.6031
	(0.000)	(0.000)	(0.126)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
rank	-12.7324	-8.1645	-8.0896	-6.8916	-7.2885	-8.0290	-7.3696	-9.2664	-5.8760	-6.7454
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
kur	-12.3332	-6.7736	-7.7489	-6.8561	-7.6480	-7.8315	-6.6772	-8.8368	-7.8564	-8.6054
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
skw	-12.3451	-7.5406	-7.9753	-5.7338	-7.8162	-8.4360	-6.9558	-5.8690	-6.8361	-8.2356
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q05	-10.2598	-7.8391	-8.6953	-5.6976	-7.8601	-8.0394	-7.1901	-7.3590	-9.4343	-8.8238
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q10	-9.6121	-7.3237	-8.1298	-5.9147	-7.4385	-9.2167	-7.2456	-6.6092	-8.8654	-7.9041
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q20	-9.8748	-6.0291	-6.6535	-6.4320	-8.3694	-9.8059	-7.5230	-6.3614	-10.6432	-7.1970
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q30	-4.7739	-6.7136	-8.2277	-6.9905	-8.2217	-5.9457	-6.6584	-4.2185	-9.6534	-5.0430
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.008)	(0.000)	(0.000)
q40	-9.0937	-5.0414	-7.2383	-5.6642	-7.5628	-7.5935	-6.1152	-3.1261	-7.5061	-7.6953
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.111)	(0.000)	(0.000)
q50	-8.3186	-4.3501	-6.5651	-5.5552	-6.1780	-9.1166	-5.9936	-3.6556	-6.2759	-8.0043
	(0.000)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.034)	(0.000)	(0.000)
q60	-4.2865	-3.8642	-6.8818	-7.7783	-6.4810	-7.4314	-6.3246	-3.6184	-6.3399	-6.7380
	(0.005)	(0.020)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.037)	(0.000)	(0.000)
q70	-1.9574	-2.5087	-6.1557	-4.8997	-5.9801	-6.4516	-3.9806	-5.6874	-6.6900	-4.4908
	(0.611)	(0.343)	(0.000)	(0.001)	(0.000)	(0.000)	(0.015)	(0.000)	(0.000)	(0.004)
q80	-3.1590	-2.9303	-4.7917	-8.1616	-7.8226	-4.5003	-4.1783	-5.5063	-7.0438	-6.3762
	(0.097)	(0.161)	(0.002)	(0.000)	(0.000)	(0.004)	(0.009)	(0.000)	(0.000)	(0.000)
q90	-3.7781	-6.7083	-6.1802	-7.7920	-8.4331	-5.2557	-6.8749	-6.2127	-5.9022	-7.6965
	(0.021)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
q95	-6.3470	-7.1615	-7.4286	-7.3475	-7.9032	-4.6887	-8.1790	-6.7734	-6.3377	-5.6170
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.000)	(0.000)	(0.000)	(0.000)

Notes: Augmented Dickey-Fuller test is applied; lag-selection with BIC; pvalues in brackets.

Table 2
Trend tests (Globe, 1880-2023, Globe regions,1960-2023)

Characteristic	Globe (1880-2023)	Globe (1960-2023)	Arctic Polar Circle	Europe	North-America	South-America	Asia	Africa	Australia	Antarctica
mean	0.0117	0.0304	0.0536	0.0342	0.0280	0.0061	0.0330	0.0252	0.0191	0.0020
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.021)	(0.000)	(0.000)	(0.000)	(0.534)
max	-0.0018	0.0212	0.0234	0.0373	0.0490	0.0060	0.0212	0.0480	0.0223	-0.0056
	(0.389)	(0.000)	(0.006)	(0.000)	(0.000)	(0.494)	(0.000)	(0.000)	(0.001)	(0.256)
min	0.0194	-0.0082	0.0424	0.0725	0.0790	0.0139	0.0424	0.0154	0.0353	-0.0082
	(0.022)	(0.494)	(0.001)	(0.000)	(0.000)	(0.001)	(0.002)	(0.091)	(0.000)	(0.494)
std	-0.0024	-0.0076	-0.0137	-0.0025	-0.0072	-0.0007	-0.0094	0.0064	-0.0006	-0.0011
	(0.000)	(0.000)	(0.000)	(0.414)	(0.000)	(0.608)	(0.001)	(0.000)	(0.606)	(0.676)
iqr	-0.0005	-0.0062	-0.0355	-0.0032	-0.0032	-0.0011	-0.0124	0.0046	0.0000	0.0072
	(0.695)	(0.029)	(0.000)	(0.430)	(0.486)	(0.760)	(0.000)	(0.094)	(0.989)	(0.490)
rank	-0.0212	0.0294	-0.0190	-0.0352	-0.0301	-0.0079	-0.0212	0.0326	-0.0131	0.0026
	(0.017)	(0.002)	(0.253)	(0.030)	(0.132)	(0.431)	(0.138)	(0.000)	(0.087)	(0.852)
kur	-0.0005	0.0021	0.0040	-0.0025	-0.0042	-0.0005	0.0017	-0.0005	-0.0005	0.0004
	(0.168)	(0.098)	(0.000)	(0.104)	(0.007)	(0.321)	(0.027)	(0.556)	(0.142)	(0.674)
skw	0.0001	0.0001	-0.0021	0.0036	0.0018	0.0001	-0.0003	-0.0006	-0.0005	-0.0001
	(0.601)	(0.759)	(0.000)	(0.000)	(0.000)	(0.664)	(0.319)	(0.235)	(0.248)	(0.740)
q05	0.0182	0.0468	0.0548	0.0490	0.0451	0.0074	0.0485	0.0134	0.0150	0.0012
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)	(0.893)
q10	0.0158	0.0475	0.0713	0.0458	0.0377	0.0055	0.0508	0.0139	0.0186	0.0136
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.004)	(0.000)	(0.000)	(0.000)	(0.055)
q20	0.0131	0.0374	0.0748	0.0378	0.0311	0.0081	0.0472	0.0192	0.0170	0.0019
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.867)
q30	0.0120	0.0291	0.0787	0.0281	0.0219	0.0074	0.0368	0.0211	0.0213	0.0088
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.009)	(0.000)	(0.000)	(0.000)	(0.235)
q40	0.0110	0.0241	0.0723	0.0246	0.0211	0.0048	0.0299	0.0233	0.0226	0.0019
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.120)	(0.000)	(0.000)	(0.000)	(0.794)
q50	0.0101	0.0251	0.0585	0.0253	0.0230	0.0048	0.0281	0.0272	0.0197	-0.0010
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.145)	(0.000)	(0.000)	(0.000)	(0.868)
q60	0.0100	0.0252	0.0420	0.0256	0.0245	0.0082	0.0270	0.0275	0.0197	0.0002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.025)	(0.000)	(0.000)	(0.000)	(0.981)
q70	0.0127	0.0293	0.0379	0.0272	0.0236	0.0068	0.0278	0.0318	0.0189	0.0009
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.059)	(0.000)	(0.000)	(0.000)	(0.888)
q80	0.0118	0.0276	0.0361	0.0353	0.0232	0.0018	0.0281	0.0303	0.0174	0.0024
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.712)	(0.000)	(0.000)	(0.000)	(0.593)
q90	0.0086	0.0242	0.0294	0.0424	0.0240	0.0073	0.0218	0.0340	0.0195	0.0066
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.013)	(0.000)	(0.000)	(0.000)	(0.076)
q95	0.0078	0.0216	0.0293	0.0430	0.0257	0.0044	0.0199	0.0337	0.0177	-0.0021
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.294)	(0.000)	(0.000)	(0.000)	(0.721)

Notes: OLS estimates and HAC $t_{\beta=0}$ from regression: $C_t = \alpha + \beta t + u_t$; pvalues in brackets.

Table 3
Model description

Name	Acrimonious	Model
Mean model	mean	$C_t = \beta_0 + u_t$
Linear trend model	linear-trend	$C_t = \beta_0 + \beta_1 t + u_t$
Polynomial trend model	pol-trend	$C_t = \sum_{i=o}^{k} \beta_i t^i + u_t$
Polynomial trend model average slope (1)	pol-trend-av-sl	$C_t = \beta_0 + \tau t + u_t$
Logarithmic polynomial trend model	pol-trend-log	$C_t = \sum_{i=o}^{k} \beta_i (log(t))^i + u_t$
Structural break model	$C_t = \alpha_0 + \alpha_1 D_u + \beta_1 t + \beta_2 D_t + e_t$	$D_u = 1(t > TB)$, $D_t = 1(t > TB)(t - TB)$(2)
Polynomial trend autoregressive model	pol-trend-arp	$C_t = \sum_{i=o}^{k} \beta_i t^i + \sum_{j=1}^{p} \phi_j C_{t-j} + u_t$
Polynomial trend autoregressive model average slope (1)	pol-trend-arp-av-sl	$C_t = \beta_0 + \tau t + \sum_{j=1}^{p} \phi_j C_{t-j} + u_t$
Autoregressive model	arp	$C_t = \sum_{j=1}^{p} \phi_j C_{t-j} + u_t$
Random walk model	rw	$C_t = C_{t-1} + u_t$
Random walk drift model	rwd	$C_t = \alpha + C_{t-1} + u_t$
Local Level Model, IMA(1,1)	ima	$C_t = \alpha + C_{t-1} + (1 - \theta L) u_t$
Fractional model	arfima	$\Phi(L)(1 - L)^d C_t = \Theta(L)\epsilon_t$
Large autoregressive model	arp20	$C_t = \sum_{j=1}^{20} \phi_j C_{t-j} + u_t$

Notes: (1) The $average - slope$ models are time varying linear trend models derived form their corresponding polynomial trend models in the following way:

$$C_t = \beta_0 + \tau_T t + u_t \tag{8}$$

where

$$\tau_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial t} C_t \tag{9}$$

(2) TB is the period in which the structural break occurs and 1(.) the indicator function.

Table 4
BIC model selection for the mean (The Globe, CRU data, 1880-2023)

Models	BIC	weights
mean	-0.9753	0.0420
linear-trend	-1.8821	0.0661
pol-trend (k=5)	-2.2704	0.0803
pol-trend-av-sl (k=5)	-2.2704	0.0803
pol-trend-log (k=8)	-2.1682	0.0763
struct-breaks	-2.1853	0.0769
pol-trend-arp (k=3, p=1)	-2.2653	0.0801
pol-trend-arp-av-sl (k=3, p=1)	-2.2653	0.0801
arp (p=2)	-2.0724	0.0727
rw	-1.9017	0.0668
rwd	-1.8741	0.0658
ima	-2.2360	0.0789
arfima	-2.0991	0.0737
arp20	-1.6917	0.0601

Notes: The weights used to combine forecast in the rest of the paper are calculated as follows: $w_j = \frac{e^{-1/2BIC_j}}{\sum_{j=1}^{m} e^{-1/2BIC_j}}$ where BIC_j is the BIC criterium of each model from j=1...m in sample.

Table 5
Mean temperature forecasts (The Globe, CRU data, 1880-2023)

Horizon	Benchmark model: linear trend	Selected model: pol-trend	Selected model: pol-trend-av-sl
1	12.53	13.24	13.26
	(11.91,13.16)	(12.75,13.73)	(12.77,13.75)
10	12.64	13.42	13.38
	(12.01,13.26)	(12.94,13.91)	(12.90,13.87)
25	12.81	12.84	13.59
	(12.19,13.44)	(12.36,13.33)	(13.10,14.08)
50	13.10	6.22	13.93
	(12.48,13.73)	(5.73,6.71)	(13.45,14.42)
100	13.69	-66.70	14.62
	(13.06,14.31)	(-67.19,-66.21)	(14.14,15.11)

Notes: The average value of the *mean* in the full sample is 11.69; the average value of the Mean in 1880-1900 (the baseline model for the Paris Agreement) is 11.12; the average value of the Mean in the AR5 reference period (1986-2005) is 12.17; in the AR6 period (1995-2014) is 12.39; Standard 95% confidence intervals.

Table 6
Long-term forecasts for the mean (The Globe, CRU data, 1880-2023)

Models/horizon	h=1	h=25	h=50	h=100
mean	11.69	11.69	11.69	11.69
	(10.67,12.70)	(10.67,12.70)	(10.67,12.70)	(10.67,12.70)
linear-trend	12.53	12.81	13.10	13.69
	(11.91,13.16)	(12.19,13.44)	(12.48,13.73)	(13.06,14.31)
pol-trend	13.24	12.84	6.22	-66.70
	(12.75,13.73)	(12.36,13.33)	(5.73,6.71)	(-67.19,-66.21)
pol-trend-av-sl	13.26	13.59	13.93	14.62
	(12.77,13.75)	(13.10,14.08)	(13.45,14.42)	(14.14,15.11)
pol-trend-log	13.48	15.96	20.03	33.43
	(12.03,14.93)	(14.51,17.41)	(18.58,21.48)	(31.98,34.88)
struct-breaks	13.50	14.03	14.58	15.67
	(13.02,13.98)	(13.55,14.51)	(14.10,15.05)	(15.20,16.15)
pol-trend-arp	13.38	15.80	17.63	70.49
	(12.88,13.88)	(15.29,16.30)	(17.13,18.12)	(70.03,70.96)
pol-trend-arp-av-sl	13.36	13.84	13.27	51.37
	(12.86,13.86)	(13.33,14.34)	(12.78,13.77)	(50.90,51.83)
arp	13.20	13.36	13.26	14.70
	(12.60,13.80)	(12.48,14.23)	(12.35,14.17)	(13.91,15.49)
rw	13.47	13.47	13.47	13.47
	(12.83,14.10)	(12.53,14.40)	(12.37,14.57)	(12.29,14.65)
rwd	13.48	13.84	14.20	14.94
	(12.85,14.12)	(13.05,14.63)	(13.36,15.04)	(14.38,15.50)
ima	13.20	13.56	13.92	14.66
	(12.42,13.99)	(12.68,14.43)	(13.10,14.75)	(14.17,15.15)
arfima	13.07	12.32	12.16	12.03
	(6.39,19.74)	(5.64,18.99)	(5.48,18.84)	(5.36,18.71)
arp20	13.20	13.61	13.92	19.72
	(12.67,13.73)	(12.71,14.51)	(13.33,14.52)	(19.56,19.89)
combined0	13.20	13.71	13.73	16.91
	(12.47,13.93)	(12.85,14.56)	(12.85,14.60)	(16.12,17.70)
combined1	13.24	13.61	13.87	19.79
	(12.53,13.96)	(12.76,14.46)	(12.95,14.79)	(18.97,20.62)
combined2	13.29	13.50	13.58	16.74
	(12.59,13.99)	(12.62,14.38)	(12.66,14.51)	(15.92,17.56)
combined3	13.29	13.50	13.63	15.11
	(12.59,13.99)	(12.59,14.42)	(12.64,14.62)	(14.24,15.98)

Notes: The average value of the *mean* in the full sample is 11.69; the average value of the *mean* in 1880-1900 (the baseline model for the Paris Agreement) is 11.12; the average value of the *mean* in the AR5 reference period (1986-2005) is 12.17; in the AR6 period (1995-2014) is 12.39. “Combination0” weights the models using the BICs obtained in-sample; “combined1” removes two extreme values ; “combined2” removes four extreme values and “combined3” six extreme values. In these last cases BIC-weights are properly recalculated. Confidence intervals for combined models are built with estimated errors following the procedure described in the text.

Table 7
Long-term forecasts for the *mean* with "Pareto" superior models (The Globe, CRU data, 1880-2023, w=100)

Models/horizon	h=1	h=10	h=25
mean	-	-	-
	(-,-)	(-,-)	(-,-)
linear-trend	-	-	12.81
	(-,-)	(-,-)	(12.19,13.44)
pol-trend	-	-	-
	(-,-)	(-,-)	(-,-)
pol-trend-av-sl	13.26	13.38	13.59
	(12.77,13.75)	(12.90,13.87)	(13.10,14.08)
pol-trend-log	13.48	14.26	-
	(12.03,14.93)	(12.81,15.71)	(-,-)
struct-breaks	-	-	-
	(-,-)	(-,-)	(-,-)
pol-trend-arp	-	-	-
	(-,-)	(-,-)	(-,-)
pol-trend-arp-av-sl	-	-	-
	(-,-)	(-,-)	(-,-)
arp	-	-	-
	(-,-)	(-,-)	(-,-)
rw	13.47	-	-
	(12.83,14.10)	(-,-)	(-,-)
rwd	13.48	-	13.84
	(12.85,14.12)	(-,-)	(13.05,14.63)
ima	13.20	13.34	13.56
	(12.42,13.99)	(12.48,14.19)	(12.68,14.43)
arfima	-	12.54	-
	(-,-)	(11.75,13.34)	(-,-)
arp20	-	-	-
	(-,-)	(-,-)	(-,-)
combined	13.37	13.39	13.46
	(12.64,14.10)	(12.68,14.10)	(12.57,14.35)

Notes: combined uses BIC weights.
The confidence intervals are computed with forecast errors following the procedure described in the text.
The average value of the *mean* in the full sample is 11.69; the average value of the *mean* in 1880-1900 (the baseline model for the Paris Agreement) is 11.12; the average value for the AR5 reference period (1986-2005) is 12.17; the average value for the AR6 reference period (1995-2014) is 12.39.

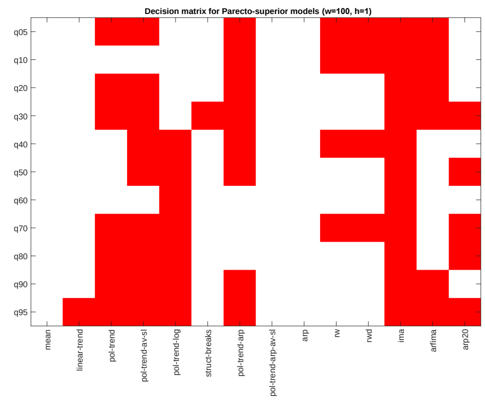
Table 8
Long-term density forecasts with Pareto superior model for all quantiles (the Globe, CRU data, 1880-2023, w=100)

	h=1	h=10	h=25
q05	-1.70	-1.10	-2.06
	(-4.19,0.80)	(-2.91,0.71)	(-3.95,-0.17)
q10	1.37	1.94	0.83
	(-0.67,3.41)	(0.45,3.43)	(-0.78,2.44)
q20	4.87	5.34	4.58
	(3.49,6.25)	(4.35,6.32)	(3.50,5.65)
q30	7.59	8.00	7.50
	(6.63,8.55)	(7.32,8.67)	(6.76,8.24)
q40	10.63	11.11	10.38
	(9.78,11.48)	(10.48,11.73)	(9.68,11.08)
q50	13.28	13.51	12.98
	(12.53,14.03)	(12.97,14.06)	(12.34,13.63)
q60	15.87	16.19	15.47
	(15.09,16.64)	(15.67,16.72)	(14.80,16.13)
q70	18.83	19.15	18.31
	(18.06,19.61)	(18.59,19.70)	(17.58,19.03)
q80	21.73	22.12	21.27
	(20.95,22.52)	(21.55,22.69)	(20.52,22.01)
q90	25.50	25.67	25.17
	(24.66,26.33)	(25.09,26.26)	(24.47,25.86)
q95	27.83	27.90	27.72
	(27.15,28.51)	(27.41,28.39)	(27.15,28.29)

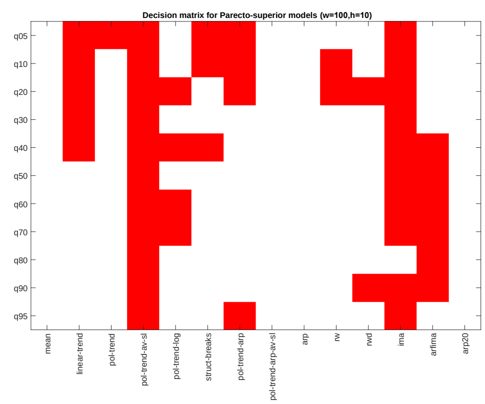
Notes: The method selects the Pareto-superior models for all the quantiles and combines them according with the BIC weights. The selected model are: *ima*, *pol-trend-av-sl* and *linear-trend* for h=1,10, 25, respectively.

Figure 6
Selected Pareto-superior models

(a) h=1



(b) h=10



(c) h=25

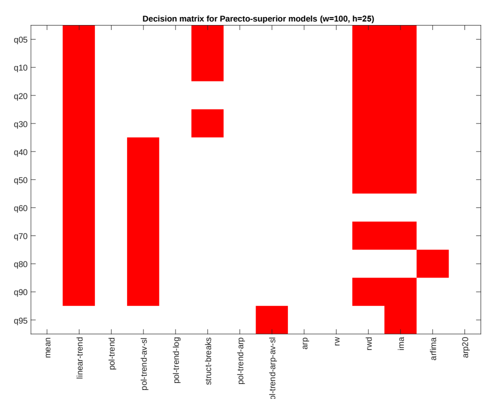


Table 9

Temperature quantile increases by reference periods and horizons, w=100 (CRU data, Globe 1880-2023)

quantile	periods	h=1, 2024	h=10, 2033	h=25, 2048	h=77, 2100
q05	1880-1900	3.19	3.79	2.82	3.77
		(-0.22,6.59)	(1.99,5.59)	(0.94,4.71)	(1.88,5.66)
	1986-2005	1.20	1.80	0.83	1.78
		(-2.21,4.60)	(-0.00,3.60)	(-1.05,2.72)	(-0.11,3.67)
	1995-2014	1.03	1.63	0.67	1.61
		(-2.37,4.44)	(-0.17,3.44)	(-1.22,2.55)	(-0.28,3.50)
q10	1880-1900	2.92	3.49	2.38	3.20
		(0.46,5.39)	(2.01,4.98)	(0.78,3.99)	(1.59,4.81)
	1986-2005	1.29	1.86	0.75	1.57
		(-1.17,3.76)	(0.38,3.35)	(-0.85,2.36)	(-0.04,3.18)
	1995-2014	1.18	1.75	0.64	1.46
		(-1.29,3.64)	(0.26,3.23)	(-0.97,2.24)	(-0.15,3.07)
q20	1880-1900	2.30	2.76	2.00	2.68
		(0.78,3.81)	(1.78,3.74)	(0.93,3.07)	(1.61,3.76)
	1986-2005	1.02	1.49	0.73	1.41
		(-0.49,2.54)	(0.51,2.47)	(-0.35,1.80)	(0.33,2.48)
	1995-2014	0.80	1.26	0.50	1.18
		(-0.72,2.31)	(0.28,2.24)	(-0.57,1.57)	(0.11,2.25)
q30	1880-1900	1.90	2.31	1.82	2.44
		(0.93,2.88)	(1.64,2.99)	(1.08,2.55)	(1.70,3.18)
	1986-2005	0.72	1.13	0.63	1.26
		(-0.25,1.70)	(0.46,1.80)	(-0.10,1.37)	(0.52,2.00)
	1995-2014	0.51	0.92	0.42	1.05
		(-0.47,1.48)	(0.24,1.59)	(-0.32,1.16)	(0.30,1.79)
q40	1880-1900	1.86	2.33	1.61	2.18
		(0.58,3.14)	(1.72,2.95)	(0.92,2.30)	(1.48,2.88)
	1986-2005	0.87	1.35	0.63	1.20
		(-0.41,2.16)	(0.73,1.97)	(-0.07,1.32)	(0.50,1.89)
	1995-2014	0.59	1.06	0.34	0.91
		(-0.70,1.87)	(0.44,1.68)	(-0.36,1.03)	(0.21,1.61)
q50	1880-1900	1.73	1.96	1.43	1.96
		(0.45,3.00)	(1.42,2.50)	(0.79,2.07)	(1.31,2.60)
	1986-2005	0.91	1.14	0.61	1.14
		(-0.37,2.18)	(0.60,1.68)	(-0.03,1.25)	(0.50,1.78)
	1995-2014	0.62	0.85	0.32	0.85
		(-0.66,1.89)	(0.31,1.39)	(-0.32,0.96)	(0.20,1.49)
q60	1880-1900	1.76	2.09	1.36	1.88
		(0.65,2.88)	(1.57,2.61)	(0.70,2.02)	(1.22,2.55)
	1986-2005	0.99	1.32	0.59	1.11
		(-0.12,2.11)	(0.80,1.84)	(-0.07,1.25)	(0.45,1.78)
	1995-2014	0.79	1.11	0.38	0.90
		(-0.33,1.90)	(0.59,1.63)	(-0.28,1.05)	(0.24,1.57)
q70	1880-1900	2.29	2.61	1.77	2.42
		(1.39,3.19)	(2.06,3.16)	(1.04,2.49)	(1.70,3.15)
	1986-2005	1.27	1.58	0.74	1.40
		(0.37,2.17)	(1.03,2.13)	(0.02,1.46)	(0.68,2.13)
	1995-2014	0.99	1.31	0.47	1.13
		(0.09,1.89)	(0.76,1.86)	(-0.26,1.19)	(0.40,1.85)
q80	1880-1900	2.10	2.49	1.64	2.25
		(1.20,3.00)	(1.92,3.06)	(0.90,2.38)	(1.51,2.99)
	1986-2005	1.23	1.62	0.76	1.38
		(0.33,2.13)	(1.05,2.19)	(0.02,1.50)	(0.63,2.12)
	1995-2014	0.92	1.31	0.45	1.07
		(0.02,1.82)	(0.74,1.88)	(-0.29,1.19)	(0.32,1.81)
q90	1880-1900	1.50	1.68	1.17	1.62
		(0.66,2.34)	(1.10,2.26)	(0.48,1.87)	(0.92,2.31)
	1986-2005	0.83	1.00	0.50	0.94
		(-0.01,1.66)	(0.42,1.59)	(-0.20,1.19)	(0.25,1.64)
	1995-2014	0.60	0.78	0.27	0.71
		(-0.24,1.44)	(0.20,1.36)	(-0.42,0.96)	(0.02,1.41)
q95	1880-1900	1.20	1.27	1.09	1.50
		(0.52,1.89)	(0.78,1.76)	(0.53,1.66)	(0.93,2.06)
	1986-2005	0.54	0.61	0.43	0.84
		(-0.14,1.23)	(0.12,1.10)	(-0.13,0.99)	(0.27,1.40)
	1995-2014	0.32	0.39	0.21	0.61
		(-0.37,1.00)	(-0.10,0.87)	(-0.36,0.77)	(0.05,1.18)

Notes: The predictions, and consequently the estimated temperature increases over the reference periods, were derived by combining the Pareto-superior models for each prediction horizon. For $h=77$, the model selection corresponding to $h=25$ was applied; Confidence intervals for combined models are built with estimated errors following the procedure described in the text.

Figure 7
Temperature increases in 2100 by reference periods

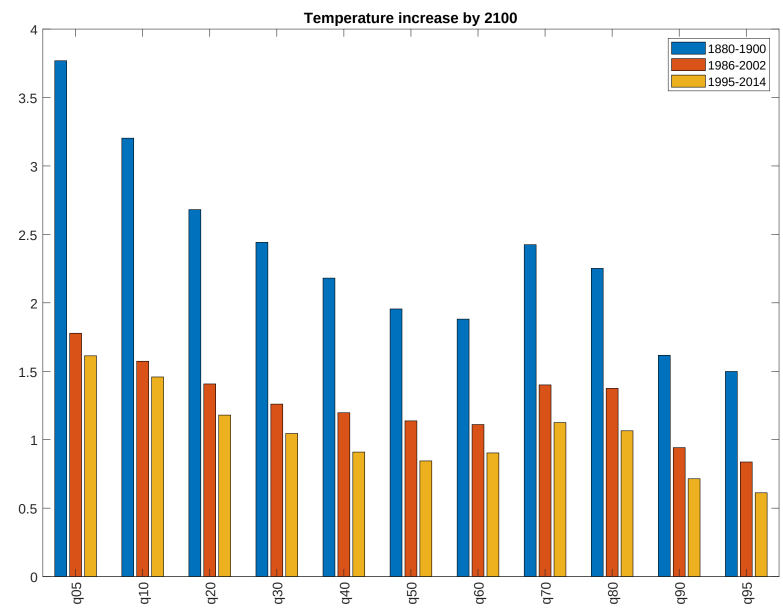
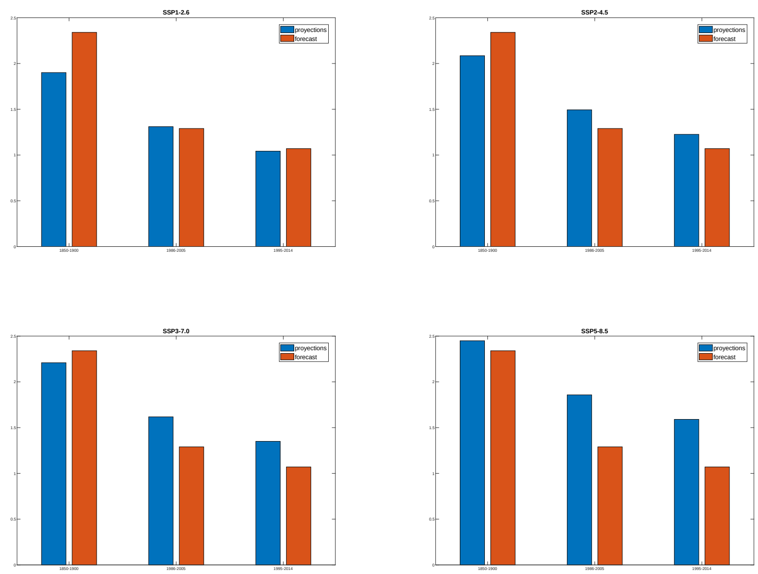


Figure 8

Comparing predictions and projections for the mean temperature

(a) h=25, 2048



(b) h=77, 2100

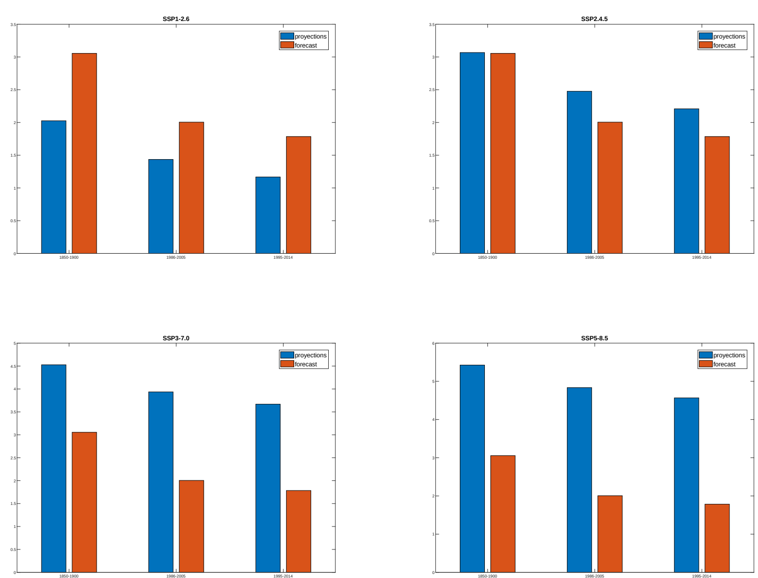


Table 10
Long-term density forecasts with Pareto superior model for all quantiles (the Globe, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-14.02	-13.60	-12.90
	(-15.66,-12.38)	(-15.24,-11.96)	(-14.54,-11.25)
q10	-5.97	-5.54	-4.83
	(-7.15,-4.79)	(-6.72,-4.37)	(-6.01,-3.65)
q20	1.31	1.65	2.21
	(0.30,2.32)	(0.63,2.66)	(1.20,3.22)
q30	5.44	5.70	6.13
	(4.78,6.09)	(5.04,6.35)	(5.48,6.79)
q40	8.84	9.05	9.42
	(8.36,9.32)	(8.57,9.53)	(8.93,9.90)
q50	12.12	12.35	12.72
	(11.68,12.56)	(11.90,12.79)	(12.28,13.16)
q60	15.05	15.27	15.65
	(14.65,15.45)	(14.87,15.67)	(15.25,16.05)
q70	18.02	18.28	18.72
	(17.59,18.45)	(17.85,18.71)	(18.29,19.15)
q80	21.26	21.51	21.93
	(20.82,21.71)	(21.07,21.96)	(21.48,22.37)
q90	25.55	25.77	26.13
	(25.15,25.94)	(25.37,26.16)	(25.73,26.52)
q95	27.77	27.97	28.29
	(27.46,28.08)	(27.66,28.27)	(27.98,28.60)

Notes: The selected models are *linear-trend* for all horizons.

Table 11
Long-term density forecasts with Pareto superior model for all quantiles (the Arctic, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-31.94	-31.45	-30.63
	(-34.22,-29.66)	(-33.73,-29.17)	(-32.91,-28.35)
q10	-27.05	-26.41	-25.34
	(-28.96,-25.14)	(-28.31,-24.50)	(-27.25,-23.43)
q20	-20.15	-19.48	-18.36
	(-21.94,-18.36)	(-21.27,-17.69)	(-20.15,-16.57)
q30	-13.21	-12.50	-11.32
	(-15.26,-11.16)	(-14.55,-10.46)	(-13.37,-9.27)
q40	-7.80	-7.14	-6.06
	(-9.67,-5.92)	(-9.02,-5.27)	(-7.93,-4.19)
q50	-3.40	-2.87	-2.00
	(-4.74,-2.06)	(-4.21,-1.53)	(-3.34,-0.66)
q60	0.20	0.58	1.21
	(-0.81,1.21)	(-0.44,1.59)	(0.19,2.22)
q70	3.68	4.02	4.59
	(2.73,4.63)	(3.07,4.98)	(3.64,5.54)
q80	7.30	7.63	8.17
	(6.18,8.42)	(6.51,8.75)	(7.05,9.29)
q90	11.64	11.91	12.35
	(10.76,12.53)	(11.02,12.79)	(11.46,13.23)
q95	13.82	14.08	14.52
	(12.79,14.84)	(13.05,15.10)	(13.49,15.54)

Notes: The Pareto-superior model is *linear-trend* for all the horizons.

Table 12
Long-term density forecasts with Pareto superior models for all quantiles (Europe, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-2.19	-1.75	-1.02
	(-4.55,0.16)	(-4.11,0.60)	(-3.38,1.34)
q10	0.50	0.91	1.60
	(-1.58,2.58)	(-1.16,2.99)	(-0.48,3.68)
q20	3.58	3.92	4.48
	(2.07,5.08)	(2.41,5.42)	(2.98,5.99)
q30	5.78	6.03	6.45
	(4.86,6.70)	(5.11,6.95)	(5.53,7.37)
q40	8.07	8.29	8.66
	(7.29,8.85)	(7.51,9.07)	(7.88,9.44)
q50	10.57	10.80	11.18
	(9.67,11.47)	(9.90,11.70)	(10.28,12.08)
q60	12.97	13.20	13.58
	(12.18,13.75)	(12.41,13.98)	(12.80,14.37)
q70	15.25	15.50	15.90
	(14.46,16.04)	(14.71,16.28)	(15.12,16.69)
q80	17.77	18.08	18.61
	(16.90,18.63)	(17.22,18.95)	(17.75,19.48)
q90	20.63	21.01	21.64
	(19.67,21.58)	(20.05,21.96)	(20.69,22.60)
q95	23.03	23.42	24.07
	(22.13,23.94)	(22.52,24.33)	(23.16,24.97)

Notes: The selected model is *linear-trend* for all the horizons.

Table 13
Long-term density forecasts with Pareto superior model for all quantiles (North America, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-10.78	-10.35	-9.67
	(-12.98,-8.59)	(-12.17,-8.53)	(-11.50,-7.85)
q10	-4.26	-3.87	-3.30
	(-6.58,-1.95)	(-5.76,-1.97)	(-5.20,-1.40)
q20	1.59	2.08	2.54
	(0.05,3.12)	(0.83,3.32)	(1.30,3.79)
q30	5.92	6.17	6.50
	(4.78,7.06)	(5.24,7.10)	(5.57,7.43)
q40	9.35	9.58	9.90
	(8.48,10.23)	(8.87,10.29)	(9.19,10.61)
q50	12.71	12.86	13.20
	(11.92,13.49)	(12.20,13.51)	(12.55,13.86)
q60	15.92	16.07	16.44
	(15.20,16.64)	(15.46,16.68)	(15.83,17.05)
q70	19.21	19.30	19.65
	(18.46,19.96)	(18.63,19.97)	(18.99,20.32)
q80	22.38	22.51	22.85
	(21.51,23.24)	(21.78,23.24)	(22.12,23.58)
q90	25.47	25.73	26.09
	(24.72,26.21)	(25.12,26.33)	(25.48,26.69)
q95	27.56	27.83	28.22
	(26.84,28.27)	(27.25,28.42)	(27.64,28.80)

Notes: The selected model are {*linear-trend*,*pol-trend-log*}; *linear-trend*; *linear-trend* for h=1, 10 and 25, respectively.

Table 14
Long-term density forecasts with Pareto superior model for all quantiles (South America, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	7.17	7.25	7.47
	(6.24,8.10)	(6.28,8.23)	(6.72,8.22)
q10	8.73	8.80	8.96
	(7.80,9.66)	(7.82,9.77)	(8.21,9.71)
q20	11.24	11.33	11.57
	(10.32,12.15)	(10.37,12.29)	(10.83,12.30)
q30	13.38	13.47	13.68
	(12.69,14.08)	(12.74,14.19)	(13.13,14.23)
q40	15.62	15.68	15.82
	(14.65,16.60)	(14.66,16.70)	(15.03,16.61)
q50	17.85	18.00	18.04
	(16.89,18.80)	(17.02,18.99)	(17.27,18.82)
q60	19.95	20.17	20.28
	(19.01,20.88)	(19.22,21.11)	(19.53,21.02)
q70	21.98	22.17	22.26
	(21.12,22.84)	(21.30,23.05)	(21.57,22.95)
q80	24.06	24.28	24.14
	(23.19,24.93)	(23.43,25.13)	(23.43,24.85)
q90	26.36	26.52	26.65
	(25.73,26.99)	(25.87,27.16)	(26.16,27.15)
q95	27.39	27.63	27.57
	(26.60,28.18)	(26.87,28.39)	(26.93,28.21)

Notes: The selected models are: {*mean*, *linear-trend*}, {*mean*, *linear-trend*, *pol-trend-av-sl*} and *linear-trend* for h=1,10,25, respectively.

Table 15

Long-term density forecasts with Pareto superior model for all quantiles (Asia, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-20.08	-19.62	-18.90
	(-24.07,-16.09)	(-25.56,-13.68)	(-21.53,-16.26)
q10	-12.42	-11.91	-11.15
	(-15.56,-9.27)	(-16.35,-7.46)	(-13.19,-9.10)
q20	-3.59	-3.24	-2.54
	(-5.68,-1.50)	(-5.32,-1.17)	(-3.87,-1.21)
q30	2.24	2.64	3.03
	(0.29,4.19)	(1.21,4.07)	(2.21,3.85)
q40	6.86	7.00	7.47
	(5.63,8.09)	(2.45,11.56)	(6.82,8.11)
q50	10.94	11.07	11.49
	(10.00,11.87)	(9.45,12.70)	(10.90,12.08)
q60	14.73	14.89	15.29
	(14.08,15.39)	(13.99,15.79)	(14.86,15.72)
q70	18.15	18.28	18.72
	(17.39,18.90)	(17.02,19.54)	(18.21,19.22)
q80	21.99	22.09	22.56
	(21.14,22.85)	(19.39,24.78)	(22.01,23.11)
q90	26.78	26.89	27.22
	(26.12,27.43)	(26.09,27.68)	(26.81,27.63)
q95	28.47	28.68	28.91
	(27.97,28.97)	(28.17,29.19)	(28.59,29.24)

Notes: The selected model are: {*linear-trend*, *pol-trend-log*}, {*linear-trend*, *pol-trend-av-sl*} and *linear-trend* for h=1,10, 25, respectively.

Table 16

Long-term density forecasts with Pareto superior model for all quantiles (Africa, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	10.92	11.04	11.24
	(9.88,11.96)	(10.00,12.08)	(10.21,12.28)
q10	12.20	12.33	12.54
	(11.44,12.97)	(11.57,13.09)	(11.78,13.30)
q20	14.10	14.27	14.56
	(13.41,14.78)	(13.58,14.95)	(13.87,15.24)
q30	15.95	16.14	16.46
	(15.28,16.63)	(15.47,16.82)	(15.79,17.13)
q40	17.71	17.92	18.27
	(17.14,18.28)	(17.35,18.49)	(17.69,18.84)
q50	19.71	19.95	20.36
	(19.10,20.32)	(19.35,20.56)	(19.76,20.97)
q60	21.56	21.80	22.22
	(20.96,22.15)	(21.21,22.40)	(21.62,22.81)
q70	23.89	24.18	24.66
	(23.27,24.52)	(23.56,24.81)	(24.03,25.28)
q80	26.07	26.34	26.79
	(25.43,26.71)	(25.70,26.98)	(26.16,27.43)
q90	28.56	28.87	29.38
	(27.81,29.32)	(28.11,29.63)	(28.62,30.14)
q95	31.75	32.05	32.56
	(30.86,32.64)	(31.16,32.95)	(31.66,33.45)

Notes: The selected model is *linear-trend* for all horizons.

Table 17

Long-term density forecasts with Pareto superior model for all quantiles (Australia, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	10.92	11.04	11.24
	(9.88,11.96)	(10.00,12.08)	(10.21,12.28)
q10	12.20	12.33	12.54
	(11.44,12.97)	(11.57,13.09)	(11.78,13.30)
q20	14.10	14.27	14.56
	(13.41,14.78)	(13.58,14.95)	(13.87,15.24)
q30	15.95	16.14	16.46
	(15.28,16.63)	(15.47,16.82)	(15.79,17.13)
q40	17.71	17.92	18.27
	(17.14,18.28)	(17.35,18.49)	(17.69,18.84)
q50	19.71	19.95	20.36
	(19.10,20.32)	(19.35,20.56)	(19.76,20.97)
q60	21.56	21.80	22.22
	(20.96,22.15)	(21.21,22.40)	(21.62,22.81)
q70	23.89	24.18	24.66
	(23.27,24.52)	(23.56,24.81)	(24.03,25.28)
q80	26.07	26.34	26.79
	(25.43,26.71)	(25.70,26.98)	(26.16,27.43)
q90	28.56	28.87	29.38
	(27.81,29.32)	(28.11,29.63)	(28.62,30.14)
q95	31.75	32.05	32.56
	(30.86,32.64)	(31.16,32.95)	(31.66,33.45)

Notes: The selected model is *linear-trend* for all horizons.

Table 18

Long-term density forecasts with Pareto superior model for all quantiles (the Antarctic, CRU data, 1960-2023, w=25)

	h=1	h=10	h=25
q05	-58.52	-58.53	-58.52
	(-60.62,-56.43)	(-61.07,-55.99)	(-60.62,-56.43)
q10	-53.82	-53.82	-53.82
	(-56.12,-51.51)	(-56.65,-51.00)	(-56.12,-51.51)
q20	-26.59	-26.59	-26.59
	(-29.15,-24.03)	(-29.72,-23.46)	(-29.15,-24.03)
q30	-19.00	-19.00	-19.00
	(-20.87,-17.13)	(-21.30,-16.71)	(-20.87,-17.13)
q40	-16.47	-16.48	-16.47
	(-18.17,-14.78)	(-18.55,-14.40)	(-18.17,-14.78)
q50	-14.51	-14.51	-14.51
	(-16.00,-13.01)	(-16.34,-12.68)	(-16.00,-13.01)
q60	-12.36	-12.36	-12.36
	(-13.81,-10.91)	(-14.13,-10.59)	(-13.81,-10.91)
q70	-9.50	-9.50	-9.50
	(-10.85,-8.16)	(-11.15,-7.86)	(-10.85,-8.16)
q80	-5.59	-5.59	-5.59
	(-6.66,-4.52)	(-6.89,-4.29)	(-6.66,-4.52)
q90	-2.19	-2.19	-2.19
	(-3.12,-1.27)	(-3.33,-1.06)	(-3.12,-1.27)
q95	-0.79	-0.81	-0.79
	(-1.85,0.27)	(-2.07,0.44)	(-1.85,0.27)

Notes: The selected models are: *mean*, {*mean*, *arp*}, *mean*, for h=1,10,25, respectively.

Table 19
Descriptive values of quantiles by regions (CRU data 1960-2023)

quantile	statistics	Globe	Arctic	Europe	NorthAmerica	SouthAmerica	Asia	Africa	Australia	Antarctic
q05	mean 1986-2005	-15.15	-33.76	-3.64	-11.71	7.06	-21.12	10.35	10.21	-58.85
	mean 1995-2014	-15.07	-33.13	-3.48	-11.52	7.08	-21.35	10.69	10.33	-58.57
q10	mean 1986-2005	-7.22	-29.63	-0.89	-4.91	8.68	-13.66	11.68	11.54	-53.98
	mean 1995-2014	-6.97	-28.64	-0.73	-4.91	8.70	-13.58	11.93	11.68	-53.76
q20	mean 1986-2005	0.25	-22.86	2.44	1.07	11.17	-5.06	13.49	13.30	-26.89
	mean 1995-2014	0.42	-21.81	2.60	1.12	11.18	-4.66	13.73	13.52	-26.83
q30	mean 1986-2005	4.57	-15.93	4.86	5.45	13.23	1.10	15.31	15.35	-18.99
	mean 1995-2014	4.83	-14.97	5.17	5.47	13.25	1.52	15.56	15.62	-18.69
q40	mean 1986-2005	8.13	-10.32	7.32	8.81	15.49	5.93	16.98	17.25	-16.51
	mean 1995-2014	8.39	-9.44	7.64	8.88	15.55	6.28	17.23	17.56	-16.39
q50	mean 1986-2005	11.32	-5.28	9.73	11.94	17.65	10.03	18.93	19.21	-14.63
	mean 1995-2014	11.63	-4.59	10.13	12.04	17.77	10.42	19.20	19.49	-14.52
q60	mean 1986-2005	14.24	-1.21	12.14	15.02	19.68	13.91	20.80	20.93	-12.24
	mean 1995-2014	14.49	-0.65	12.51	15.19	19.83	14.22	21.03	21.20	-12.47
q70	mean 1986-2005	17.09	2.36	14.36	18.22	21.75	17.13	23.00	22.68	-9.40
	mean 1995-2014	17.42	2.83	14.68	18.45	21.87	17.58	23.30	22.91	-9.59
q80	mean 1986-2005	20.37	6.01	16.64	21.47	23.86	20.99	25.21	24.84	-5.70
	mean 1995-2014	20.70	6.56	17.01	21.69	24.00	21.43	25.51	25.04	-5.62
q90	mean 1986-2005	24.78	10.69	19.35	24.71	26.19	26.06	27.51	27.20	-2.20
	mean 1995-2014	25.06	10.94	19.77	24.94	26.36	26.37	27.94	27.33	-2.14
q95	mean 1986-2005	27.13	12.94	21.75	26.80	27.11	27.89	30.53	28.66	-0.82
	mean 1995-2014	27.34	13.23	22.17	27.05	27.34	28.09	31.15	28.73	-0.99

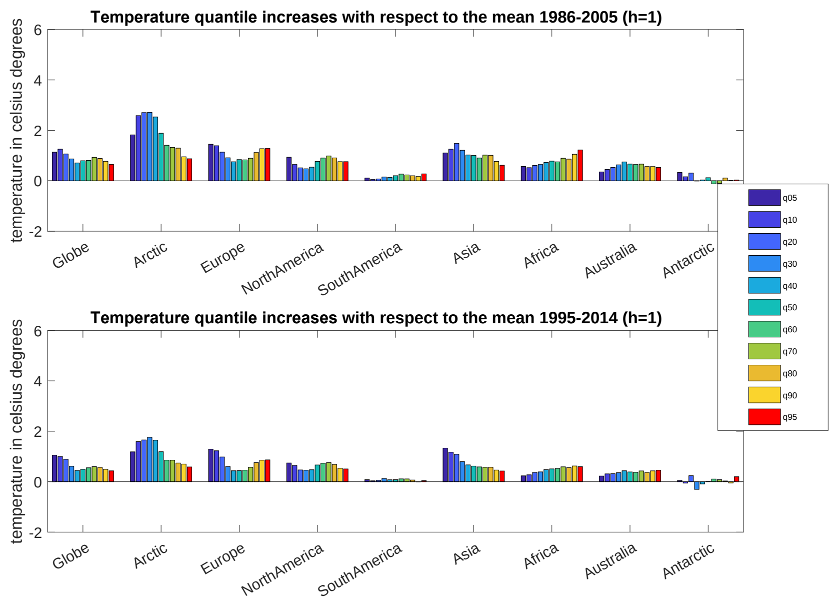


Figure 9
Comparing regional forecast (h=1, 1960-2023)

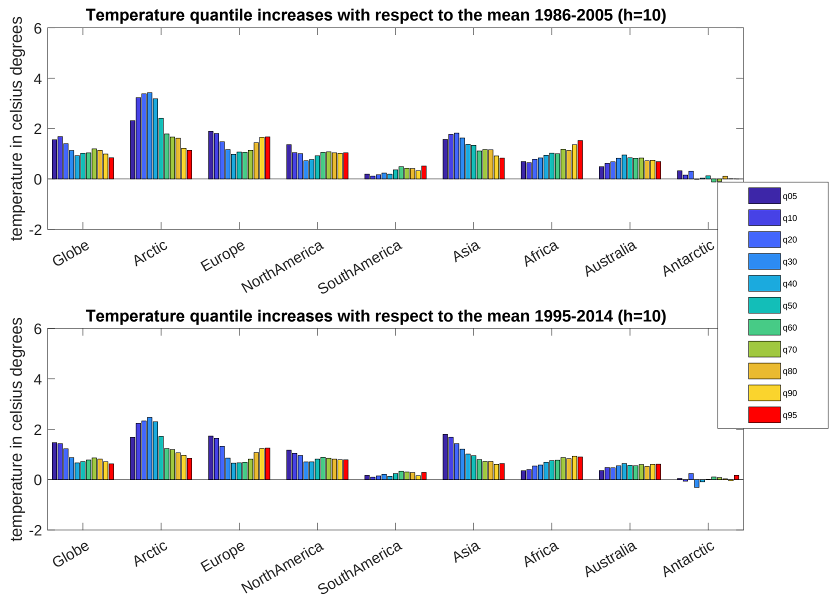


Figure 10
Comparing regional forecast (h=10, 1960-2023)

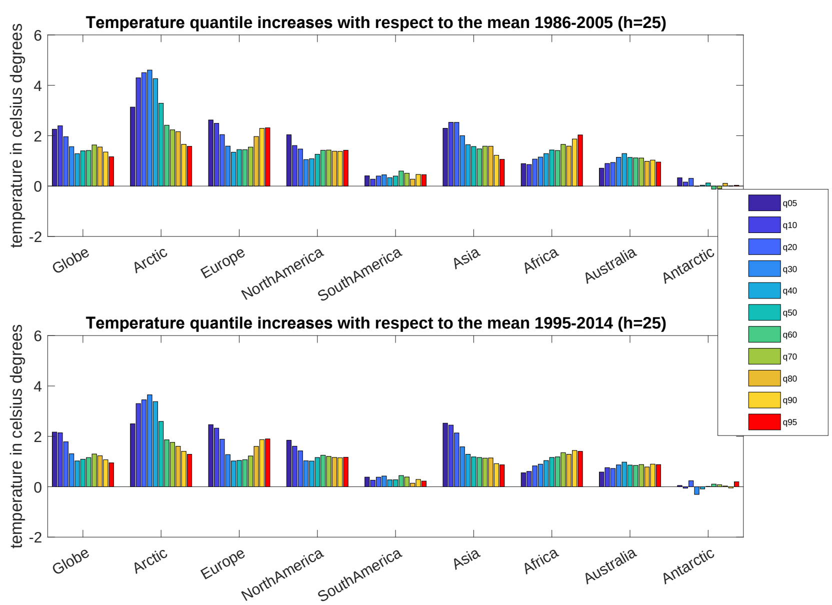


Figure 11
Comparing regional forecast (h=25, 1960-2023)

Table 20
Forecasted temperature increases in 2100

quantile	statistics	Globe	Arctic	Europe	NorthAmerica	SouthAmerica	Asia	Africa	Australia	Antarctic
q05	mean 1986-2005	4.58	6.37	5.07	4.19	0.03	4.80	1.61	1.55	0.30
	mean 1995-2014	4.49	5.74	4.92	3.99	-0.00	5.01	1.23	1.41	-0.05
q10	mean 1986-2005	4.79	7.84	4.68	3.35	0.01	5.15	1.56	1.76	0.06
	mean 1995-2014	4.52	6.84	4.52	3.34	-0.01	5.07	1.30	1.64	0.08
q20	mean 1986-2005	3.77	8.24	3.94	2.95	-0.05	4.94	2.14	1.81	0.24
	mean 1995-2014	3.59	7.16	3.78	2.90	-0.07	4.55	1.90	1.62	0.16
q30	mean 1986-2005	2.98	8.58	3.02	2.16	-0.00	3.81	2.19	2.29	-0.00
	mean 1995-2014	2.73	7.62	2.71	2.12	0.00	3.41	1.96	2.02	-0.30
q40	mean 1986-2005	2.13	8.08	2.65	2.14	0.08	3.08	2.50	2.10	0.05
	mean 1995-2014	1.86	7.20	2.32	2.08	-0.05	2.73	2.25	1.88	-0.11
q50	mean 1986-2005	2.32	6.44	2.71	2.32	0.03	2.89	2.80	2.13	0.20
	mean 1995-2014	2.01	5.74	2.30	2.20	-0.02	2.50	2.54	1.89	0.09
q60	mean 1986-2005	2.80	4.61	2.72	2.61	0.12	2.79	2.83	2.15	-0.16
	mean 1995-2014	2.53	4.04	2.35	2.44	-0.07	2.48	2.58	1.90	0.07
q70	mean 1986-2005	3.25	4.19	2.81	2.62	0.10	3.03	2.93	2.15	-0.07
	mean 1995-2014	2.92	3.71	2.49	2.40	-0.05	2.58	2.65	1.93	-0.01
q80	mean 1986-2005	2.87	3.81	3.71	2.57	0.15	3.00	3.04	1.95	0.13
	mean 1995-2014	2.54	3.26	3.34	2.35	0.01	2.55	2.75	1.75	-0.08
q90	mean 1986-2005	2.41	3.02	4.48	2.67	0.08	2.27	3.17	2.09	-0.05
	mean 1995-2014	2.12	2.77	4.07	2.45	-0.10	1.97	2.82	1.97	-0.04
q95	mean 1986-2005	2.25	3.00	4.49	2.63	0.21	2.06	3.73	1.92	0.03
	mean 1995-2014	2.04	2.71	4.08	2.39	-0.04	1.87	3.34	1.86	0.21

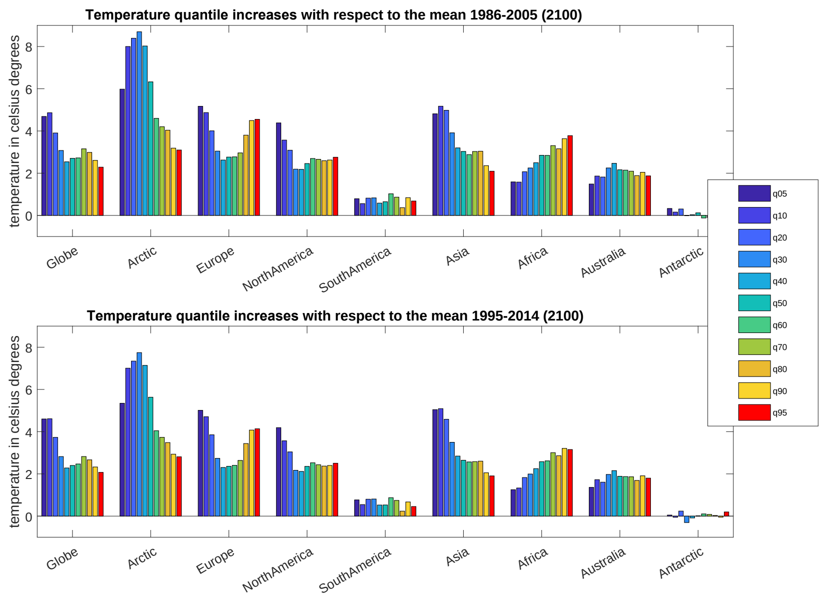


Figure 12
Comparing regional forecast 2100

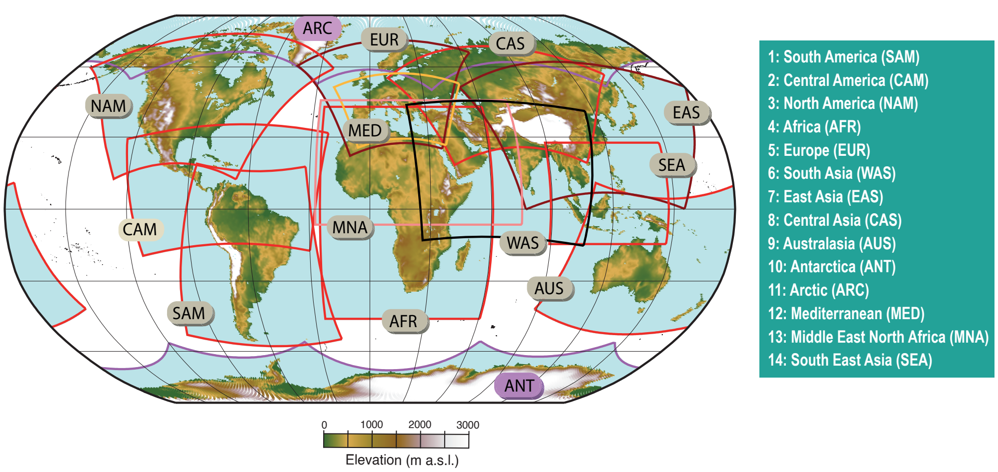


Figure Atlas.6 | CORDEX domains showing the curvilinear domain boundaries resulting from the original rotated domains. The topography corresponding to the standard CORDEX 0.44° resolution is shown to illustrate the orographic gradients over the different regions.

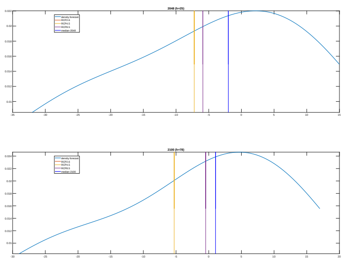
Figure 13
Atlas regional classification

Notes: Source: Gutierrez et al. (2021).

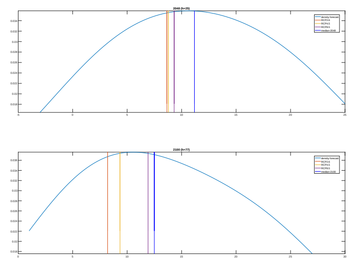
Figure 14

AR6 Atlas average projections and density forecasts

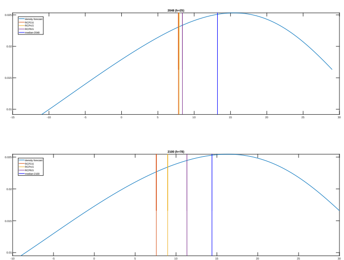
(a) Arctic Polar Circle



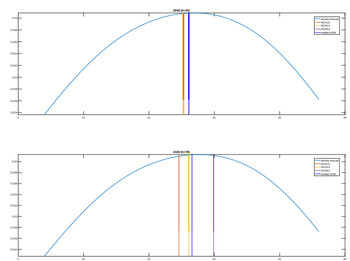
(b) Europe



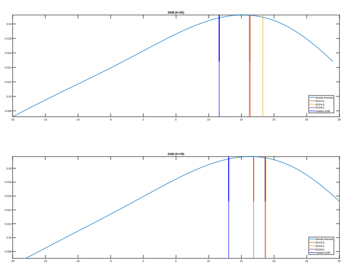
(c) North America



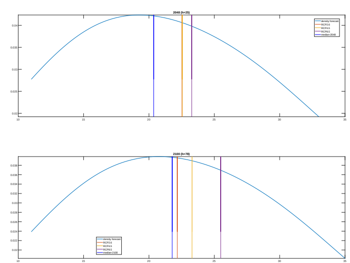
(d) South America



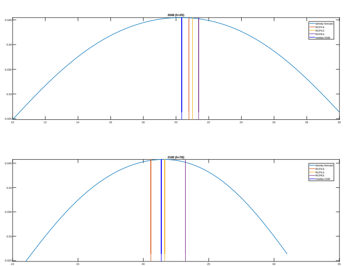
(e) Asia



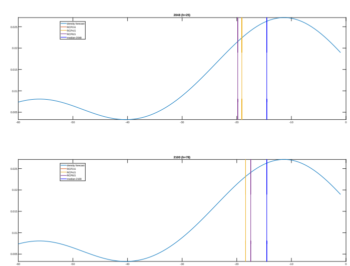
(f) Africa



(g) Australia



(h) Antarctic



7 Appendix: Additional exercises

7.1 A synthetic control experiment

As a complement to the previous prediction exercises, this section employs a synthetic control experiment, a methodology recently highlighted for its utility in prediction as an attribution tool (see Botosaru et al., 2023). The control group is derived from data spanning 1880 to 1960, a period prior to the significant acceleration of greenhouse gas concentrations in the atmosphere and the widespread adoption of climate change mitigation measures (see Global Carbon Budget, 2024). Using the methodology detailed in earlier sections, predictions are extended to the present, and the results are compared against observed temperature data.

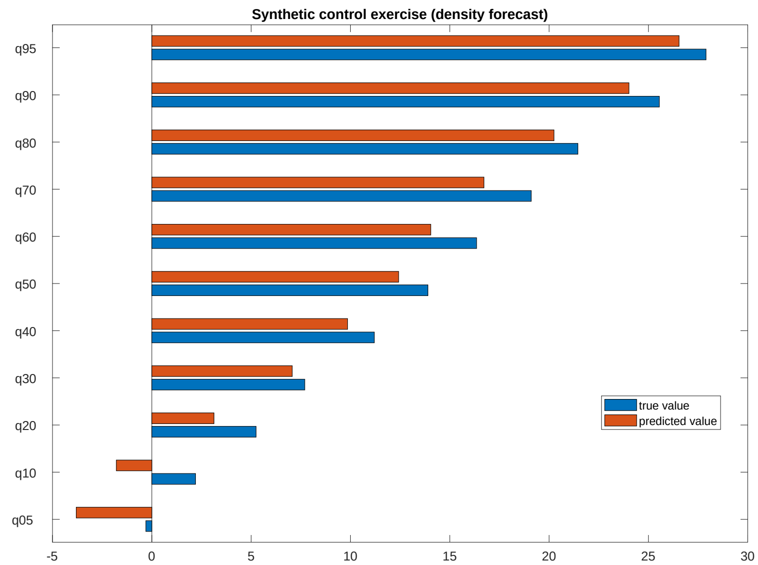
The counterfactual framing of this experiment is straightforward: What would have occurred in the absence of increased emissions concentrations and mitigation measures? Since these factors act in opposing directions—emissions driving warming and mitigation efforts aiming to reduce it—the experiment seeks to quantify the extent to which mitigation efforts have counteracted the effects of elevated emissions. Figure A-1 displays the density forecast of the projections. The results are unequivocal: in all cases, current observed temperatures exceed those projected under a business-as-usual (BAU) scenario extrapolated from the trends of the first half of the 20th century. This divergence is evident across all quantiles, though with varying intensities. The lower quantiles show the most significant discrepancies between observed and predicted temperatures, as highlighted in the lower panel of Figure A-1. This is in line with Chen et al. (2023) who show that a $\Delta C02$ accumulation affects the lower than the upper quantiles. These findings suggest that while mitigation measures have had some effect, they have not been fully effective in offsetting the warming effects of elevated emissions.

This experiment underscores the complex interplay between emissions growth and mitigation efforts. Although the latter have likely prevented more severe warming, the persistent gap between observed and counterfactual scenarios highlights the need for more robust and effective measures to combat global warming.

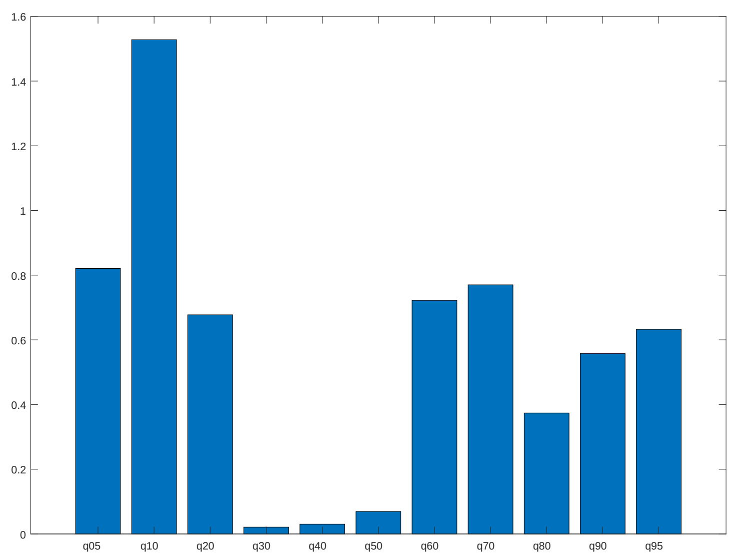
Figure A-1

Synthetic control exercise: Density forecast

(a) True and forecasted temperature by quantiles in 2023



(b) Differences by quantiles in the period 1961-2023



7.2 A short-run exercise in real time

Climate, by nature, is a long-term phenomenon. Accordingly, our methodological approach has been designed with this principle in mind. However, in this section, and without claiming to act as "weathermen", we demonstrate the effectiveness of our methods over shorter timeframes and in real-time scenarios. To this end, we conduct prediction exercises for the Globe. [27]. In all cases, data up to 1999 serve as the reference period, with real-time forecasts recursively calculated through 2024. The proposed methods are applied to both the mean and the distribution density.

For global forecasts, we use the same dataset as the rest of this study, covering the period 1880–1999 as the reference. Predictions are computed recursively for the years 2000–2024 with *h=1*. A fan plot (Figure A-2) illustrates the forecasts through 2024. The predictions generally align with observed trends but display non-negligible variability. Figure A-3 compares predictions from the benchmark linear model, the BIC-selected model, and various combinations of models. While the linear model performs adequately, combinations yield more precise predictions. Evaluation performance using RMSE (2000–2023), shows that model combinations, particularly those based on Pareto-superior models and the BIC criterion, achieve the lowest RMSEs. Figure A-4 compares real-time density forecasts against actual data.

Two key observations emerge: Predictions successfully capture the observed rightward shift in the temperature distribution. Predicted distributions tend to smooth out observed variability, reflecting the trend-based nature of the methodology.

[27] An application for the Central England and Madrid-Retiro station is available upon request.

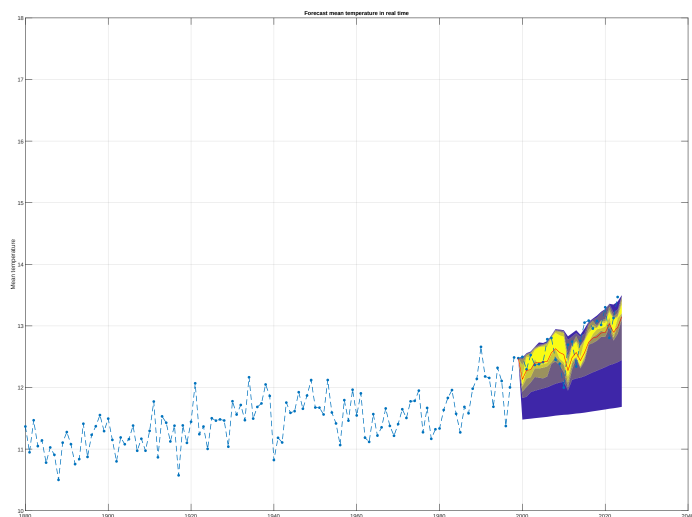


Figure A-2

Forecasted mean temperature in real time with individual models (Globe, 2000-2024)

Notes: The forecasts are made with each of the 14 models with recursive information for the period 2000-2024. The dots are the historical lines and the filled line (in red) indicates the mean of the forecasts.

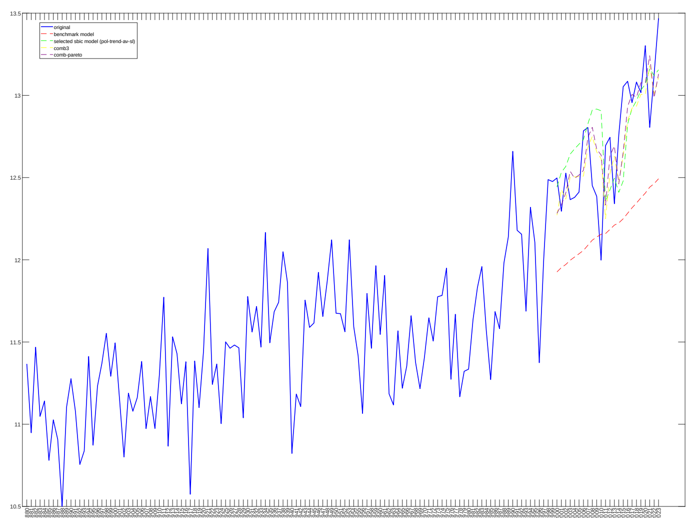


Figure A-3
Forecasted mean temperature in real time with combinations of models (Globe, 2000-2024)

Notes: Benchmark model refers to "linear trend"; selected model is "pol-trend-av-sl"; "comb3"removes the six extreme values; "comb-pareto" combines the selected Pareto-superior models by using the BIC weights.

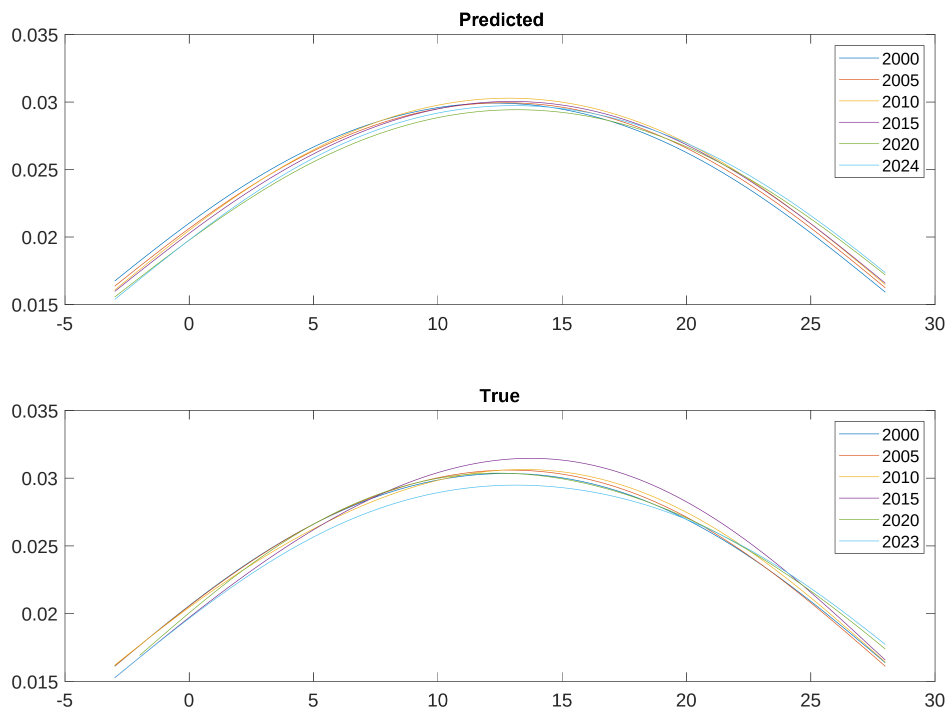


Figure A-4
Density forecast in real time (Globe, 2000-2024)