Decline in Temperature Variability on Svalbard✑

SONDRE HØLLELAND AND HANS ARNFINN KARLSEN

University of Bergen, Bergen, Norway

(Manuscript received 13 March 2020, in final form 5 July 2020)

ABSTRACT

The variability in the temperature on Svalbard, Norway, has been decreasing over the last four decades. This may be due to the reduction in sea ice, transitioning the regional climate to a more stable, coastal one. We quantify this transition in terms of decreasing volatility in a daily average temperature time series at Svalbard Airport from 1976 to 2019. We use two different approaches: a nonstochastic model and a time-dependent generalized autoregressive conditional heteroskedasticity (GARCH) model. These parametric approaches include a time-dependent trend, where the slope depends on the day of the year. For Svalbard, the slope has a minimum in late August and the steepest slope during winter is estimated to be $-0.1^{\circ}\text{C}^2\,\text{yr}^{-1}$. The nonstochastic model, for which the conditional and unconditional variances are the same, only depends on the marginal distribution and is perhaps the easiest to interpret. The GARCH model extends the nonstochastic model by including short-range temporal dependence in the volatility and is thus more locally adapted. Volatility modeling is important for a complete statistical description of the temperature dynamics on Svalbard as an Arctic representative. In combination with increasing temperatures, the volatility reduction makes the extremely cold days during winter occur less frequently. Although we focus exclusively on the Svalbard Airport series, the models should be suitable for other temperature or climatic time series.

1. Introduction

For temperature time series, shifts in mean temperature over time are a main focus, but we can also ask: what about the variation? Variation, variability, or volatility is often measured in variance or standard deviation. We analyze the time series of daily average temperatures measured at Svalbard Airport (Norway; red dot in Fig. 1) from 1976 to 2019, with main focus on the evolution of the volatility. The location is chosen for its proximity to the North Pole (78.2°N) and is thereby a representative of the Arctic, showing enhanced effects of climate change. It is also a particularly interesting location because of the relatively large difference between summer and winter day-to-day volatility.

Over the last decades the yearly mean temperature on Svalbard has been increasing extensively. Isaksen et al. (2016) suggest that this increase in temperature is driven by sea ice decline, higher sea surface temperature, and a general background warming and Kohnemann et al. (2017) reach a similar conclusion. Onarheim et al. (2014) discuss causes for loss of sea ice north of Svalbard during winter. They found that warm Atlantic Ocean water is likely to have caused the sea ice loss. Declining extent of sea ice has consequences for the regional climate in the Arctic, but research is also being done on the effect at lower latitudes (Screen et al. 2015). The report *Climate in Svalbard 2100* (Hanssen-Bauer et al. 2019), commissioned by the Norwegian Environment Agency, discusses past, present, and projected future climate on Svalbard under different emission scenarios. The report is an assessment of existing literature and model results [cf. Hanssen-Bauer et al. (2019) and references therein] related to climate on Svalbard, but also presents some new results from atmosphere, ocean, and hydrological models. Under the scenario RCP8.5 (referred to as "business as usual" or "high emissions") they project an increase in annual mean temperature of almost 10°C, a 65% increase in annual precipitation, and a 20% increase in heavy rainfalls from the reference period 1971–2000 to 2071–2100. The projections are based on physical climate models and, although they are associated with high uncertainty, they show where the Svalbard

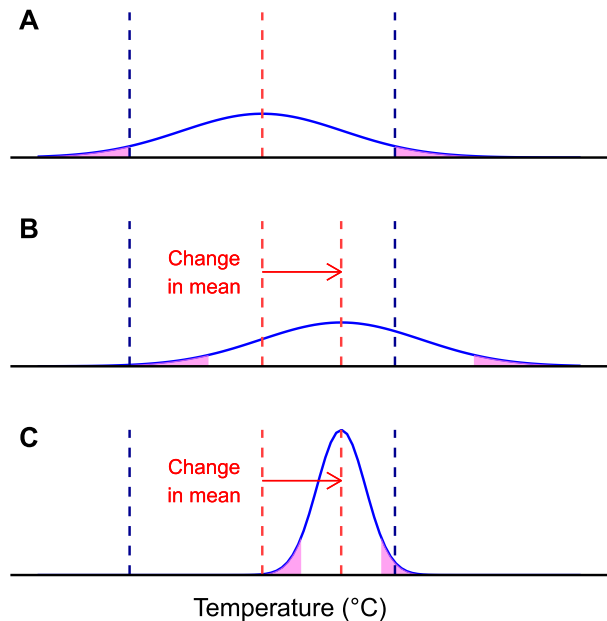FIG. 1. Map of Svalbard, with Svalbard Airport shown in red.

FIG. 2. Marginal temperature distributions for (a) a specific day in a reference period and (b),(c) a distant future after a climatic change. From (a) to (b), the mean increases, and from (b) to (c) the variance is decreased. The pink-shaded areas are the distributional extreme regions, defined by quantiles in the new climate marginal distribution. The climatic extremes are defined by being outside the interval indicated by the dashed blue vertical lines, given by quantiles in the reference distribution in (a).

climate is heading. However, the authors do not present a trend for temperature volatility. Screen (2014) finds that subseasonal cold-season temperature variability has significantly decreased over the last decades in the mid- to high-latitude Northern Hemisphere. He claims this is partly due to that northerly winds and associated cold days are warming more rapidly than southerly winds and warm days. This decrease in variability is consistent with our findings on Svalbard. On the global scale, Huntingford et al. (2013) find that the time-evolving standard deviation of globally averaged temperatures is stable.

Variance is a distributional characteristic that is highly relevant for occurrence of extremes, and here we distinguish between climatic and distributional extremes. Climatic extremes are temperature observations exceeding a quantile in a reference distribution. Distributional extremes are extremes with respect to the marginal distribution of the temperature on a given day, irrespective of any effect of global warming. A distributional extreme is not necessarily a climatic extreme and vice versa. The mean temperature is increasing globally, and a distributional extreme is a large deviation from this changing mean. We have illustrated these concepts in Fig. 2, where a distributional extreme is an observation in the pink shaded area and a climatic extreme is outside the vertical blue lines. Figure 2a illustrates the reference distribution. When only the mean temperature increases, the probability of upper tail climatic extremes increases substantially, while the distributional extreme probability remains unchanged (Fig. 2b). If the variance decreases at the same time, what is defined as a distributional extreme becomes less extreme, while the lower-tail probability of climatic extremes will

decrease further (Fig. 2c; relative to Fig. 2b). Although the figure is somewhat exaggerated to illustrate the point, we find that on Svalbard the mean temperature has increased and, for the summer, the variance has not changed (Fig. 2b). For the rest of the year, the variance has decreased (Fig. 2c). This will have consequences for the occurrence of climatic extremes and the magnitude of distributional extremes for the temperature. We investigate this on Svalbard in terms of conditional and unconditional variance.

Conditional variance is the variance conditioned on past information, while the unconditional variance is based on the marginal distribution of the variables and depends only on the day of the year. Typically, the immediate past history will be of highest influence on the conditional volatility. If considering a series of independent observations, the two variability concepts are equal, but this is rarely the case for time series.

A series of uncorrelated variables can be dependent with significant correlation when transformed to a squared process. This indicates conditional heteroskedasticity and an *autoregressive conditional heteroskedasticity* (ARCH; Engle 1982) or *generalized* ARCH (GARCH; Bollerslev 1986) model may be suited. These models are primarily used in fields like finance and

econometrics, where modeling of, for example, stock returns is of interest. Stock returns are known for being leptokurtic with time-dependent volatility and extreme returns appear in clusters. These are some characteristics of a GARCH model. In temperature time series there are seasonal components in both the conditional and unconditional variance, and this is not necessarily captured by standard GARCH models. We use a GARCH model with a time-dependent trend, where the slope parameter follows a cosine curve, and with harmonic functions to describe the general seasonal component. Except for the trend, our volatility model is similar to that of Campbell and Diebold (2005). There are certainly other ways of modeling development in volatility, but a GARCH-type model preserves the uncorrelatedness of the residuals, while modeling the second-order behavior of the squared process in a linear way. Dupuis (2012, 2014) uses a deterministic seasonal volatility and a stationary exponential GARCH model, respectively, in preprocessing temperature series before using extreme value methods on the residuals. She studies daily extreme temperatures and removes trend and season in the temperature level, as we do, though in a somewhat different way. By combining tail properties of the residual distribution with climate change, Dupuis (2012) obtains an increased probability of climatic extremes for four U.S. cities. Others have also applied GARCH models on temperature time series (e.g., Tol 1996; Taylor and Buizza 2004). Although our models can be used for prediction, the main goal here is to quantify the rate of change in the daily variance.

We describe in section 2 the Svalbard temperature time series data, and the method is presented in section 3. The estimated models are discussed in section 4, and we make some concluding remarks in section 5.

2. Data description

The data are collected, stored, and maintained by the Norwegian Meteorological Institute (MET Norway), and can be downloaded (http://eklima.met.no). Nordli et al. (2014) describe the daily average surface air temperature time series at Svalbard Airport, which is located near the outer part of Adventfjorden (see Fig. 1). The station, using an MI-33 screen, recorded its first temperature in August 1975 and is still operating. From 1 January 2005, the daily mean is calculated as the average temperature of every hour of the day. For observations prior to this date, Köppen's formula (Köppen 1888) is used, that is,

$$\text{daily mean} = \overline{T} - k(\overline{T} - T_{\min}),$$

where $\overline{T}$ is the average of the temperatures measured at 0600, 1200, and 1800 UTC, $T_{\min}$ is the minimum

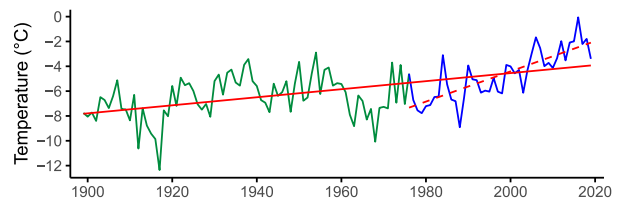

FIG. 3. Annual means of the full composite series of Nordli et al. (2014) from 1899 to 2019. The blue part of the curve is the years of available daily observations (from 1976 onward). The red solid line is the linear regression line of the full series, and the red dashed line is the linear regression for the years following 1976. The estimated trends are respectively 3.2° and 12.2°C per century.

temperature of the day, and k is a factor depending on the month and location of the station. On 5 October 2010, the screen was changed to an MI-74 and relocated to a site farther away from the runway of the airport to prevent thermal influence. Parallel measurements of the old and new site were performed from 27 October 2010 to 8 November 2011, and, according to Nordli et al. (2014), all monthly mean differences between the two sites during this period fall in the interval $[-0.09°C, 0.06°C]$. Therefore, they conclude that the temperature series is homogeneous through the relocation and screen type shift of October 2010. MET Norway has chosen Svalbard Airport as the only Reference Climate Series on Spitsbergen, the largest island of Svalbard.

A Svalbard Airport composite series of monthly mean temperatures back to September 1898 was reconstructed by Nordli et al. (2014), adding to the composite series of Nordli (2010), and is available online (http://eklima.met.no). The reconstruction is performed by predicting the temperature at Svalbard Airport from measurements at other stations for the period 1911–75. For the period 1898–1911, the authors use observations from hunting and research expeditions. To estimate transfer functions between the temperature at Svalbard Airport and the expedition locations, temporary stations were placed at these historical sites from 2010 to 2012, as part of the Arctic Climate and Environment of the Nordic Seas and the Svalbard–Greenland Area (AWAKE) project (Nordli et al. 2014). Annual means of this series are plotted in Fig. 3 with a simple linear regression line with slope 3.2°C per century. Nordli et al. (2014) report a yearly trend of 2.6°C per century for the period prior to 2012, so including the years following have increased the trend. In our analysis, we use daily observations, which are only available from August 1975. As can be seen in Fig. 3, this period, if considered to be isolated, has a much steeper linear trend of 12.2°C per century, relative to the full series. Similarly, Hanssen-Bauer et al. (2019) find a linear trend of 10.1°C per century for the period 1971–2017 at Svalbard Airport. We have chosen 1976–2019 because of availability
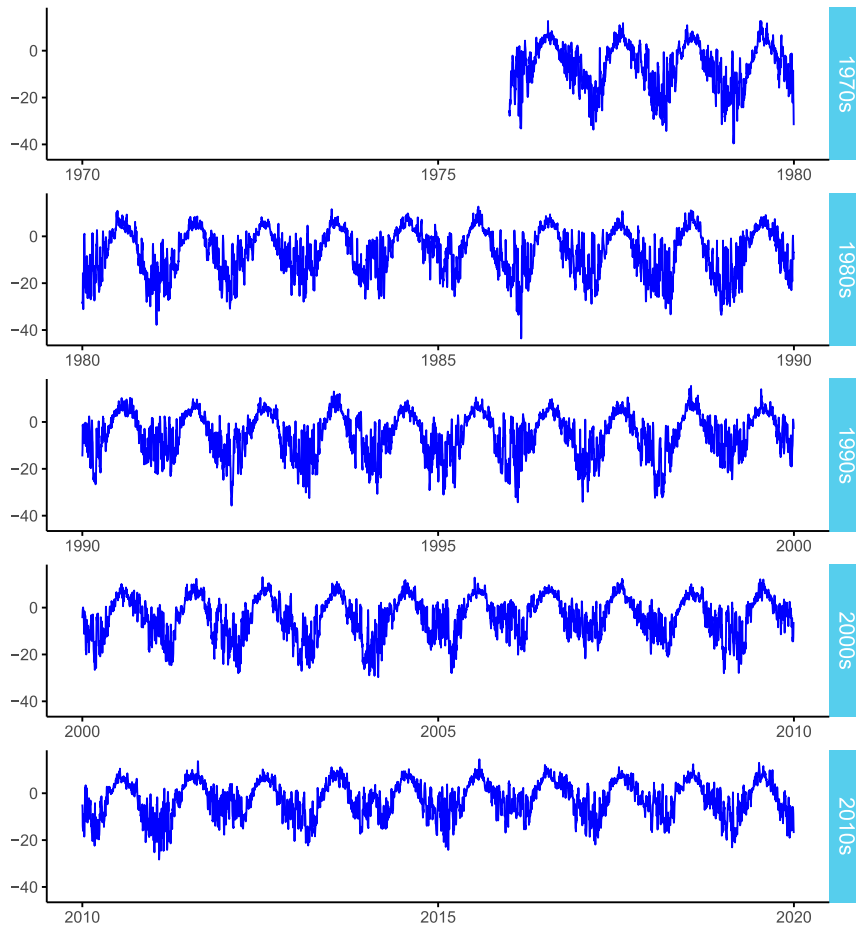
FIG. 4. Daily mean air temperatures at Svalbard Airport from 1976 to 2019, organized by decade.

of daily measurements, because modeling conditional volatility, as different from unconditional volatility, requires time dependence in data and many observations. Thus, data that are based on month or year are less relevant for such modeling because of lower dependence and fewer data. It is also on the daily scale that we are interested in making inference about the variability. A consequence of this is that we have picked a period during which the temperature has increased excessively.

The data contain no missing values, but MET Norway rates observations in terms of quality on a discrete scale from 0 to 7, where 0–1 is *OK* [acceptable], 2–4 is *slightly uncertain*, 5 is *very uncertain*, 6 is *very uncertain, based on model data*, and 7 is *erroneous*. In our data there are 0.3% of the observations with status 6 and 0.07% with status 5, but we treat them all equally. For simplicity, we have removed all observations on 29 February in leap years, so that every year has 365 days. The data are presented in Fig. 4 for the entire period where each panel is a decade. Notice that in Fig. 4, the yearly

maximum temperatures are fairly stable for the entire period, while the minimum temperatures seem much more volatile. The observations prior to 1 January 1976 (1 August–31 December 1975) are used for the regression model and the nonparametric volatility models (see section 3) but are discarded from the main analysis. We also remove observations after 31 December 2019, giving a total of 16 060 observations over 44 years.

3. Method

Let Y_t denote the daily average temperature at time t. Our primary interest is on the volatility of the process and we therefore split our model into two parts, a regression model and a volatility model. The residuals of the regression model are input to the volatility model. We consider the regression model primarily as a detrending and deseasonalizing step to get a white noise process with time-dependent variance. For the seasonal effects of the models, let $\psi_t = \psi_t^k(\eta, \zeta)$ denote a finite

TABLE 1. Parameter estimates and standard errors (SDE) for the regression model and the parametric volatility models. The standard errors are the square root, diagonal elements of the inverted Hessian matrix. An asterisk indicates that $0.01 < p$ value < 0.05; otherwise the p value < 0.001.

| Regression Model | | | Volatility Model | | | | | |
| | | | Nonstochastic | | | GARCH | | |
Parameter	Estimate	SDE	Parameter	Estimate	SDE	Parameter	Estimate	SDE
ϕ_1	0.553	0.089	ν	9.319	0.646	ν	10.470	0.805
ϕ_2	0.954	0.025	ω	11.207	0.206	ω	11.033	0.323
ϕ_3	-0.556	0.071	κ_A	-0.096	0.011	κ_A	-0.084	0.015
θ_1	0.381	0.088	c_1	10.053	0.240	α	0.060	0.006
θ_2	-0.836	0.091	d_1	5.399	0.243	β	0.845	0.017
θ_3	-0.234	0.011	c_2	-0.401^*	0.194	c_1	10.473	0.320
μ	-7.325	0.310	d_2	2.201	0.218	d_1	3.615	0.462
γ	0.121	0.012	c_3	-1.898	0.122	d_2	1.902	0.253
a_1	-8.550	0.217	d_3	-0.567	0.120	c_3	-2.277	0.144
b_1	-4.977	0.217						
a_2	1.876	0.203						
b_2	0.951	0.203						

Fourier series of order k with parameters $\eta = \{\eta_{1:k}\}$ and $\zeta = \{\zeta_{1:k}\}$; that is,

$$\psi_t = \psi_t^k(\eta, \zeta) = \sum_{j=1}^k \eta_j \cos(2\pi j t/\tau) + \sum_{j=1}^k \zeta_j \sin(2\pi j t/\tau),$$

where $\tau = 365$ is the period.

a. Regression model

The regression model is an autoregressive model with exogenous covariates, given by

$$Y_t = \hat{Y}_t + X_t, \quad \mu_t = \mu + \gamma t \tau^{-1} + \psi_t \quad \text{and}$$

$$\hat{Y}_t = \mu_t + \sum_{j=1}^p \phi_j(Y_{t-j} - \mu_{t-j}) + \sum_{j=1}^q \theta_j X_{t-j}, \quad (1)$$

where t is the day index, $\psi_t = \psi_t^r(a, b), a = \{a_{1:r}\}, b = \{b_{1:r}\}$, X_t is the residual process, p, q, and r are the orders of the autoregressive, moving average, and Fourier parts of the model, respectively, and $(\{\phi_{1:p}\}, \{\theta_{1:q}\}, \mu, \gamma, a, b)$ is the parameter vector. The mean model μ_t consists of a linear trend parameter γ, measured in degrees Celsius per year, and a finite Fourier series ψ_t that captures seasonal effects, and the one-step predictor $\hat{Y}_t$ also includes an autoregressive moving-average (ARMA) part for short-range linear correlation. In meteorology, anomalies are usually deviations from a normal temperature. The residual process $\{X_t\}$ is also an anomaly series, but this is relative to the regression model. We are mainly interested in the volatility of $\{X_t\}$. Two different parametric approaches to modeling the conditional variance are considered. In addition, we use two nonparametric methods as benchmarks for the parametric models.

b. Parametric volatility

We assume that $X_t = h_t^{1/2} Z_t$, where $\{Z_t\}$ is the iid innovation series following a standardized t distribution with ν degrees of freedom and $\{h_t\}$ is the conditional variance and the term of interest here. The first model is a first-order time-dependent GARCH model for the volatility; that is,

$$h_t = \sigma_t^2(1 - \alpha - \beta) + \alpha X_{t-1}^2 + \beta h_{t-1},$$
$$\sigma_t^2 = \omega + \kappa_t t \tau^{-1} + \psi_t, \quad \text{and}$$
$$\kappa_t = \kappa_A + 2^{-1}(\kappa_B - \kappa_A)\{1 - \cos[2\pi(t - t_0)/\tau]\}, \quad (2)$$

where t_0 is a hyperparameter and $\psi_t = \psi_t^s(c, d)$. We assume $\alpha + \beta < 1$, which guarantees stability of the stochastic part. In (2), the parameter vector is $(\nu, \omega, \kappa_A, \kappa_B, \alpha, \beta, c, d)$, s is the order of the finite Fourier series, and $c = \{c_{1:s}\}$ and $d = \{d_{1:s}\}$ are seen as nuisance parameters. *The climate change of volatility κ_t is parameterized to allow for different rates of change throughout the year* by assuming a cosine curve. We estimate the minimum and maximum of the curve, κ_A and κ_B, respectively, and require these to be one-half year apart. This specific formulation is partly inspired by inspecting the empirical data. Note that the scale of κ_t is degrees Celsius squared per year and that σ_t^2 is not the variance of X_t, but $\mathbb{E} h_t$ is.

The second parametric model is a special case of (2) with $\alpha = \beta = 0$ and

$$h_t = \sigma_t^2. \quad (3)$$

Under this model, h_t will also be the unconditional variance, $\mathbb{E} h_t = \sigma_t^2$, since σ_t^2 is not stochastic. Estimation
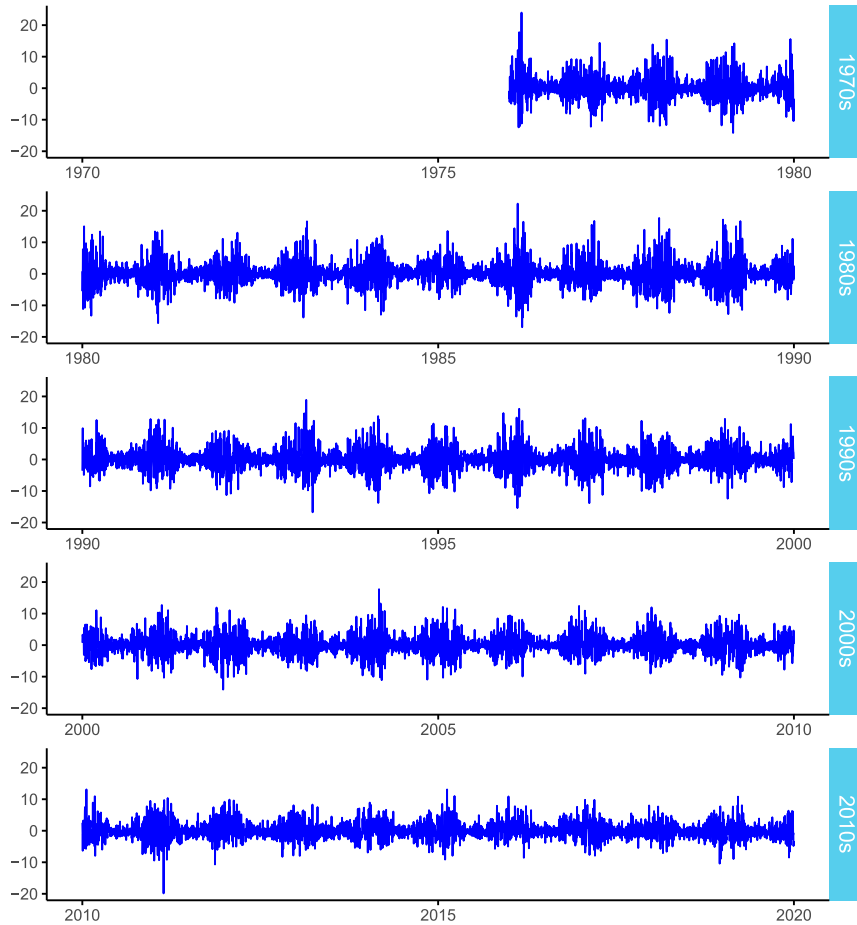
FIG. 5. The regression model residual series $\{X_t\}$ by decade.

and model selection of (3) is done independently of (2), although we use the same parameter symbols.

In (2), notice that the term $\sigma_t^2 \varphi$, with $\varphi = 1 - \alpha - \beta$, corresponds to the constant term in a standard GARCH(1, 1) model. This parameterization is used to be able to compare the climate change parameters of (2) and (3). For the model (2), we get by iterating

$$h_t = \varphi\sigma_t^2 + \alpha X_{t-1}^2 + \beta h_{t-1} = \varphi\sigma_t^2 + (\alpha Z_{t-1}^2 + \beta)h_{t-1}$$
$$= \varphi\sum_{j=0}^{\infty}\sigma_{t-j}^2\prod_{k=1}^{j}(\alpha Z_{t-k}^2 + \beta).$$

The unconditional variance is given by

$$\mathbb{E}\,h_t = \varphi\sum_{j=0}^{\infty}\sigma_{t-j}^2\prod_{k=1}^{j}\mathbb{E}\,(\alpha Z_{t-k}^2 + \beta) = \varphi\sum_{j=0}^{\infty}(\alpha + \beta)^j\sigma_{t-j}^2,$$

by the independence of $\{Z_t\}$. That is, the unconditional variance is a weighted average of the deterministic part of (2). We have that

$$\mathbb{E}\,h_{\tau+t} - \mathbb{E}\,h_t = \varphi\kappa_t\sum_{j=0}^{\infty}(\alpha + \beta)^j = \kappa_t,$$

where we have exploited that $\sigma_{t+\tau}^2 - \sigma_t^2 = \kappa_t$ as a result of periodicity of ψ_t^s and κ_t. Thus, the yearly change in unconditional variance of both (2) and (3) is κ_t, and κ_t can be compared across (2) and (3). The same holds for the intercepts.

c. Nonparametric volatility

For the nonparametric, we apply a *moving variance* (MVAR) with window width m,

$$\hat{h}_t = (m-1)^{-1}\sum_{j=1}^{m}X_{t-j}^2, \qquad (4)$$

and an exponentially weighted moving average (EWMA) model, given by

$$\hat{h}_t = \frac{1-\lambda}{1-\lambda^{t-1}}\sum_{k=0}^{t-2}\lambda^k X_{t-k-1}^2, \quad \lambda = e^{-1/\xi}, \ \ t \geq 2. \quad (5)$$
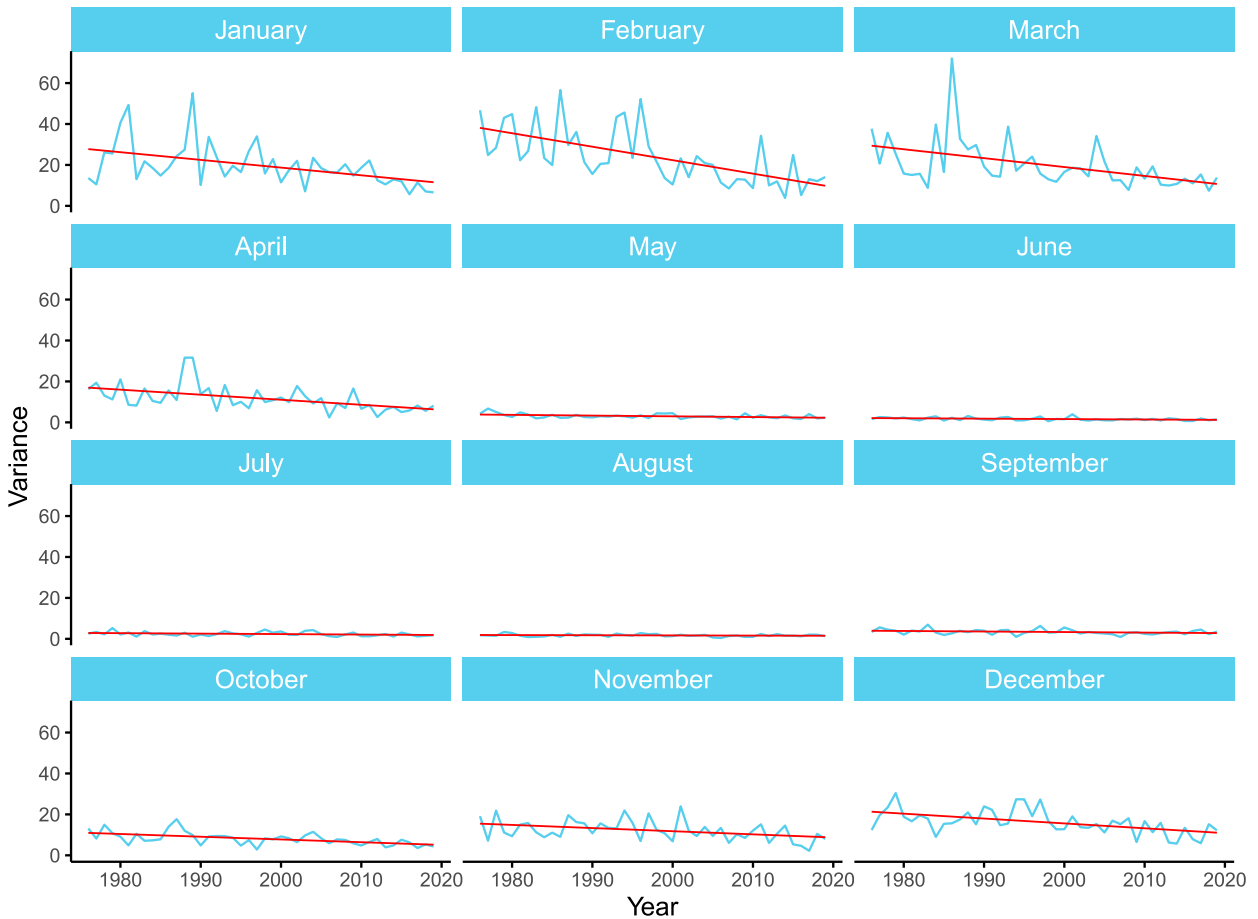
FIG. 6. Empirical variance of $\{X_t\}$ by year and month. The red lines are linear regressions based on the blue observations. At a 5% significance level, the negative trends for July, August, and September are not statistically significant.

Here, $\xi > 0$ is a smoothing parameter and, for $t = 1$, $\hat{h}_1 = X_1^2$. We refer to (2)–(5) as GARCH, nonstochastic, MVAR, and EWMA, respectively.

d. Parameter estimation and model selection

The deterministic part of the regression model (1) is estimated by linear regression, whereupon the errors are modeled as an ARMA(p, q) by maximum likelihood with conditional least squares for initial estimates (Brockwell and Davis 1991, 256–258). Estimation of the parametric volatility models is done by maximizing a conditional Student's t likelihood. For the nonstochastic model the estimation is carried out under the restriction $\alpha = \beta = 0$. We use the MVAR value for 1 January 1976 as initial value.

Because of the trend in the GARCH model, the model is nonstationary if $\kappa_t \neq 0$. If $\kappa_t \equiv 0$, it is periodically stationary. With a negative trend we have the issue that when $t \to \infty$ the variance will become negative. Hence, the model cannot be extrapolated infinitely, and asymptotic theory here has an issue. Nevertheless, it is

possible to extend the model in such a way that asymptotic arguments can be exploited.

Model selection is done by backward stepwise selection, that is, by first including many terms and then gradually removing the insignificant ones. Following this procedure, we select the model with the lowest Akaike information criterion (AIC; Brockwell and Davis 1991, 302–306).
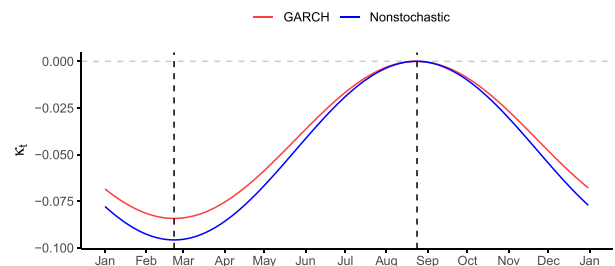

FIG. 7. Fitted $\kappa_t = \kappa_A 2^{-1}\{1 + \cos[2\pi(t - h)/\tau]\}$: The vertical lines indicate the minimum and maximum of κ_t, respectively, at 22 February and 24 August.
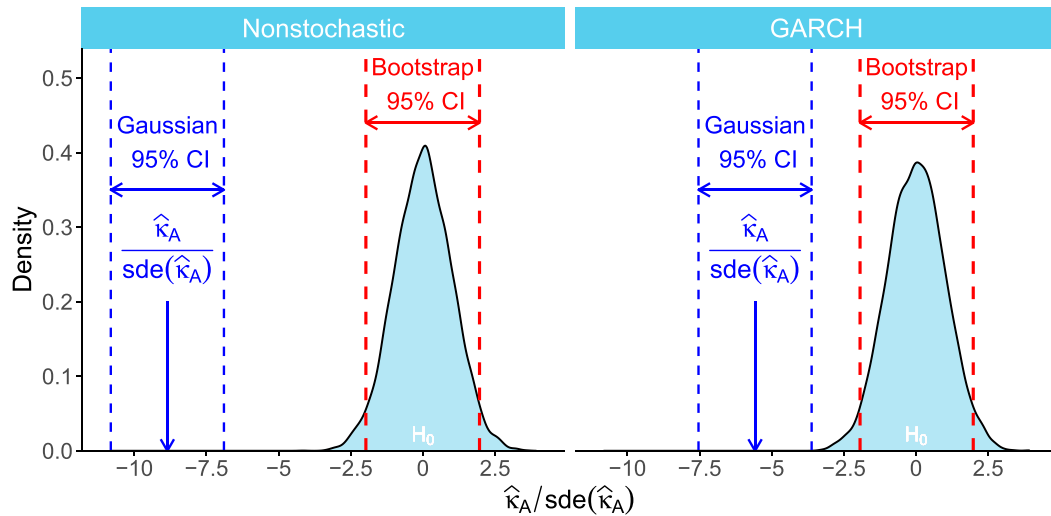
FIG. 8. Bootstrapped distribution of both of the standardized κ_A estimators under H_0: $\kappa_A = 0$. The vertical lines are the point estimates from Table 1 (vertical solid blue arrows) with their respective 95% Gaussian confidence interval (dashed blue lines) and the 95% confidence intervals of the bootstrapped null distribution (dashed red lines).

For the nonparametric we must choose a smoothing parameter ξ for EWMA and a window width m for the MVAR. These are considered as hyperparameters. The smoothing parameter ξ is chosen by minimizing the mean square error (MSE); that is,

$$\text{MSE}(\xi) = n^{-1} \sum_{t=1}^{n} \left[\frac{X_t^2}{\hat{h}_t(\xi)} - 1 \right]^2. \quad (6)$$

The window width m is set to capture seasonal variation and to get a relatively smooth estimate. The former is an argument for a small window, and the latter speaks to a large one. The selection is based on trial and error with these criteria in mind.

4. Results

The estimation results for the parametric approaches are presented in Table 1, where the selected orders are $p = q = 3$ and $r = 2$ for the regression model and $s = 3$ for both volatility models. In the model selection, we found that κ_B in both models and c_2 and d_3 in the GARCH model were not significantly different from zero and are thus fixed to zero. All estimates in Table 1 are significant with p values smaller than 10^{-3}, except for c_2 in the nonstochastic model with a p value of 0.038. These p values are based on a Gaussian approximation. The regression model is not of particular interest here, but notice that the trend parameter γ is estimated to 0.121°C per day, which corresponds to a temperature

increase of 12.1°C per century. This is consistent with the yearly aggregated trend displayed in Fig. 3. The resulting residuals $\{X_t\}$ are plotted in Fig. 5.

We have parameterized κ_t to take into account that changes are not homogeneous for all seasons. To see this effect on $\{X_t\}$, we have plotted the empirical variance for every month and year in Fig. 6. Here, the steepest decline in variance is found in February, and we have therefore set the hyperparameter $t_0 = 53$, which implies that 22 February is the day of largest yearly negative change. Other days in February were also tested, but the 22nd provided the lowest AIC among the candidates. At this day, the nonstochastic and GARCH models estimate changes of $\kappa_A = -0.096°$ and $-0.084\,°\text{C}^2\,\text{yr}^{-1}$, respectively. That κ_B is zero corresponds to not having a trend in August, which fits well with the linear trends in Fig. 6 being nonsignificant for the summer months. Figure 7 shows the fitted curve of κ_t for every day of the year. One pitfall of this simple cosine is that it assumes symmetry in the fall and spring. Although zero trend during summer seems pleasing, the somewhat restrictive cosine does not appear to go deep enough during winter, compared to the empirical rates of change. However, it is more robust, and it serves its purpose in terms of successfully separating summer and winter.

To ensure that the climate change parameter of volatility is significant without relying on a Gaussian approximation that is based on asymptotic arguments, we have bootstrapped the null distribution of $\hat{\kappa}_A$. That is, we have simulated the parametric models using the estimated parameters, except that $\kappa_A = 0$, 10 000 times and
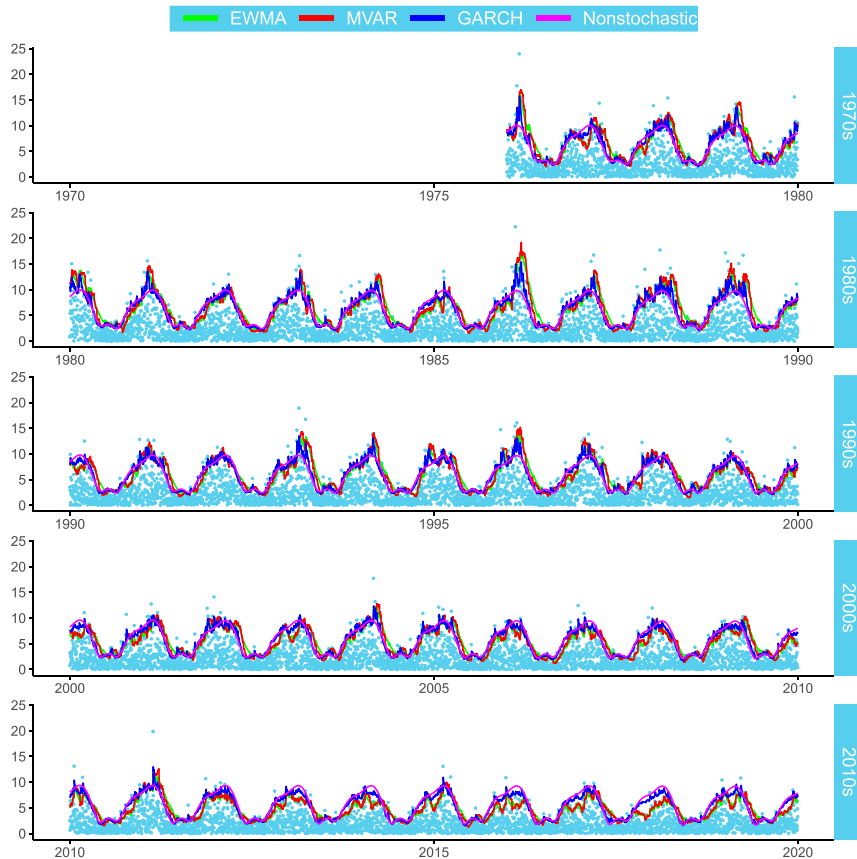
FIG. 9. Upper limit of 95% one-step prediction intervals based on the fitted t distributions and conditional variance estimates using the four methods, i.e., $\{t_v(0.975)\hat{h}_t^{1/2}\}$, with $t_v(0.975)$ being the 97.5% percentile of the standardized t_v distribution. The light-blue points are the absolute value of X_t, and ideally only 5% of them should be above the prediction band.

estimated the parameters. The nonparametric density estimates of $\hat{\kappa}_A/\text{SDE}(\hat{\kappa}_A)$, where SDE is the standard error, are presented in Fig. 8 with 95% bootstrapped confidence intervals under the null hypothesis, H_0: $\kappa_A = 0$. We have also included the point estimates for the original data (Table 1) with the respective Gaussian confidence intervals. As can be seen, the estimates are far away from the 95% H_0 intervals and the null hypotheses are clearly rejected in both cases.

In Fig. 9, we have plotted the different volatility estimates as 95% upper bound one-step prediction bands together with the absolute value of X_t. For the EWMA estimated process, we choose $\hat{\xi} = 21.01$ by minimizing the MSE in (6), and for the MVAR we let the window size $m = 30$. The bands in the figure are based on the fitted t_v distribution for the parametric, while for the MVAR and EWMA we fitted a t distribution to the standardized residuals using MLE with 7.89 and 7.43 degrees of freedom, respectively. Overall, the four different approaches behave similarly. The nonstochastic model, which implies that the residuals of model (1) are

independent, describes the data well. The GARCH model follows the nonstochastic closely for most of the time but is more adapted to the observations. The EWMA and 30-day MVAR are similar to each other. Their main purpose is to validate the parametric approaches, as they overall do. The nonstochastic estimate seems to be slightly too low during winter early in the series and too high toward the end, but this estimate is also more robust against outliers relative to the others. For $\{X_t\}$ in Fig. 5 and $\{|X_t|\}$ in Fig. 9, we clearly see a pattern of highly volatile winters and stable summers with a smooth transition between the seasons. The same pattern is also present in the fitted prediction bands of Fig. 9. The marginal coverage probabilities of the 95% standardized t_v-prediction intervals are 94.8% and 94.7% for nonstochastic and GARCH, and 93.7% and 93.9% for MVAR and EWMA, respectively.

To further evaluate the fitted models, we consider the standardized empirical residuals, $\hat{Z}_t = \hat{h}_t^{-1/2}X_t$ for $t = 1, \ldots, n$, where $\hat{h}_t$ is the fitted conditional volatility of either (2) or (3). They should be uncorrelated and
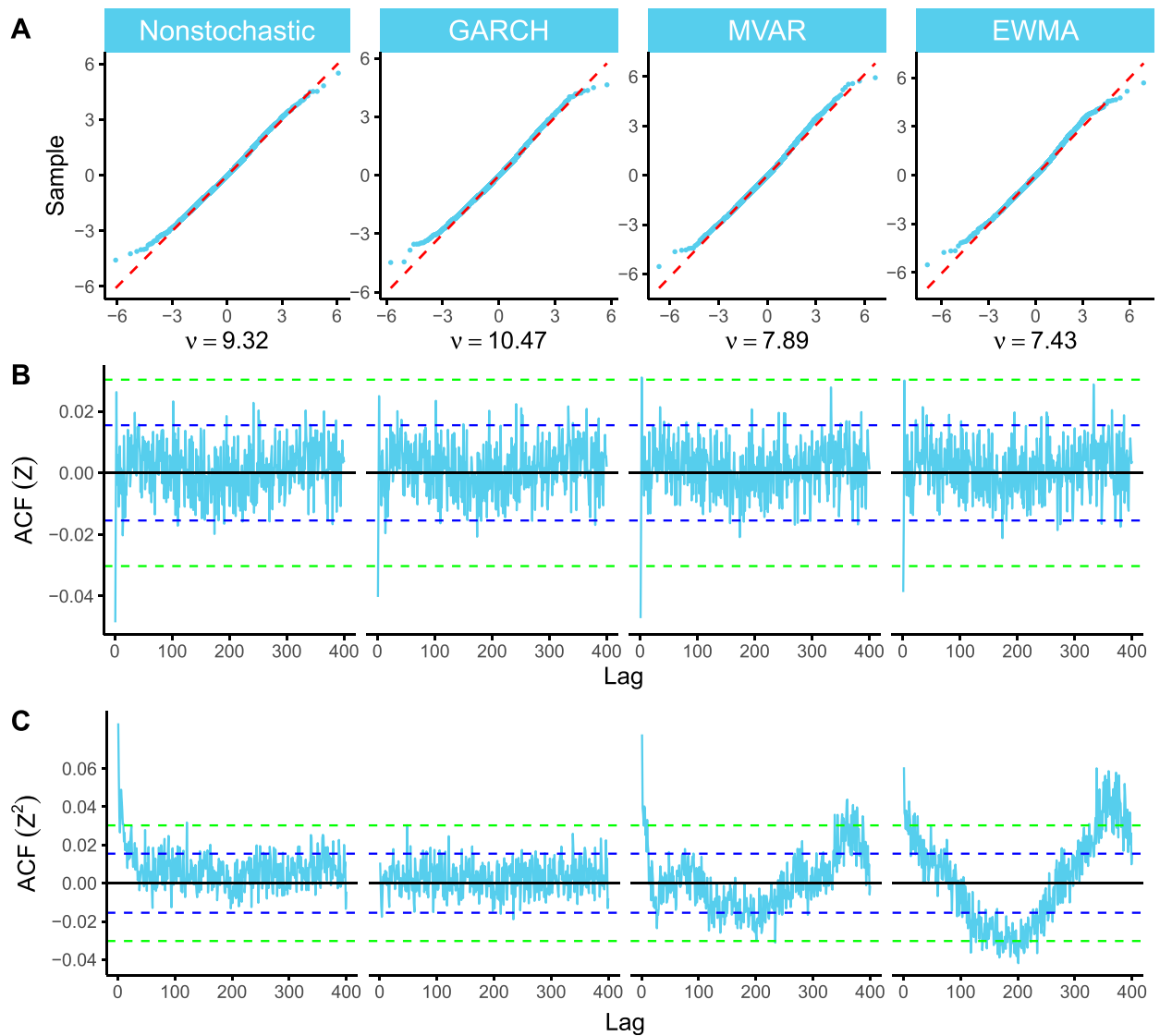
FIG. 10. (a) The QQ plots for each model residuals against the respective t_ν-distribution quantiles. For nonstochastic and GARCH models, ν is the estimated degrees of freedom from Table 1, and for MVAR and EWMA we have estimated a standardized t distribution to the empirical residuals using MLE. Also shown are sample autocorrelation functions for (b) $\hat{Z}_t$ and (c) $\hat{Z}_t^2$ up to lag 400 for every method. Bartlett's confidence intervals of 95% with (green dashed lines) and without (blue dashed lines) Bonferroni adjustment for multiple testing are also included.

standard Student's t distributed, with ν degrees of freedom. In Fig. 10a, quantile–quantile (QQ) plots for the model residuals against the respective t_ν quantiles are presented. The nonparametric benchmark residuals are also included, although they are not required to be t distributed. Sample autocorrelations of $\hat{Z}_t$ and $\hat{Z}_t^2$ with lags up to 400 for each of the methods are found in Figs. 10b and 10c. For all models, the quantiles of the standardized residuals are consistent with the corresponding t distribution, except for somewhat lighter tails. According to the observed autocorrelations of Z_t, none of the residuals are uncorrelated, but the

correlations are weak, with the strongest around -0.05. Since the period of the original data is 365, we include 400 lags in the plotted ACF, and in doing so we have a multiple testing problem. If we Bonferroni adjust (Dunn 1961) the significance level to account for multiple testing, we only get significant deviations from white noise for the first lag for the parametric and the first and second lags for the nonparametric. With such a high number of observations (16 060), we do not expect to keep the white noise hypothesis completely. Apart from the mentioned trace of correlation, we see a tendency of a nonadjusted periodicity in the standardized residuals from the sample
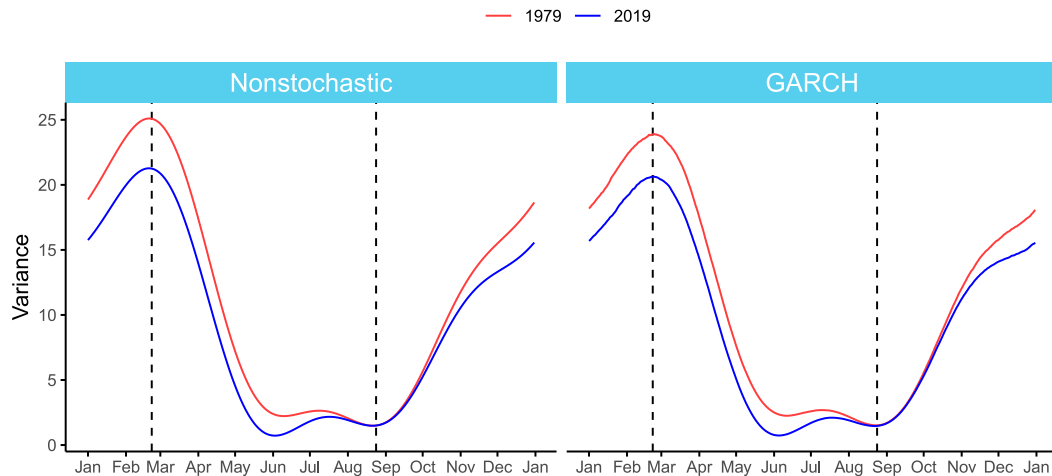
FIG. 11. Comparison of unconditional variances from the nonstochastic and GARCH models for 1979 and 2019. The GARCH model variance is a mean of 10 000 Monte Carlo simulations from the fitted model; 22 February and 24 August are indicated by vertical dashed lines.

autocorrelations. More interesting is the autocorrelation of the squared residuals that has more structure. For the nonstochastic, there are some notable correlations for the first lags, before stabilizing between the bands. The EWMA shows signs of an almost perfect periodic autocorrelation, while the periodic structure of the MVAR is more complex. In complete contrast, the GARCH model has no significant correlations for the squared residuals with Bonferroni adjusted 95% confidence intervals, but also just few with the standard Bartlett's. In total, we recommend the GARCH model based on the standardized residuals.

Distributional extremes are related to the marginal distribution of the temperatures, as discussed in the introduction. We therefore need the unconditional variance, which for the nonstochastic model is the same as the conditional. For the GARCH model, however, the unconditional variance is given by $\mathbb{E}\,h_t$. This expectation can be approximated by Monte Carlo simulation from the fitted model. That is, we simulate 10 000 realizations of the model and calculate the mean of h_t for every day. In Fig. 11 we present both the nonstochastic variance- and Monte Carlo–approximated GARCH variance estimates for each day of the years 1979 and 2019. As expected with the models we have used, we see that the 1979 variance curves are above the 2019 throughout winter and spring, whereas during late summer and early autumn the difference is small or zero. At the peak in February, the distance between the curves is consistent with $\kappa_A \times 40$ years ≈ -3.8 and $-3.7°C^2$, respectively. We also see here that the GARCH and nonstochastic models give consistent unconditional estimates, although the winter peak in variance of the GARCH is slightly lower than the nonstochastic. This is also a good figure

for seeing the seasonal aspects of the variability in temperature on Svalbard.

We have created density plot animations for each day of these years, and for every year on the days 22 February and 5 June, that are available as online supplemental material. In the former animation, one can clearly see the seasonal variations, and the latter visualizes the development, as the models describe it, over the years. The visual change in density from 1979 to 2019 during winter is minor, because the variance is high and thus the changes are relatively smaller. Around June the relative changes are larger, and the density plots reflect this result, whereas in August the change is zero. In terms of changes in the distribution of the temperature, the change in mean is by far the most important, but there is also an effect from the change in variance. When the mean increases and the variance decreases (e.g., in winter), it takes away more probability mass from the left than the right tail of the reference distribution (Fig. 2c) and together they contribute to decreasing the occurrence of really cold winter days.

5. Concluding remarks and discussion

In the specific case of Svalbard, we see a deviation from periodic stationarity in terms of a negative seasonally dependent trend κ_t in the volatility. The summer temperatures on Svalbard are increasing without any significant change in variance. For the rest of the year, they are both increasing and getting more stable in terms of a decreasing volatility. According to the models, the distributional extreme values are less extreme in the new climate, with summer being the exception. The left climatic extremes become less likely, both due to the

shift in mean and decreasing volatility, while for the right climatic extremes, the decreasing volatility contributes in reducing their magnitude. As mentioned earlier, Hanssen-Bauer et al. (2019; see also references therein) predict a warmer and increasingly wet climate on Svalbard, with more open, ice-free waters. Our contribution to the discussion about climate on Svalbard is that, except for the summer, the day-to-day variation of the temperature is on a decreasing trend. The decreasing volatility, that to some extent is a consequence of the sea ice retreat, could be conceived as a transition from a continental winter climate to a coastal winter climate. In this respect, κ_t quantifies the transition between these two different climate types. The decline can, however, not continue forever and must necessarily subside. If we were to extrapolate the parametric trend, the variance would at some point turn negative, which is impossible.

Such a trend appears to be of climatological interest and is connected to climate change. We are pleased that these relatively simple models manage to capture many characteristics of the raw data. The stable summer temperatures and volatile winters are apparent in the fitted prediction bands of Fig. 9 and the unconditional variances of Fig. 11. The parametric models give a strong negative trend in winter and weak in summer (Fig. 7), consistent with Fig. 6. However, the cosine curve is most likely too restrictive and further development of the models should allow for more advanced dynamics in κ_t.

It should also be possible to apply the method that we have used to other climatological series, but it requires a portion of high-quality daily-based data. It is an open question as to whether the climate change of volatility is nonzero at other locations.

Supplemental material

Animations of density plots based on the unconditional variances can be found in the online supplemental material for every day of the years 1979 and 2019 and for 22 February and 5 June in every year between 1976 and 2019.

REFERENCES

Bollerslev, T., 1986: Generalized autoregressive conditional heteroskedasticity. *J. Econ.*, **31**, 307–327, https://doi.org/10.1016/0304-4076(86)90063-1.

Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods.* 2nd ed. Springer Series in Statistics, Springer-Verlag, 580 pp.

Campbell, S. D., and F. X. Diebold, 2005: Weather forecasting for weather derivatives. *J. Amer. Stat. Assoc.*, **100**, 6–16, https://doi.org/10.1198/016214504000001051.

Dunn, O. J., 1961: Multiple comparisons among means. *J. Amer. Stat. Assoc.*, **56**, 52–64, https://doi.org/10.1080/01621459.1961.10482090.

Dupuis, D. J., 2012: Modeling waves of extreme temperature: The changing tails of four cities. *J. Amer. Stat. Assoc.*, **107**, 24–39, https://doi.org/10.1080/01621459.2011.643732.

——, 2014: A model for nighttime minimum temperatures. *J. Climate*, **27**, 7207–7229, https://doi.org/10.1175/JCLI-D-14-00101.1.

Engle, R. F., 1982: Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987–1007, https://doi.org/10.2307/1912773.

Hanssen-Bauer, I., and Coauthors, 2019: Climate in Svalbard 2100—A knowledge base for climate adaptation. NCCS Tech. Rep. 1, 105 pp., https://www.miljodirektoratet.no/publikasjoner/2019/februar-2019/climate-in-svalbard-2100–a-knowledge-base-for-climate-adaptation/www.miljodirektoratet.no.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox, 2013: No increase in global temperature variability despite changing regional patterns. *Nature*, **500**, 327–330, https://doi.org/10.1038/nature12310.

Isaksen, K., Ø. Nordli, E. J. Førland, E. Łupikasza, S. Eastwood, and T. Niedźwiedź, 2016: Recent warming on Spitsbergen—Influence of atmospheric circulation and sea ice cover. *J. Geophys. Res. Atmos.*, **121**, 11 913–11 931, https://doi.org/10.1002/2016JD025606.

Kohnemann, S. H., G. Heinemann, D. H. Bromwich, and O. Gutjahr, 2017: Extreme warming in the Kara Sea and Barents Sea during the winter period 2000–16. *J. Climate*, **30**, 8913–8927, https://doi.org/10.1175/JCLI-D-16-0693.1.

Köppen, W., 1888: Über die Ableitung wahrer Tagesmittel aus den Beobachtungen um 8h am, 2h pm und 8h pm (On the derivation of true daily means from the observations at 8 am, 2 pm, and 8 pm). *Ann. Hydrogr. Marit. Meteor.*, **16**, 341–354.

Nordli, Ø., 2010: The Svalbard Airport temperature series. *Bull. Geogr. Phys. Geogr. Ser.*, **3**, 5–25, https://doi.org/10.2478/bgeo-2010-0001.

——, R. Przybylak, A. E. Ogilvie, and K. Isaksen, 2014: Long-term temperature trends and variability on Spitsbergen: The extended Svalbard Airport temperature series, 1898–2012. *Polar Res.*, **33**, 21349, https://doi.org/10.3402/polar.v33.21349.

Onarheim, I. H., L. H. Smedsrud, R. B. Ingvaldsen, and F. Nilsen, 2014: Loss of sea ice during winter north of Svalbard. *Tellus*, **66A**, 23–33, https://doi.org/10.3402/tellusa.v66.23933.

Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nat. Climate Change*, **4**, 577–582, https://doi.org/10.1038/nclimate2268.

——, C. Deser, and L. Sun, 2015: Projected changes in regional climate extremes arising from Arctic sea ice loss. *Environ. Res. Lett.*, **10**, 084006, https://doi.org/10.1088/1748-9326/10/8/084006.

Taylor, J. W., and R. Buizza, 2004: A comparison of temperature density forecasts from GARCH and atmospheric models. *J. Forecasting*, **23**, 337–355, https://doi.org/10.1002/for.917.

Tol, R. S., 1996: Autoregressive conditional heteroscedasticity in daily temperature measurements. *Environmetrics*, **7**, 67–75, https://doi.org/10.1002/(SICI)1099-095X(199601)7:1<67::AID-ENV164>3.0.CO;2-D.