Adaptive now- and forecasting of global temperatures under smooth structural changes

Robinson Kruse-Becher

FernUniversitaet in Hagen, Lehrstuhl für Angewandte Statistik, Postfach 940, 58084 Hagen Germany.

E-mail: robinson.kruse-becher@fernuni-hagen.de

Summary. Accurate short-term now- and forecasting of global temperatures is an important issue and helpful for policy design and decision making in the public and private sector. We compose a raw mixed-frequency data set from weather stations around the globe (1920-2020). First, we document smooth variation in average monthly and annual temperature series by applying a dynamic stochastic coefficient model. Second, we use adaptive cross-validated forecasting methods which are robust to smooth changes of unknown form in the short-run. Therein, recent and past observations are weighted in a mean squared error-optimal way. Overall, it turns out exponential smoothing methods (with bootstrap aggregation) often performs best. Third, by exploiting monthly data, we propose a simple procedure to update annual nowcasts during a running calendar year and demonstrate its usefulness. Further, we show that these findings are robust with respect to climate zones. Finally, we investigate now- and forecasting of climate volatility via a range-based measure and a quantile-based climate risk measure.

1. Introduction

While long-run effects of climate change are well-studied, less is known about short horizons. In this paper, we focus on short-term forecasting of global temperatures. Short-term effects of climate change are of immediate concern to the private and public sector. Therefore, accurate short-term forecasts are valuable to decision making and policy design in near future, see e.g. Burke et al. (2015) and Cashin et al. (2017) for economic policy issues related to climate. Short-term forecasts are important in economics per se, see e.g. Giannone et al. (2008) for macroeconomic nowcasting which is nowadays at the core of research activities in e.g. central banks.

The forecasts in this work are generated by adaptive methods which are robust under recent and ongoing structural changes, see Giraitis et al.

(2013). These appear to be suitable for temperature series under continuous and ongoing climate change. The idea is to obtain robust forecasts under a wide range of possible underlying data generating processes. These might contain (non-)stationary stochastic and deterministic trends with structural breaks. Forecasts are generated by weighting recent and past observations in which the tuning parameter (e.g. a window size or a parameter controlling the weighting function) is selected via cross-validation. Thereby, the particular underlying data generating process does not need to be modelled explicitly. Giraitis et al. (2013) show the theoretical validity of such procedures. Rossi (2021) reviews many approaches to forecasting under instabilities and recommends the adaptive forecasting procedures by Giraitis et al. (2013) in case of small and continuous breaks, while other methods are more appropriate when breaks are large, discrete and abrupt.

The temperature data analyzed in this work is much better characterized by small and smooth shifts rather than large and abrupt breaks. Therefore, the applied adaptive methods are potentially helpful for improved forecasting. Raw station data is retrieved from the CRUTEM (Climatic Research Unit TEMperature) 5 data base and processed to form a balanced data set with high quality temperature measurements. In a first step of our preliminary analysis we fit a time-varying autoregressive model with a time-varying random attractor to the temperature series. The model and its nonparametric estimation technique are proposed in Giratis et al. (2014). The estimated model fits the data well and the results support the notion of smoothly evolving average temperatures.

We tackle the issue of one-year ahead forecasts and propose a simple but effective method for constructing nowcasts of average temperatures during a running calendar year. The resulting nowcasts can be updated month by month and the empirical investigation shows that it only takes two realized monthly observations (those from January and February) to significantly outperform the best forecasts using only annual data (obtained from a linear trend model with residual bagged exponential smoothing). Obviously, there is a trade-off between annual forecasts with relatively low estimation uncertainty, but no updating, and continuously updated averaged monthly forecasts with increased estimation uncertainty. The empirical results show that updating with realized values outweighs the remaining estimation uncertainty quite quickly during a running calendar year. Thereby, more accurate nowcasts can be offered. In general, these nowcasts have a monotonic mean squared error improvement over the calendar months by construction. We suggest a simple and consistent estimator for the calendar month at which the switch to monthly updated nowcasts occurs.

Our results for forecasting annual temperatures in the twelve different calendar months reveal a particular pattern in predictability: while predictability is relatively low between October and April, the opposite is found for the warmer months May to September. This pattern is also mirrored in the autocorrelation patterns of the calendar months series. In spite of some heterogeneity in the best performing approaches, mostly some variant of bagged exponential smoothing delivers the most accurate forecasts. Given that January and February values are relatively hard to forecast, updating the nowcasts with realized values explains the observed strong improvements in mean squared nowcast errors for the first calendar months. We also take the temporal hierarchy between monthly and annual forecasts into account and find that the so-called forecast reconciliation approach leads to slight improvements. These findings are robust with respect to climate zones.

We also focus on the range and lower and upper five percent quantiles of temperature distributions (for which we find similar patterns). These are particularly critical for climate change risk.

The remainder of the paper is organized as follows: Section 2 describes the data set and the construction of average temperatures from raw station data. In Section 3, a preliminary analysis is given with emphasis on robust trend testing and the estimation of the time-varying model by Giratis et al. (2014). The adaptive forecasting techniques are given in Section 4. The construction and updating of nowcasts is located in Section 5. Empirical results are reported in Section 6. Conclusions are drawn in Section 7.

2. Data

Data is obtained from CRUTEM 5 in NetCDF4 format. Using the R packages "ncdf4" and "ncdf4.helpers", data is converted from the nc format to regular time series data. These data are raw station data similar to the ones used in Gonzalo and Gadea (2020) for an earlier CRUTEM release.

We set the following standards to the selected weather stations in order to ensure excellent temperature data recording quality, minor measurement error and only a small proportion of missing data. In our experience, data quality varies considerably across weather stations around the globe. Criteria are in particular as follows:

(i) Start date of recording in 1920 (or earlier);
(ii) Continuous recording until 2020;
(iii) One hundred percent coverage of data in all twelve calendar months;

(iv) Less than five percent of missing data points in each calendar month.

Starting in 1920 offers a good balance between the length of the resulting annual series and their quality. In sum, we study 1,152 out of 10,639 available weather stations. These are of high available quality with continuous recordings and form a balanced sample with only few missing data points. The balancedness of the sample ensures the comparability across time points which is an important feature. Thereby, we rule out spurious effects due to a different composition of weather stations across time. This would make a comparison difficult. Moreover, our handling even enables to track individual weather stations (cf. Gadea and Gonzalo, 2021 and He et al., 2021).

The weather stations are located in the Americas (936), Europe (118), Russia (49), Asia (19), Pacific (17), Far East (11) and Africa (2). Clearly, many excluded stations do not fulfill the strong quality requirements. In fact, from the total sample of 10,639 stations, 3,229 are remaining when considering the start date restriction (i); 2,012 when additionally accounting for the end date (ii). The coverage restriction (iii) reduces the sample to 1,709 stations. The requirement regarding the completeness (iv) of the time series leaves us with $N = 1,152$ stations. Missing data is linearly imputed for all series.†

We consider the period from 1920 (January) to 2020 (December) yielding $T = 101$ annual observations. Exactly as in the CRUTEM data base, the annual temperature is calculated as an average from monthly observations (January to December) from the respective calendar year:

$$y_t^a = \frac{1}{N} \sum_{i=1}^{N} y_{t,i}^a \tag{1}$$

where y_t^a denotes the annual cross-sectional average computed from N weather stations measuring the individual annual average temperature $y_{t,i}^a$ with $i = 1, 2, ...N$. The individual annual average per station is defined as

$$y_{t,i}^a = \frac{1}{12} \sum_{m=1}^{12} y_{t,i}^m$$

where $y_{t,i}^m$ denotes the monthly recorded temperature in month $m = 1, 2, .., 12$ at station i in year t. This brings us directly to

$$y_t^a = \frac{1}{12N} \sum_{m=1}^{12} \sum_{i=1}^{N} y_{t,i}^m \ . \tag{2}$$

†A complete list of weather stations with WMO id is available upon request.

From the raw station data, we construct the average (global) temperature. This measurement is of highest interest to academics, policy makers, practitioners and households. Besides, the variation around the mean is informative as well (see e.g. Gonzalo and Gadea, 2020 and Diebold and Rudebusch, 2022). Following Diebold and Rudebusch (2022), we consider the range. Other variation measures are less persistent leading to less predictability and less responsive forecasts. Finally, we also consider the lower and upper five percent quantile as a climate-at-risk measure. These are discussed in Section 6.3.

3. Trend analysis

Table 1 reports realized values of the annual and monthly temperature average time series for the years 1920, 1970 and 2020. Predictions from the individually selected forecasting method (in terms of lowest MSE) for the year 2021 are also given.

Looking at the values of temperatures in different years increases are clearly visible for the annual average, but also for most of the individual calendar months. We start our trend analysis by running a robust trend test by Gonzalo and Gadea (2020). The authors suggest robust OLS-based HAC inference (see e.g. Newey and West (1987) and Andrews (1991)) in a simple linear regression framework:

$$y_t = \alpha + \beta t + u_t \ . \tag{3}$$

Their t-test for $H_0 : \beta = 0$ is shown to be robust against various forms of deterministic and stochastic trends. Results are reported in Table 1, column t_{HAC}. Critical values are taken from the asymptotic distribution which standard normal. All reported statistics are positive and significant (at least at the nominal significance level of five percent) except for the October series. Overall, the existence of upward trends is confirmed for almost all series suggesting the existence of global warming.

Next, we apply a dynamic stochastic coefficient process as proposed by Giraitis et al. (2014) to model the temperature series. Thereby, we decompose the climate series into a random persistent attractor and a dynamic part with a time-varying autoregressive coefficient:

$$y_t - \mu_t = \rho_{t-1}(y_{t-1} - \mu_{t-1}) + u_t.$$

The model can be re-expressed as a time-varying autoregressive process containing a time-varying intercept α_t:

$$y_t = \alpha_t + \rho_{t-1} y_{t-1} + u_t \tag{4}$$

Table 1. Trend test results

Series	1920	1970	2020	2021 (pred)	t_{HAC}
Annual	9.98	9.79	11.12	11.26	4.13
Jan	-4.58	-2.65	-0.46	-0.74	2.84
Feb	-0.20	-0.66	1.65	1.11	2.66
Mar	6.04	5.12	5.15	4.05	3.95
Apr	8.67	9.37	8.74	10.32	4.13
May	15.06	13.60	17.19	15.15	4.42
Jun	19.62	19.10	20.56	20.02	2.65
Jul	21.06	21.42	22.66	22.50	2.25
Aug	21.13	20.43	21.92	21.65	4.10
Sep	15.52	16.69	18.59	18.03	2.15
Oct	11.96	11.52	11.44	11.93	1.14
Nov	4.51	4.78	4.41	5.63	3.55
Dec	0.85	-1.28	1.62	0.31	2.35

with $\alpha_t = \mu_t - \rho_{t-1}\mu_{t-1}$. Such a model has been applied previously to inflation and real exchange rates. Both time-varying quantities α_t and ρ_t can be estimated consistently by nonparametric estimation methods and point-wise confidence intervals are provided. As a result, the estimated random attractor is informative about the time-varying underlying intercept. The dynamic autoregressive parameter provides a time-varying measure for the persistence in the dynamic component.

Figure 1 displays the estimation of the persistent random attractor and the dynamic AR coefficient for the annual series, while Figures 2 and 3 show the ones for annual calendar month series. Point-wise confidence intervals reported at the 90% level. We use a normal kernel and a bandwidth of $\sqrt{T}$ as the theory suggests, see Theorem 2.4 and Corollary 2.3 in Giraitis et al. (2014).

The directions of random attractors are in conjunction with signs of estimated linear OLS slopes in the linear trend test regression. Throughout, the shapes of estimated random attractors track the trajectories well and support the notion of smooth and continuous structural changes in temperature series rather than large and abrupt breaks. The estimated dynamic persistence shows some heterogeneous results with different patterns but typically indicated low or moderate levels of persistence. Clearly, static persistence ignoring any trends (red vertical line) is reduced as expected when accounting for smoothly varying attractors. Diagnostic tests on residuals (Ljung-Box tests with one and five lags) reveal no remaining

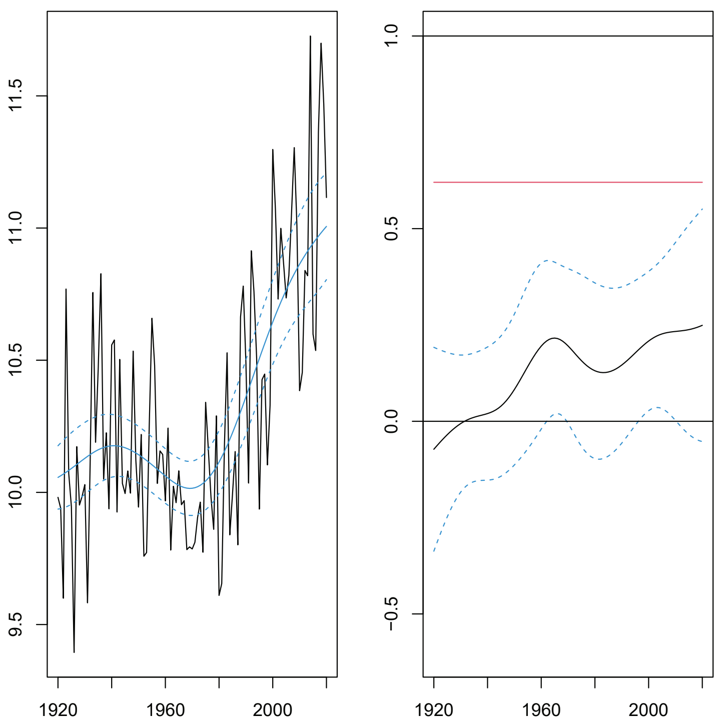

Fig. 1. Time-varying attractor and persistence, annual average

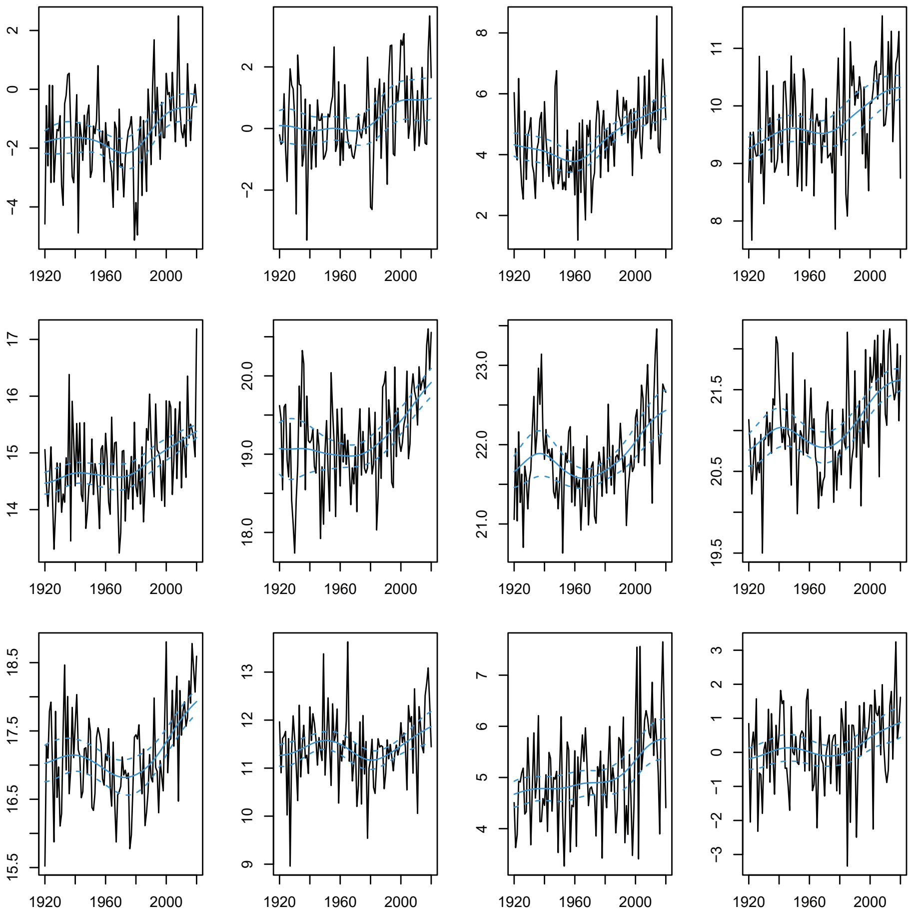

Fig. 2. Time-varying attractor, calendar months

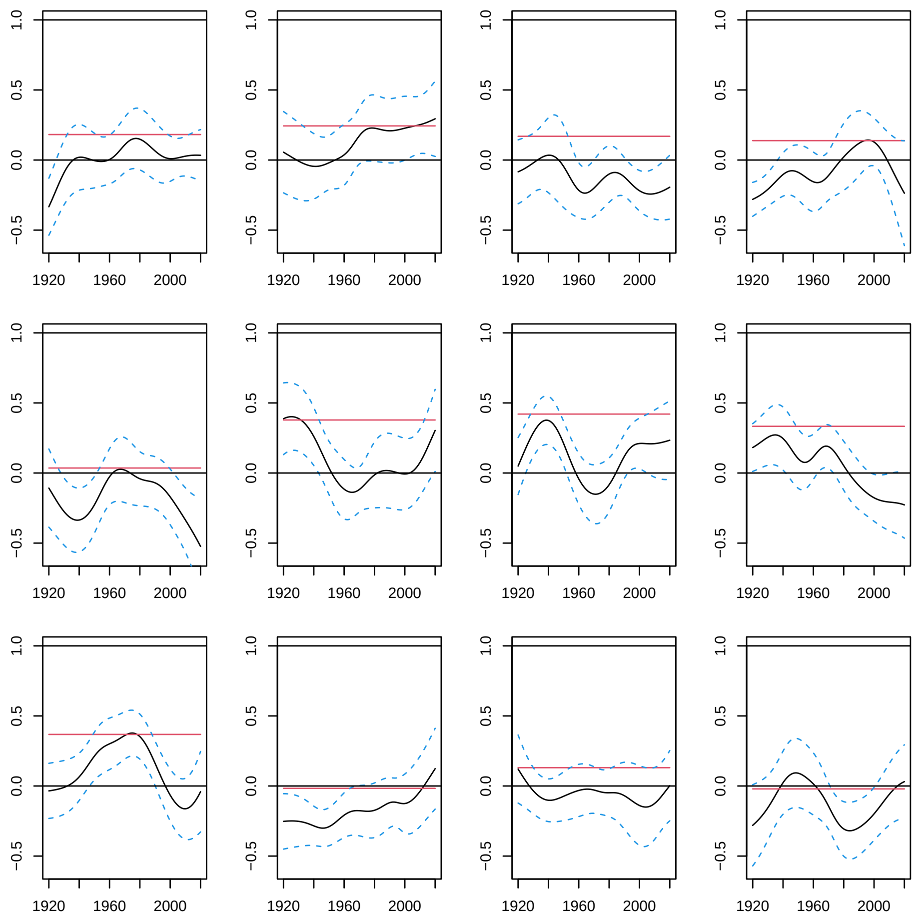

Fig. 3. Time-varying persistence, calendar months

autocorrelation at conventional significance levels.‡ Our evidence clearly points towards smooth and continuous changes and trends for which the following robust forecasting devices are appropriate, see the excellent survey by Rossi (2021) and the discussion therein.

4. Adaptive forecasting techniques

General setup We consider a simple location framework following Giraitis et al. (2013):

$$y_t = \beta_t + u_t. \tag{5}$$

The noise term u_t is a martingale difference sequence being independent of β_t. Here, we follow the general framework and do not specify the trend component β_t, but rather allow for a wide class of stochastic or deterministic processes with minimal structure. β_t does not have to be smooth and may contain a unit root. Importantly, the nature of β_t is neither assumed to be known specifically, nor estimated.§

The adaptive forecasting scheme is about weighting the most recent and past observations in a flexible way:

$$\widehat{y}_{t+1|t,H} = \sum_{i=0}^{t-1} w_{t,i,H} y_{t-i}. \tag{6}$$

The weights $w_{t,i,H}$ are restricted to be positive and to sum up to one. They are, in general, depending on some tuning parameter H for which we consider cross-validation on the in-sample one-step ahead forecasts, see below.

Different weighting schemes are available, e.g. parametric exponential weighting and nonparametric schemes. Besides, some simple routines are considered. Among these is the mean of all available observations $y_1, y_2, ..., y_t$ with equal weights (Mean). Such a forecast would be optimal in case of a pure white noise process. Next, just using the last observation y_t gives a driftless random walk forecast (Last). Besides these two extremes of either equally weighting all observations or just focusing on

‡Results are not reported to conserve space and are available upon request.

§Inoue et al. (2017) suggest several extensions (e.g. multi-step forecasts) to the framework of Giraitis et al. (2013), see also Farmer et al. (2022) for a general linear predictive regression. These authors have different aspects in view, while we follow Giraitis et al. (2013) and focus on simple procedures with optimal weighting of recent and past data through cross-validation. Their methods fit the empirical situation of interest in this work very well.

the last one, all possible variations can be averaged, see Pesaran and Timmermann (2007).

$$\widehat{y}_{t+1|t} = \frac{1}{t}\sum_{H=1}^{t} \widehat{y}_{t+1|t,H}$$

with

$$\widehat{y}_{t+1|t,H} = \frac{1}{H}\sum_{i=t-H+1}^{t} y_i \ .$$

This approach (Avg) combines all possible averages and might offer some robustness with respect to unknown persistence of the underlying time series. Finally, we consider a triangular weighting scheme (Tri), see Giraitis et al. (2013).

Exponential smoothing has proven to be successful in short-term forecasting in general and under small and continuous breaks in particular, see e.g. Petropoulos et al. (2022). We therefore consider exponential smoothing weights

$$w_{t,i,H} = \frac{\gamma^i}{\sum_{j=1}^{t}\gamma^j} \tag{7}$$

with $\gamma = \exp(-H^{-1})$. Here, the tuning parameter H controls the degree of down-weighting past observations. Rather than using fixed or predetermined values, we resort to cross-validation for which Giraitis et al. (2013) have established a theoretical framework showing that data-driven selection of H, i.e. $\widehat{H}$, yields MSE-optimal weights.¶ Their results hold under a wide range of processes, e.g. stationary, stochastic and deterministic trends and structural breaks. Such forecasting devices have proven to be successful for many macroeconomic time series undergoing smooth structural changes, but less is known for climate time series.

Additionally to the regular exponential smoothing procedure, we consider bagging, see Breimann (1996), Inoue and Kilian (2008) and Hillebrand and Medeiros (2010) for time series applications. Here, forecasts of multiple bootstrap samples are aggregated (via averaging) to forecast the original series. Bergmeir et al. (2016) have shown that bagged exponential smoothing yields accurate forecasts. Related works are Dantas and Oliveira (2018) and Petropoulos et al. (2018). On the contrary, Barrow et

¶The cross-validation uses $\widehat{H} = \arg\min_H Q_{T,H}$ with $Q_{T,H} = \frac{1}{T}\sum_{t=1}^{T}(y_t - \widehat{y}_{t|t-1,H})^2$ based on in-sample observations. Giraitis et al. (2013) show that the minimal MSE can be obtained by cross-validation under a wide range of possible processes including deterministic and stochastic trends and structural breaks.

al. (2020) find bagging to perform worse. To the best of our knowledge, no results are yet available for temperature series. We thus provide new evidence on the empirical performance of bagging for exponential smoothing.

The procedure works as follows: A variance-stabilizing Box-Cox transformation with data-driven parameter selection (see Guerrero, 1993) is applied in a first step. As our data is non-seasonal, a Loess-based procedure (see Cleveland et al., 2017) is used for decomposition. Residuals are bootstrapped via the moving block bootstrap by Kuensch (1989) to form bootstrapped series for which exponential smoothing is used to generate forecasts which are finally averaged.

As an alternative to parametric exponential smoothing, we consider a nonparametric weighting method. For instance, it is not required that weights decrease monotonically which can be helpful during multiple structural changes. The weights are determined by minimizing the one-step ahead forecast MSE subject to the constraint that they sum up to one. The Lagrangian leads to a system of equations which are solved by imposing $\beta_t = \widehat{\beta}_t$ for $t = 1, 2, ..., T-1$ and $\beta_T = \widehat{\beta}_T = \widehat{\beta}_{T-1}$. We use a local constant model with cross-validated bandwidth choice. For technical details on the procedure, we refer to Giraitis et al. (2013).

Next, we consider the possibility of a parametric conditional mean model, e.g. a linear deterministic trend, i.e. $\widehat{y}_{t+1|t} = \widehat{\alpha} + \widehat{\beta}(t+1)$. A linear trend might approximate for instance (logarithmic) carbon-dioxide emissions and real gross domestic product per capita. We study the resulting pure linear trend forecasts (Trd) in separation, but also in combination with exponential or nonparametric weighting applied to the residuals of the linear trend model (Trd+ES, Trd+ES* and Trd+NP). By doing so, eventually neglected features can be picked up. Thus, residuals can be processed in an adaptive way controlling for potential deviations from white noise in a robust way.

As a benchmark forecast, we take a first-order autoregressive model (AR1). This benchmark covers many possible situations ranging from a white noise to a random walk.|| Similarly to the linear trend model, we also consider exponential or nonparametric weighting for the residuals, labeled

||We have also considered the simple robust predictor suggested in Martinez et al. (2022) (see also Castle et al., 2015 and Hendry, 2018) which is given as $\widehat{y}^a_{t+1|t} = y^a_t + \rho \Delta y^a_t$. It always performs better than the AR1 benchmark in terms of MSE, but worse than exponential smoothing in all cases. This might be due to the fact that their procedure performs very well for rather large and abrupt breaks. Our data is instead characterized by small and smooth changes. The authors also study a smooth robust predictor. It is, however, unclear how to

as AR1+ES (or AR1+ES*) and AR1+NP, respectively.

As a final layer of robustness, we equally weight all forecasts from different sources, see Elliott and Timmermann (2013) for an excellent survey article. This technique does require additional estimation and is known to be difficult to be beaten by more sophisticated weighting schemes. The combination of all forecasts with equal weights is labeled as 'Comb'.

5. Nowcasting annual averages by monthly data

The methods used in this work draw heavily from Giraitis et al. (2013). These are designed for non-seasonal (e.g. annual) data and focus on one-step ahead forecasts. Therefore, our primary interest lies in annual one-step ahead short-term forecasts. Besides, we are interested in nowcasts for the running calendar year. In order to enable nowcasts from non-seasonal monthly data, we consider twelve annual time series for the different calendar months. Hence, we forecast the one-step ahead annual average from a time series of e.g. January temperature series. By doing so, we obtain non-seasonal annual data and can apply the same set of methods as for the annual averaged series. Importantly, by collecting January to December forecasts we can construct nowcasts during the running calendar year by consecutively replacing month forecasts with realized values.

As a by-product, we can also compare the annual forecast obtained from annual series with one obtained from averaging twelve monthly forecasts. Naturally, the latter one has a much higher uncertainty stemming from the fact that twelve (instead of a single) estimations are needed. Nonetheless, such a comparison is interesting per se and allows us to 'reconcile' the forecasts (see e.g. Athanasopoulos et al., 2017). Clearly, there is a temporal hierarchy in the data and the forecasts as annual observations are defined as averages of monthly observations. Interestingly, equally averaging (i) the annual forecast and (ii) the aggregated monthly one, leads to a reconciled forecast exploiting the temporal structure of the forecasts, see also Hollyman et al. (2021).**

select the window size optimally. One way might be to apply cross-validation techniques. This extension is left for future research.

**We have also considered other ways of forecast reconciliation, see Athanasopoulos et al. (2017). The reconciled forecast can be interpreted as a GLS estimator which involves the structural scaling matrix and the covariance matrix of reconciliation errors. More sophisticated variants of reconciliation beyond the simple structural scaling with equal variances did not improve on the MSE and are therefore not further considered.

The annual temperature average is defined as

$$y_t^a = \frac{1}{12} \sum_{m=1}^{12} y_t^m$$

where y_t^m denotes the contribution of month $m = \{1, 2, ..., 12\}$. Now, y_t^a can be forecasted from an annual series which is the most obvious choice, i.e. $\widehat{y}_{t+1|t}^a$. Another possibility would be to forecast the individual monthly contributions and to take their average to form the annual forecast. As mentioned, the monthly forecasts can be obtained from annual series again, i.e. $\widehat{y}_{t+1|t}^m$. The annual forecast is then obtained as

$$\widehat{\overline{y}}_{t+1|t}^a = \frac{1}{12} \sum_{m=1}^{12} \widehat{y}_{t+1|t}^m.$$

Averaging is a unique and preserving linear transformation and also allows the construction of nowcasts as follows. Let $\widehat{y}_{t+1|t}^{a,m}$ denote the annual nowcasts in month m. Precisely, data up to month m is available. Due to lags in reporting, this does not necessarily mean that such a nowcast is actually computed in month m. The nowcast can simply be constructed as follows:

$$\widehat{y}_{t+1|t}^{a,m} = \frac{1}{12} \left(\sum_{i=1}^{m} y_{t+1}^i + \sum_{i=m+1}^{12} \widehat{y}_{t+1|t}^i \right). \tag{8}$$

Here, the realized monthly temperatures for months 1 to m are used and the remaining $n = 12 - m$ months are forecasted. Setting $m = 0$ yields the annual averaged forecast $\widehat{\overline{y}}_{t+1|t}^a$ based on monthly forecasts. As m increases, more and more realized values are incorporated and less and less forecasts are needed. Setting $m = 12$ yields the realized annual average y_{t+1}^a.

It is clearly expected that the nowcast MSE, i.e.

$$\omega(a, m) = E\left[(y_{t+1}^a - \widehat{y}_{t+1|t}^{a,m})^2 \right], \tag{9}$$

is monotonically decreasing in m, i.e. $\omega(a, m+1) \leq \omega(a, m)$, see Fosten and Gutknecht (2020) for a related study. Moreover, the nowcast MSE $\omega(a, m)$ approaches zero as m approaches 12. In our empirical study, we investigate at which $\widetilde{m}$ it is recommendable to switch from the annual forecast $\widehat{y}_{t+1|t}^a$ to the continuously updated nowcast $\widehat{y}_{t+1|t}^{a,m}$. Clearly, for low values of m, the annual forecast has the advantage of reduced uncertainty, while the nowcast updated with realized values gains attraction as m increases. We determine $\widetilde{m}$ by first constructing a confidence interval (based on HAC standard errors) for the MSE of annual forecast $\omega(a)$ and then taking the first m for which the nowcast MSE $\omega(a, m)$ is below the lower confidence bound, i.e.

$$\widetilde{m} = \inf_{0 \leq m \leq 11} \{m : \omega(a, m) < \overline{\omega}(a) + se_{HAC}(\overline{\omega}(a)) q_{\alpha_T}\}. \tag{10}$$

At this point and beyond, due to monotonicity, the nowcast is performing significantly better and thus preferable. For consistency of the estimator, the size α_T needs to shrink to zero as T grows to infinity, see inter alia Phillips et al. (2011).

6. Empirical results

6.1. Forecasting results

In total, we have 16 forecast procedures in competition. The following forecast evaluation tables report the MSE relative (Rel MSE) to the AR(1) benchmark. A value below unity indicates a better performance and vice versa. Next, three statistics obtained from the Mincer-Zarnowitz (MZ) regression

$$y_{t+1} = \alpha + \beta \widehat{y}_{t+1|t} + u_{t+1} \tag{11}$$

are reported: (i) the regression R^2 measuring the degree of predictability, (ii) the Wald statistic for testing rationality of the forecasts, i.e. $H_0 : \alpha = 0 \cup \beta = 1$, (critical value at the five percent level equals 5.99) and (iii) the bias t-statistic for $H_0 : \alpha = 0 \mid \beta = 1$ (imposing a slope coefficient of unity, i.e. $\beta = 1$) with a five percent critical value of 1.96. These entries are followed by p-values for the Ljung-Box statistic with one lag applied to the one-step forecast errors and their squares. The p-value for the Jarque-Bera statistic is also reported. Hence, we consider important optimality properties of the forecast errors, namely uncorrelatedness in the first two moments and normality. Finally, the model confidence set (MCS) p-value (see Hansen et al. 2011) is reported for those models included in the model confidence set. We run this procedure with a nominal significance level of twenty-five percent, see e.g. Bennedsen et al. (2021). Missing entries indicate that the respective procedure is eliminated and not contained in the final model confidence set. All computations are carried out in the open-source statistical software R.

Table 2. Forecast evaluation results, annual average

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.82	0.92	1.25	1.55	1.00	1.12	0.74	0.82
MZ R^2	0.49	0.44	0.54	0.51	0.45	0.52	0.50	0.48
MZ Wald	102.81	12.40	36.82	66.78	12.50	30.21	3.56	5.16
MZ bias (t)	2.62	0.50	3.40	2.97	3.17	3.68	1.88	2.23
LB [e] p-val	0.00	0.10	0.00	0.00	0.11	0.01	0.23	0.04
LB [e2] p-val	0.00	0.96	0.07	0.01	0.48	0.15	0.74	0.39
JB [e] p-val	0.40	0.90	0.36	0.39	0.52	0.31	0.41	0.38
MCS p-val							0.98	0.50
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.73	0.79	0.75	0.75	0.72	0.82	0.80	0.83
MZ R^2	0.51	0.48	0.50	0.50	0.52	0.48	0.50	0.53
MZ Wald	3.55	4.47	4.76	3.95	3.98	6.13	5.78	10.34
MZ bias (t)	1.81	2.14	2.09	1.95	1.83	2.48	2.42	3.01
LB [e] p-val	0.38	0.48	0.95	0.22	0.38	0.37	0.07	0.21
LB [e2] p-val	0.73	0.55	0.61	0.67	0.71	0.52	0.40	0.61
JB [e] p-val	0.45	0.63	0.70	0.42	0.51	0.57	0.41	0.38
MCS p-val	1.00	0.62	1.00	0.87	1.00	0.32	0.62	0.37

Table 2 reports the evaluation results for the annual time series forecasts $\widehat{y}^a_{t+1|t}$. Starting point is 1971 for the forecasting period. Thus, the period from 1920 to 1970 contains the observations for the estimation sample which is used for the first forecast in 1971. From that point onward, the estimation sample is recursively extended. All optimizations and cross-validations are repeated in each step, i.e. on an annual basis.

The best performing specification is a linear trend with an additional bagged exponential smoothing (Trd+ES*) for the residuals. Its relative MSE is 0.72 and close to other well performing forecasts obtained from exponential smoothing and its variants. Nonparametric weighting also performs relatively good with relative MSE values around 0.8. The predictability (as measured by the Mincer-Zarnowitz regression R^2) is 0.52 for the best performing Trd+ES* forecast indicating a medium level of predictability in the underlying temperature series. The resulting forecasts are found to be rational and unbiased. The forecast errors appear to be uncorrelated, homoskedastic and normally distributed. The model confidence set includes ten forecasts in which the ES*, AR1+ES* and Trd+ES* reach the largest MCS p-values. From this perspective, the results are indicative that bagged exponential smoothing performs very well and best if applied to either linear trend or autoregressive residuals. Nonparametric forecasts are not unbiased and have autocorrelated errors. Furthermore, simple approaches do not perform well.

Results for the evaluation of forecasts of annual series for the twelve calendar months are not reported to save space and available from the author upon request. The individual results for the different calendar months indicate that the best performing forecasts are well specified in most cases. In a few situations, the rationality and bias statistics are significant at the five percent level, but not at the one percent level. While the best performing forecasts for the annual averaged series are obtained by the Trd+ES* approach, the results for calendar months are different and diverse. Most often, bagged exponential smoothing and its variants appear to perform best. We also find that the combination of all available forecasts (with equal weights) performs best (for January and February). Even simple linear trend forecasts without additional residual treatment are selected, albeit with very low levels of predictability (November and December).

Overall, predictability seems to be quite low for the colder months October-April. For the warmer months, larger predictability is found. Thus, there is some form of seasonality in the predictability. This is also resembled in the underlying autocorrelation of the series, see the red horizontal lines in Figures 1 and 3. June and September appear to have largest

R^2-values (0.37 and 0.36, both with bagged exponential smoothing). However, the predictability is remarkably lower in comparison to the averaged annual level. A similar pattern is seen in the relative MSEs which are closer, but still significantly below unity. Moreover, the trend test results also indicate somewhat weaker (albeit still significant) evidence against the null of no trend.

Figure 4 plots the global temperature average together with the best performing annual forecasts from the Trd+ES* approach in blue color. Next, the plot also contains the aggregated annual forecast obtained from twelve best performing individual forecasts for the respective monthly contributions in red color (i.e. $\widehat{\overline{y}}^a_{t+1|t}$). An equal weighting of the two forecasts gives the reconciled forecast exploiting the temporal hierarchy among the annual and monthly series. The annual forecast has the lowest MSE, followed by the reconciled forecast. The averaged monthly forecast performs worst among the three, see the nowcast MSE results plotted in Figure 5 for $m = 0$.

6.2. Nowcasting results

Figure 5 depicts the nowcast MSEs for the yearly forecast $\widehat{y}^a_{t+1|t}$ which is constant with respect to m by definition as there is no updating. Monthly updates result in monotonically decreasing nowcast MSEs in m. They approach zero for m being close to 12. Remarkably, even after one month (i.e. $m = 1$), the nowcast MSE is already smaller for the updated nowcast albeit insignificant (at the 10% level). However, after two months, the updated nowcast has a significantly lower MSE than the annual average, yielding $\widetilde{m} = 2$. The combination (with equal weights exploiting the temporal hierarchy) improves slightly on the pure nowcast for $m = 1$, but not for the remaining months in the year. Clearly, due to low predictability for the months January to April - as found above -, the nowcasts strongly benefit from replacing potentially difficult forecasts in a low predictability environment with realized values. The gains for the remaining months are still noticeable, but the largest magnitude in nowcast MSE reduction is observed for January and February. The loss from using the monthly forecast even for $m = 0$ is rather small as compared to the gain from using it for $m = 1, 2, ..., 11$. From a statistical viewpoint, a switch from the annual forecast to the continuously updated monthly nowcasts is advisable from February onward since $\widetilde{m} = 2$. Even though there might be a reporting lag which hinders the practitioners to exploit the realized monthly observations immediately, the advantages of updates are clearly documented here. Our recommendation is thus to switch from the annual forecast to

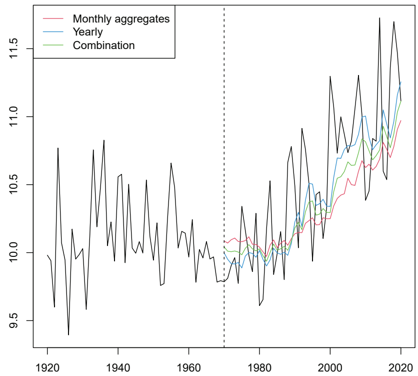

Fig. 4. Forecast reconciliation plot

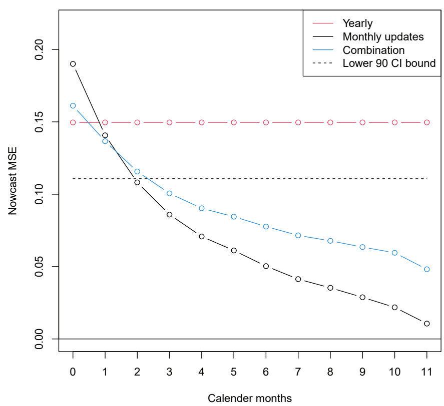

Fig. 5. Nowcast MSE comparison

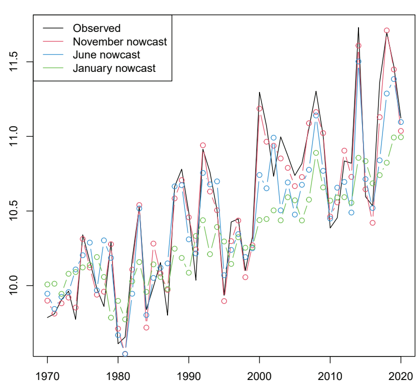

Fig. 6. Nowcast updates

the monthly updated nowcasts once the February data is made available to the forecaster.

Figure 6 plots the observed temperature average together with three different nowcasts obtained during the running calendar year. The January nowcast exploits the realized January value and uses forecasts for the remaining eleven months. These are obtained from the best performing models. The June nowcast uses the first six observed monthly values and forecasts for the months July to December. It can be seen that the forecasts track the global temperature evolution quite well. Updating the nowcasts with monthly information improves the forecast accuracy strikingly. Unsurprisingly, the November nowcast is almost perfectly following the annual averages as only the December forecast is needed in addition to the other eleven realized monthly observations.

6.3. Nowcasting climate zones

We now turn to the analysis of different climate zones in order to investigate whether there are notable differences. So far, the global temperature averages are analyzed. A study of climate zones can reveal geographical differences in the strength of climate change and its short-term predictability.

Following the Koeppen-Geiger classification scheme, we exploit longitude and latitude values to match the location of the analyzed weather stations in our sample with the climate zones. The classification results in 19 stations in polar climate, 444 stations in snow climate, 537 stations in warm climate and 150 stations in arid climate. Two stations in equatorial climate are discarded.

Figures 7–10 plot the annual temperature averages in the four different climate zones. In all four zones, a clear and significant upward trend is present, but the strongest (and least noisiest) trend is found for the polar region. This is reflected in a relatively high degree of predictability ($R^2 = 0.83$ for the best performing Trd+ES* forecast), see Table 3. Remarkably, this forecast is the only one included in the model confidence set. The results for the remaining three climate zones (snow, warm and arid) are quite similar, see Tables 4–6. Most notably, the Trd+ES* forecast is best performing in all climate zones.

Turning to the nowcasting results, we find a very similar pattern to the global averages, see Figures 11–14. The optimal switch from annual to updated forecasts occurs in February. The polar zone is a slight exception with $\widetilde{m} = 3$ (March).

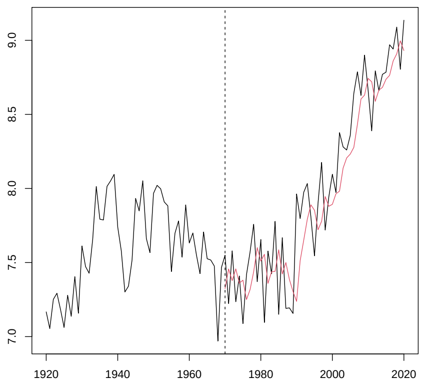

Fig. 7. Forecast plot - polar zone

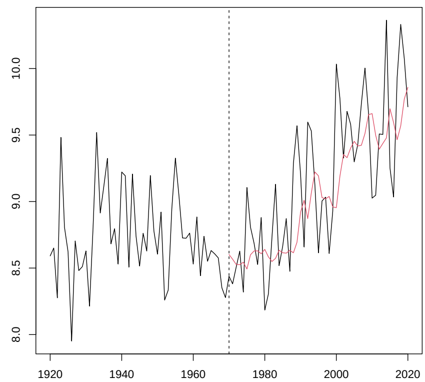

Fig. 8. Forecast plot - snow zone

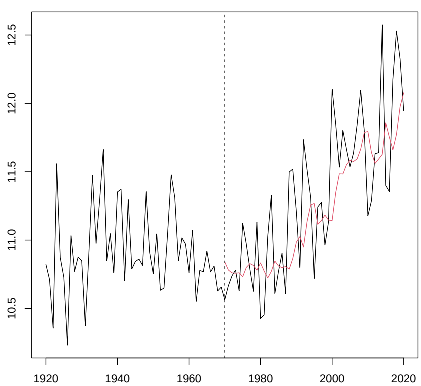

Fig. 9. Forecast plot - warm zone

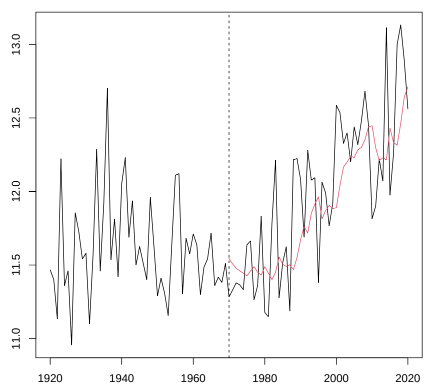

Fig. 10. Forecast plot - arid zone

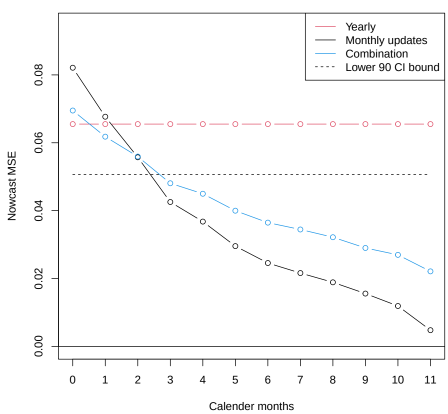

Fig. 11. Nowcast MSE comparison - polar zone

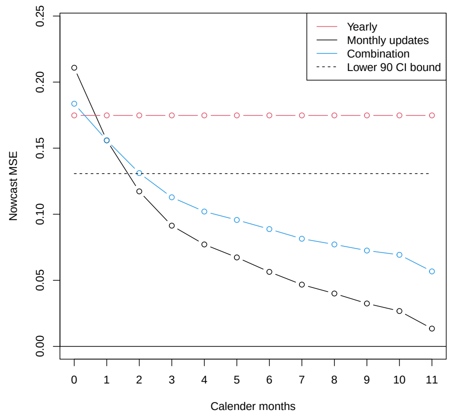

Fig. 12. Nowcast MSE comparison - snow zone

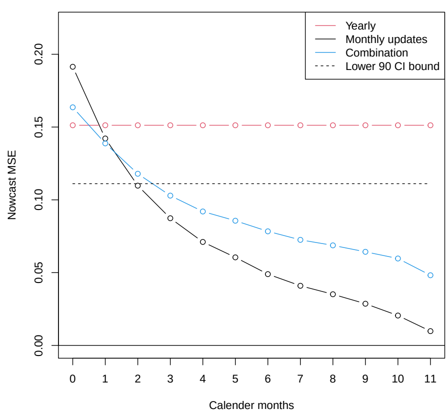

Fig. 13. Nowcast MSE comparison - warm zone

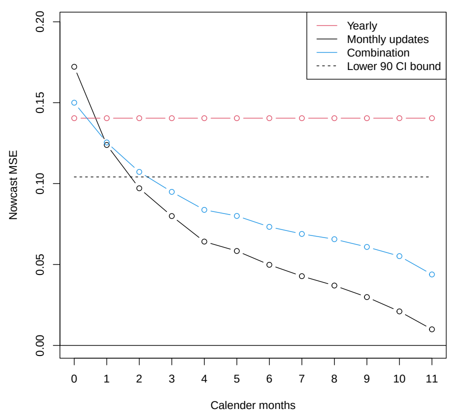

Fig. 14. Nowcast MSE comparison - arid zone

Table 3. Forecast evaluation results, annual average, polar

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	4.65	0.94	2.89	3.81	1.00	2.46	0.68	0.74
MZ R^2	0.70	0.76	0.78	0.69	0.77	0.68	0.83	0.82
MZ Wald	106.52	6.34	112.22	90.21	23.24	46.29	6.11	7.78
MZ bias (t)	1.39	1.44	1.74	1.41	3.02	1.40	2.24	2.26
LB [e] p-val	0.00	0.00	0.00	0.00	0.21	0.00	0.17	0.06
LB [e2] p-val	0.00	0.53	0.00	0.00	0.05	0.00	0.70	0.44
JB [e] p-val	0.19	0.81	0.22	0.20	0.26	0.18	0.65	0.70
MCS p-val								

	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.67	0.98	0.92	0.67	0.66	0.93	0.72	1.04
MZ R^2	0.83	0.77	0.78	0.83	0.83	0.78	0.82	0.82
MZ Wald	7.28	19.26	17.00	4.79	5.69	18.70	5.79	64.43
MZ bias (t)	2.37	2.90	2.94	2.05	2.17	2.83	1.97	3.23
LB [e] p-val	0.26	0.17	0.03	0.18	0.26	0.09	0.06	0.30
LB [e2] p-val	0.68	0.08	0.08	0.74	0.71	0.05	0.41	0.21
JB [e] p-val	0.73	0.32	0.36	0.71	0.79	0.29	0.76	0.23
MCS p-val					1.00			

Table 4. Forecast evaluation results, annual average, snow

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.77	0.97	1.24	1.51	1.00	1.12	0.80	0.91
MZ R^2	0.44	0.41	0.49	0.46	0.42	0.47	0.44	0.38
MZ Wald	100.86	14.82	33.92	64.76	12.03	28.20	2.79	4.19
MZ bias (t)	2.81	0.48	3.58	3.18	3.21	3.82	1.66	2.09
LB [e] p-val	0.00	0.17	0.01	0.00	0.13	0.01	0.16	0.10
LB [e2] p-val	0.01	0.84	0.10	0.03	0.53	0.17	0.76	0.48
JB [e] p-val	0.36	0.77	0.31	0.34	0.40	0.29	0.38	0.34
MCS p-val							0.70	
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.78	0.82	0.79	0.80	0.77	0.86	0.87	0.86
MZ R^2	0.45	0.45	0.46	0.45	0.46	0.44	0.43	0.48
MZ Wald	3.46	4.50	4.08	3.27	4.08	5.80	5.34	8.98
MZ bias (t)	1.69	2.15	2.00	1.76	1.75	2.44	2.32	2.91
LB [e] p-val	0.28	0.42	0.72	0.14	0.26	0.30	0.10	0.19
LB [e2] p-val	0.81	0.61	0.77	0.59	0.79	0.59	0.44	0.63
JB [e] p-val	0.45	0.47	0.50	0.41	0.51	0.43	0.45	0.32
MCS p-val	1.00	1.00	1.00	1.00	1.00		0.26	0.52

Table 5. Forecast evaluation results, annual average, warm

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.76	0.92	1.21	1.50	1.00	1.08	0.74	0.79
MZ R^2	0.49	0.44	0.54	0.51	0.44	0.52	0.50	0.49
MZ Wald	92.47	11.93	35.04	60.56	12.77	28.41	3.34	5.49
MZ bias (t)	2.58	0.49	3.34	2.90	3.14	3.57	1.84	2.28
LB [e] p-val	0.00	0.08	0.00	0.00	0.09	0.01	0.24	0.06
LB [e2] p-val	0.00	0.91	0.08	0.02	0.43	0.19	0.82	0.43
JB [e] p-val	0.41	0.85	0.38	0.41	0.54	0.34	0.38	0.38
MCS p-val								0.38
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.71	0.78	0.73	0.72	0.70	0.81	0.95	0.82
MZ R^2	0.51	0.48	0.51	0.51	0.52	0.48	0.36	0.52
MZ Wald	3.43	4.38	4.73	4.00	3.74	6.18	5.88	10.15
MZ bias (t)	1.83	2.12	2.12	2.00	1.86	2.48	1.83	2.93
LB [e] p-val	0.43	0.49	1.00	0.22	0.43	0.37	0.12	0.22
LB [e2] p-val	0.76	0.51	0.54	0.76	0.76	0.47	0.98	0.66
JB [e] p-val	0.43	0.63	0.71	0.43	0.47	0.61	0.01	0.43
MCS p-val	1.00	0.46	1.00	0.85	1.00		0.85	

Table 6. Forecast evaluation results, annual average, arid

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.66	0.91	1.15	1.04	1.00	1.04	0.69	0.75
MZ R^2	0.47	0.41	0.53	0.51	0.42	0.51	0.50	0.48
MZ Wald	68.56	10.71	34.29	25.36	15.25	25.36	3.08	4.59
MZ bias (t)	2.57	0.49	3.32	3.68	3.13	3.68	1.76	2.10
LB [e] p-val	0.00	0.01	0.01	0.02	0.11	0.02	0.28	0.06
LB [e2] p-val	0.00	0.31	0.05	0.11	0.50	0.11	0.53	0.22
JB [e] p-val	0.47	0.78	0.40	0.26	0.58	0.26	0.45	0.38
MCS p-val		0.30					0.72	0.37
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.67	0.73	0.71	0.70	0.66	0.76	0.74	0.78
MZ R^2	0.51	0.48	0.49	0.50	0.52	0.48	0.50	0.52
MZ Wald	3.09	4.22	4.31	3.76	3.60	6.79	5.75	11.65
MZ bias (t)	1.77	2.03	2.09	1.95	1.89	2.49	2.39	2.93
LB [e] p-val	0.53	0.72	0.78	0.24	0.54	0.59	0.10	0.32
LB [e2] p-val	0.45	0.29	0.24	0.39	0.35	0.40	0.18	0.48
JB [e] p-val	0.52	0.70	0.86	0.47	0.56	0.65	0.42	0.48
MCS p-val	1.00	1.00	1.00	1.00	1.00	0.72		

6.4. Annual range and Climate-at-Risk

In this final subsection we focus on additional measures beyond the mean. First, we consider the range of temperatures, defined as the difference between the maximal and the minimal temperature, i.e.

$$r_t^a = \max_i y_{t,i}^a - \min_i y_{t,i}^a \tag{12}$$

with $y_{t,i}^a = \frac{1}{12}\sum_{m=1}^{12} y_{t,i}^m$ being the annual temperature average at station i. Such a measure is important to judge the spread of the temperature distribution. The temperature range has been investigated in related recent studies like Diebold and Rudebusch (2022). Second, we consider the lower and upper five percent quantile of the temperature distribution characterizing the "Climate-at-Risk", labeled as $q_{t,\alpha}^a = \inf\{y_t \in \mathbb{R} : \alpha \leq F(y_t)\}$ with $\alpha = 5\%$ and $\alpha = 95\%$ and F being the distribution of y_t. The idea of considering quantiles in the context of risk is similar to the famous "Value-at-Risk" approach in financial econometrics (see e.g. Christoffersen, 2009). In financial markets, agents typically hold long positions in the portfolios and thus the risk is located in the left tail. A noticeable exception is Giot and Laurent (2003). Similarly, in macroeconomics, the concept of "Growth-at-Risk" has emerged (Brownlees and Souza, 2021). In global warming, both, the lower and the upper tail are relevant for climate risk assessment, see also Gonzalo and Gadea (2020). Due to the fact that neither the range nor the quantiles are mean-preserving functions, the nowcasting method for the average temperature series cannot be adopted. We therefore focus on the forecasts of annual series in the remainder of this work.

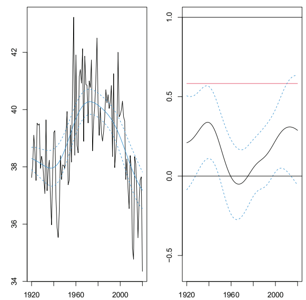

Fig. 15. Time-varying attractor and persistence, annual range

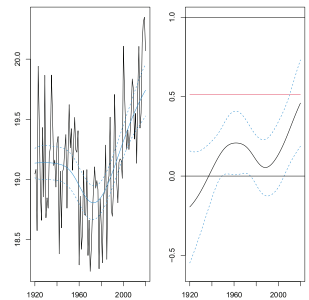

Fig. 16. Time-varying attractor and persistence, annual 95% quantile

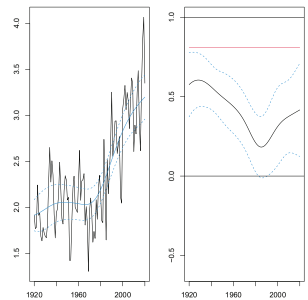

Fig. 17. Time-varying attractor and persistence, annual 5% quantile

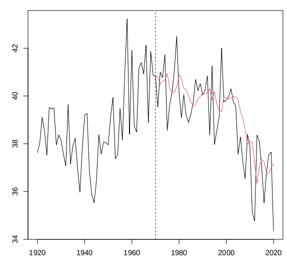

Fig. 18. Forecast plot, annual range

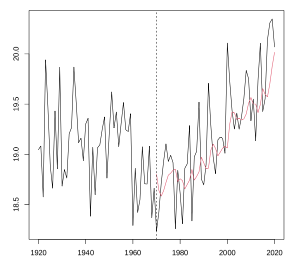

Fig. 19. Forecast plot, annual 95% quantile

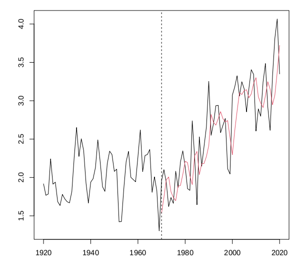

Fig. 20. Forecast plot, annual 5% quantile

Figure 15 displays the estimated random attractor and dynamic persistence for the annual range. While the range was increasing until the mid-1960s it experiences a smooth and steady decline until the end of the sample in 2020. A decreasing range coupled with an increasing mean further amplifies the severity of climate change. Moreover, the persistence increased from the 1960s onward. Figure 18 plots the forecasts from the most accurate method (in terms of MSE, relative to the benchmark: 0.83), namely the cross-validated exponential smoothing method (ES). Table 7 contains the evaluation results. The degree of predictability ($R^2 = 0.49$) is comparable to the average series. All diagnostic tests are passed. The model confidence set includes the bagged exponential smoothing forecasts, the autoregressive forecasts with exponentially smoothed residuals, the nonparametric forecasts (also in combination with an autoregressive conditional mean model) and the equally-weighted combination scheme.

Finally, the random attractor estimation results are displayed in Figures 16 and 17. Both are clearly smoothly upward trending in the recent decades. Moreover, both series are somewhat stronger autocorrelated than other annual temperature series. The forecast evaluation results for the lower and upper quantiles are reported in Tables 8 and 9. Both quantiles are characterized by medium predictability, while the best forecasting method for the upper quantile (ES*) yields a relative MSE of 0.71 as opposed to the lower quantile, where the linear the linear trend model plus ES* forecasts attain 0.88. For the upper quantile, the MCS contains ten forecasts, while the one for the lower only consists of five. Overall, (bagged) exponential smoothing also works very well for temperature series reflecting the climate risk. Figures 19 and 20 show that the movements in lower and upper climate tail risk can be reasonably tracked.

Table 7. Forecast evaluation results, annual range

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.63	1.10	1.33	1.55	1.00	2.02	0.83	0.91
MZ R^2	0.01	0.39	0.44	0.13	0.37	0.38	0.49	0.47
MZ Wald	1.30	9.63	30.86	1.94	0.88	46.35	3.40	4.67
MZ bias (t)	-0.01	-0.79	-1.35	-0.84	-0.41	-3.56	-1.74	-2.18
LB [e] p-val	0.00	0.02	0.00	0.00	0.84	0.00	0.66	0.09
LB [e2] p-val	0.20	0.89	0.06	0.04	0.69	0.01	0.71	0.14
JB [e] p-val	0.23	0.90	0.33	0.30	0.50	0.44	0.84	0.62
MCS p-val							1.00	0.74
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.85	0.91	0.92	0.89	0.90	0.88	0.99	0.91
MZ R^2	0.48	0.44	0.43	0.50	0.49	0.46	0.48	0.49
MZ Wald	3.53	3.69	3.22	8.42	8.35	2.82	9.69	5.71
MZ bias (t)	-1.71	-1.80	-1.65	-2.62	-2.56	-1.62	-3.07	-2.08
LB [e] p-val	0.75	0.24	0.23	0.68	0.77	0.40	0.13	0.57
LB [e2] p-val	0.79	0.73	0.75	0.73	0.78	0.67	0.13	0.42
JB [e] p-val	0.86	0.77	0.76	0.83	0.87	0.75	0.56	0.62
MCS p-val	0.46	0.27				1.00		0.49

Table 8. Forecast evaluation results, annual 95% quantile

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	1.57	0.90	1.24	1.48	1.00	1.44	0.73	0.82
MZ R^2	0.26	0.47	0.54	0.43	0.47	0.50	0.56	0.52
MZ Wald	10.59	11.54	52.13	59.39	20.35	62.40	6.36	8.84
MZ bias (t)	1.22	0.61	2.26	1.79	1.46	3.61	2.27	1.92
LB [e] p-val	0.00	0.02	0.00	0.00	0.01	0.00	0.28	0.02
LB [e2] p-val	0.00	0.92	0.01	0.00	0.61	0.02	0.89	0.46
JB [e] p-val	0.76	0.99	0.94	0.86	0.99	0.99	0.81	0.91
MCS p-val		0.32					0.78	0.43
	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.71	0.76	0.72	0.79	0.75	0.79	0.88	0.84
MZ R^2	0.56	0.54	0.56	0.53	0.55	0.54	0.52	0.56
MZ Wald	5.15	7.13	6.30	7.27	6.38	11.45	11.57	19.39
MZ bias (t)	2.10	2.19	2.17	2.61	2.47	1.84	3.09	2.54
LB [e] p-val	0.43	0.70	0.98	0.15	0.37	0.23	0.06	0.12
LB [e2] p-val	0.74	0.68	0.58	0.88	0.74	0.88	0.44	0.75
JB [e] p-val	0.76	0.57	0.65	0.73	0.63	0.89	0.84	0.93
MCS p-val	1.00	0.68	1.00	0.34	0.68	0.68		0.26

Table 9. Forecast evaluation results, annual 5% quantile

	Mean	Last	Avg	Tri	AR1	Trd	ES	NP
Rel MSE	3.15	1.00	1.83	2.40	1.00	1.31	0.94	1.40
MZ R^2	0.58	0.57	0.61	0.58	0.59	0.58	0.56	0.40
MZ Wald	117.41	8.92	55.96	80.69	12.96	24.59	6.67	9.15
MZ bias (t)	3.31	0.75	3.80	3.47	3.58	3.23	1.47	2.01
LB [e] p-val	0.00	0.18	0.00	0.00	0.95	0.00	0.48	0.23
LB [e2] p-val	0.00	0.63	0.00	0.00	0.84	0.05	0.57	0.20
JB [e] p-val	0.51	0.83	0.55	0.51	0.53	0.46	0.58	0.83
MCS p-val		0.74						

	ES*	AR1+ES	AR1+ES*	Trd+ES	Trd+ES*	AR1+NP	Trd+NP	Comb
Rel MSE	0.91	0.92	0.89	0.92	0.88	0.95	1.31	1.00
MZ R^2	0.57	0.60	0.60	0.57	0.58	0.60	0.42	0.60
MZ Wald	5.97	8.82	7.56	6.04	5.34	9.96	8.59	11.50
MZ bias (t)	1.50	2.81	2.49	1.26	1.28	3.10	1.62	3.42
LB [e] p-val	0.48	0.86	0.71	0.46	0.44	0.95	0.25	0.31
LB [e2] p-val	0.43	0.48	0.50	0.66	0.59	0.68	0.30	0.23
JB [e] p-val	0.64	0.77	0.77	0.61	0.67	0.65	0.81	0.59
MCS p-val			1.00	0.38	1.00			0.48

7. Conclusions and outlook

Climate change is an indisputable and challenging issue affecting global society. A major concern are average temperatures and their deviations from historic means. For a newly composed high-dimensional data set from 1920 to 2020 based on the CRUTEM 5 data base, we document smooth variation and investigate robust forecasting devices for the short-term horizon of one year ahead. Cross-validated exponential smoothing (in combination with bootstrap aggregation) turns out to be a successful and robust forecasting device. We offer a simple and robust procedure to construct and update nowcasts for a running calendar year. Results show that updating with monthly realizations significantly improves nowcasts already after two months in comparison to the best annual forecast. The analysis of climate zones reveals robustness of the previous findings and the particular strength of climate change in the polar zone. Moreover, we study the range in annual temperature distributions and the lower (upper) five percent quantile in the context of climate tail risk assessment and forecasting. Our findings are similar to the ones for the average annual temperature forecasting. While the range is decreasing over time, both quantiles are increasing. Taken these facts together with an increasing mean culminates in serious signals regarding global warming from different distributional characteristics of temperatures.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica, 59(3), 817-858.

Athanasopoulos, G., Hyndman, R. J., Kourentzes, N., & Petropoulos, F. (2017). Forecasting with temporal hierarchies. European Journal of Operational Research, 262(1), 60-74.

Barrow, D., Kourentzes, N., Sandberg, R., & Niklewski, J. (2020). Automatic robust estimation for exponential smoothing: Perspectives from statistics and machine learning. Expert Systems with Applications, 160, 113637.

Bennedsen, M., Hillebrand, E., & Koopman, S. J. K. (2021). Modeling, forecasting, and nowcasting U.S. CO2 emissions using many macroeconomic predictors. Energy Economics, 96, Article 105118.

Bergmeir, C., Hyndman, R. J., & Benítez, J. M. (2016). Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. International Journal of Forecasting, 32(2), 303-312.

Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

Brownlees, C., & Souza, A. B. (2021). Backtesting global growth-at-risk. Journal of Monetary Economics, 118, 312-330.

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

Cashin, P., Mohaddes, K., & Raissi, M. (2017). Fair weather or foul? The macroeconomic effects of El Niño. Journal of International Economics, 106, 37-54.

Castle, J. L., Clements, M. P., & Hendry, D. F. (2015). Robust approaches to forecasting. International Journal of Forecasting, 31(1), 99-112.

Christoffersen, P. (2009). Value–at–risk models. In Handbook of financial time series (pp. 753-766). Springer, Berlin, Heidelberg.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (2017). Local regression models. In Statistical models in S (pp. 309-376). Routledge.

Dantas, T. M., & Oliveira, F. L. C. (2018). Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing. International Journal of Forecasting, 34(4), 748-761.

Diebold, F. X., & Rudebusch, G. D. (2022). On the evolution of US temperature dynamics. In Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling. Emerald Publishing Limited.

Elliott, G., & Timmermann, A. (Eds.). (2013). Handbook of economic forecasting. Elsevier.

Farmer, L. E., Schmidt, L & Timmermann, A. (2022). Pockets of Predictability. Journal of Finance, forthcoming.

Fosten, J., & Gutknecht, D. (2020). Testing nowcast monotonicity with estimated factors. Journal of Business & Economic Statistics, 38(1), 107-123.

Gadea, M. D., & Gonzalo, J. (2020). Trends in distributional characteristics: Existence of global warming. Journal of Econometrics, 214(1), 153-174.

Gadea, M. D., & Gonzalo, J. (2021). A tale of three cities: climate heterogeneity. SERIEs, 13, 1-37.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. Journal of Monetary Economics, 55(4), 665-676.

Giot, P., & Laurent, S. (2003). Value-at-risk for long and short trading positions. Journal of Applied Econometrics, 18(6), 641-663.

Giraitis, L., Kapetanios, G., & Price, S. (2013). Adaptive forecasting in the presence of recent and ongoing structural change. Journal of Econometrics, 177(2), 153-170.

Giraitis, L., Kapetanios, G., & Yates, T. (2014). Inference on stochastic time-varying coefficient models. Journal of Econometrics, 179(1), 46-65.

Guerrero, V. M. (1993). Time-series analysis supported by power transformations. Journal of Forecasting, 12(1), 37-48.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453-497.

He, C., Kang, J., Teräsvirta, T., & Zhang, S. (2021). Comparing long monthly Chinese and selected European temperature series using the Vector Seasonal Shifting Mean and Covariance Autoregressive model. Energy Economics, 97, 105171.

Hendry, D. F. (2018). Deciding between alternative approaches in macroeconomics. International Journal of Forecasting, 34(1), 119-135.

Hillebrand, E., & Medeiros, M. C. (2010). The benefits of bagging for forecast models of realized volatility. Econometric Reviews, 29(5-6), 571-593.

Hollyman, R., Petropoulos, F., & Tipping, M. E. (2021). Understanding forecast reconciliation. European Journal of Operational Research, 294(1), 149-160.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. Journal of Econometrics, 196(1), 55-67.

Inoue, A., & Kilian, L. (2008). How useful is bagging in forecasting economic time series? A case study of US consumer price inflation. Journal of the American Statistical Association, 103(482), 511-522.

Kuensch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. The Annals of Statistics, 17(3), 1217-1241.

Martinez, A. B., Castle, J. L., & Hendry, D. F. (2022). Smooth robust multi-horizon forecasts. In Essays in Honor of M. Hashem Pesaran: Prediction and Macro Modeling. Emerald Publishing Limited.

Mincer, J. A., & Zarnowitz, V. (1969). The evaluation of economic forecasts. In Economic forecasts and expectations: Analysis of forecasting behavior and performance (pp. 3-46). NBER.

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. International Economic Review, 28(3), 777-787.

Pesaran, M. H., & Timmermann, A. (2007). Selection of estimation window in the presence of breaks. Journal of Econometrics, 137(1), 134-161.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., ... & Ziel, F. (2022). Forecasting: theory and practice. International Journal of Forecasting, 38(3), 705-871.

Petropoulos, F., Hyndman, R. J., & Bergmeir, C. (2018). Exploring the sources of uncertainty: Why does bagging for time series forecasting work?. European Journal of Operational Research, 268(2), 545-554.

Phillips, P. C., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?. International Economic Review, 52(1), 201-226.

Rossi, B. (2021). Forecasting in the Presence of Instabilities: How We Know Whether Models Predict Well and How to Improve Them. Journal of Economic Literature, 59(4), 1135-90.