GEOPHYSICAL RESEARCH LETTERS, VOL. , NO. , PAGES 1–13,

On the estimation of exceedance over a threshold

Xuebin Zhang

Climate Research Branch, Meteorological Service of Canada, Downsview, Ontario, Canada

Gabriele Hegerl

Nicholas School for the Environment and Earth Sciences, Duke University, Durham NC

Francis W. Zwiers

Canadian Centre for Climate Modeling and Analysis, Victoria, British Columbia, Canada

Gabriele Hegerl

Nicholas School for the Environment and Earth Sciences, Duke University, Durham NC

Abstract. Climate extremes indices have been developed and maintained for the purpose of monitoring climate change. They are also potentially useful for climate change detection studies. A number of such indices have been defined by counting the number of days in a year/season that daily values have exceeded thresholds (defined by a daily percentile) calculated for every calendar day of the year from a base period such as 1961-1990. Using Monte Carlo simulation, we demonstrate that existing methods for the estimation of those indices produce artificial jumps in the indices series at the beginning and end of the base period. This would make the indices so estimated unsuitable for the monitoring and detection of climate change. We propose a bootstrap resampling procedure to estimate the indices for the base period. Our method effectively removes the inhomogeneity.

Gabriele Hegerl, Division of Earth and Ocean Sciences, Nicholas School for the Environment and Earth Sciences, Duke University, Durham NC 27708-90227, hegerl@duke.edu

Xuebin Zhang, Climate Research Branch, Meteorological Service of Canada, 4905 Dufferin Street, Downsview, Ontario, Canada, M3H 5T4 Xuebin.Zhang@ec.gc.ca

Francis W. Zwiers, Canadian Centre for Climate Modelling and Analysis, University of Victoria, P.O. Box 1700 STN CSC, Victoria, BC, Canada, V8W 2Y2 Francis.Zwiers@ec.gc.ca

1. Introduction

Successive reports of the Intergovernmental Panel on Climate Change (IPCC) have made increasingly strong statements on the human influence on the global climate. Since the impacts of climate change may be felt more from the change in the extremes, rather than in the mean, analyzing climate extremes becomes very important. Monitoring, detection and attribution of changes in climate extremes would require daily resolution data. However, compilation, provision, and update of a globally complete and readily available full resolution daily dataset are very difficult due to several reasons. Consequently, indicators of climate extremes have been developed (e.g. Karl et al. 1999, Peterson et al. 2001) in the hope that they would be more widely obtainable. These indicators have been used to analyze changes in climate extremes for various parts of the world (e.g. Jones et al. 1999, Frich et al. 2002, Klein Tank and Können 2003, Kiktev et al. 2003)

A group of indicators are defined by computing exceedance rate of daily values (such as daily mean temperature) over (or below) thresholds (usually a percentile of daily data) calculated for every calendar day in the year. For easy comparison of indices across stations with various temporal coverage, and for easy update once new daily data are available, the thresholds are usually computed from a common base period, say 1961-1990 for all stations. Folland et al. (1999) provisionally recommended a three-steps procedure for the estimation of the thresholds: 1) to remove annual cycle by extracting the 30-year mean values of each calendar day, 2) to fit a probability distribution (e.g. 3-parameters gamma distribution to daily temperature anomalies), and 3) to compute the thresholds from the fitted probability distribution. Data from additional proximate calendar days may be added to improve the stability of the parameters for the probability distribution, but those days should be far enough apart such that data from different days are effectively independent. This method has been successfully implemented in Jones et al. (1999) who use a 5-day moving window spaced by 5-days (referred to as 5SD window hereafter). In many other applications (e.g. Frich et al. 2002, Klen Tank and Können 2003, and Kiktev et al. 2003), thresholds have been estimated using data from a 5-consecutive-day moving (referred to as 5CD) window and a plotting position formula or empirical probability distribution (referred to as EPD hereafter).

Despite the importance of those indicators in the detection and monitoring climate change, their statistical properties have not been well documented. For example, what the difference of resulting indices series would be when 5CD and 5SD windows are used? The thresholds are “adapted” to (computed from) the base period, does this cause any systematic biases in both mean and variance in the exceedances between the in-base and out-base periods? Those are important statistical characteristics that need to be well understood before the indices can be properly used for the purpose of climate change detection and monitoring. The main objective of this note is to examine, through Monte-Carlo simulations, the characteristics of time series of exceedance rate computed using existing methods. The remaining of this paper is organized in the following. We describe methods in Section 2. Results are presented in Section 3. Conclusions and discussion follow in Section 4.

2. Methods

2.1. Monte-Carlo simulation

Given that homogeneous time series is essential for the monitoring and detecting climate change, and the thresholds are computed only from a portion of the data (usually a 30-

year base period), our Monte-Carlo simulation experiment was designed to unveil possible inhomogeneity of the exceedance time series between in-base and out-base periods, i.e., bias in the estimation of mean and variance of the exceedance rates for the two periods. Daily values are usually serially correlated, and this would make the effective sample size that influences the computation of variance of the data smaller than the actual sample size, and hence underestimate the thresholds. We use an AR(1) process to generate daily data values to assess this effect.

Without loss of generality, we assume that the mean and variance of simulated daily data equal to zero and one, respectively. Let X_t an AR(1) process with a first order autocorrelation α, the AR(1) process is defined by

$$X_t = \alpha X_{t-1} + Z_t, \tag{1}$$

where Z_t is a white noise with variance

$$Var(Z_t) = 1 - \alpha^2. \tag{2}$$

The values of α are typically between 0.6 and 0.8 in Canadian daily temperature data. Here, we prescribe five different values, 0.0, 0.2, 0.4, 0.6, 0.8, for α. This would enable us to show the impacts of different effective sample sizes. For each α value, 60-years of daily data are simulated, using equation (1). The first and second 30-years periods are assumed to be in-base and out-base periods, respectively. Exceedances are computed using the methods described in the next subsection. This procedure is repeated for 1000 times. We then compare the mean and variance of the the exceedance in the two 30-year periods for the 1000 simulation.

2.2. Exceedance computation

Before the exceedance is computed, the desired thresholds for each of all calendar days must have been estimated. There are two aspects to consider when estimating the thresholds. One is the data used for determining threshold. In this study, we use both 5CD and 5SD windows. For example, to estimate threshold for January 13, data for all January 11-15 in the base period are selected when 5CD window is used, but data for January 1, 7, 13, 19, 25 in the base period would be involved when 5SD window is used. One may argue that the latter approach, while taking the effective sample size into consideration, may also have discarded useful information from the data and thus may not have used the available information efficiently. For this reason, we also use all daily data available in the time window. That is, we use data from January 1 to January 25 to estimate the threshold for January 13. This approach is termed 25-consecutive-days moving window or 25CD window. The other is how to estimate threshold from the given data set. Here we use EPD. We also fit a Gaussian distribution (referred to as FGD hereafter) to the generated data, since they follow a Gaussian distribution by design.

The exceedance is calculated in the same manner as is described in both Jones et al. (1999) and Frich et al. (2002): The exceedance rate for a giving year, regardless if the year is inside or outside of the base period, is the number of days in the year that daily values have exceeded the estimated thresholds. Note that this seemly correct approach may actually result in discontinuity in the estimated exceedance between in and out base periods. The problem is that we are using observations to estimate a threshold (say the 10th percentile of the underlying probability distribution). But our estimator (no matter how it is obtained) will be affected by sampling variability, and thus the estimated threshold can never be exact. This means that the out-of-sample (out-base period) exceedance rate will never be exactly 10%. Moreover, the estimated threshold is "adapted" to the sample (base period) from which it was obtained: information in the sample was used to estimate the threshold, so the in-base exceedance rate will be different from the out-of-base rate. To overcome this problem, we purpose a bootstrap resampling procedure to determine exceedance rate for the in-base period. 1) The 30-year base period is divided into one "out-base" year, the year for which exceedance is to be estimated, and a "base-period" consisting the remaining of 29 years from which the thresholds would be estimated. 2) A 30-year block of data is constructed by using the 29 years "base-period" data along with an additional year of data from the "base-period" (i.e., one of the years in the "base-period" is repeated). It is used to estimate thresholds. 3) The "out-base" year data are compared with these thresholds and exceedance for the "out-base" year is obtained. 4) Repeat procedures 2) and 3) for additional 28 times, by using data from different years to construct the 30-year block. 5) The exceedance for the "out-base" year is the average of the 29 estimates obtained from steps 2) to 4). In this way, the year for which exceedance is to be estimated is not used for estimating the thresholds. This effectively makes the estimation of exceedance for both in-base and out-base periods comparable as we shall see later, greatly reducing the discontinuity.

3. Results

To simplify the presentation of the results, we summarize the outcomes corresponding to the use of 5CD window and EPD under various settings of simulations (e.g. different autocorrelations) in one subsection. After that, we present results obtained by using different moving windows, different methods of computing thresholds in another subsection.

In the latter case, we only provide results corresponding to $\alpha = 0.8$ and $Z_t = 0.0$ which are more likely the case for the real data.

3.1. 5-consecutive-day moving window

Figure 1 displays relative bias in the exceedance rate estimated by using 5CD window and EPD. The bias is calculated as the difference between the nominal rate (nominal rate is 10% when 90th percentile is used as the thresholds) and average exceedance rate in 1000 simulations and then divided by the nominal rate. The biases for $\alpha = 0.0, 0.2, 0.4, 0.6, 0.8$ in equation (1) are shown. For the out-base period, there is almost no bias when $\alpha = 0$, but there is a positive bias when $\alpha > 0$, and the larger the α, the bigger the bias. This indicates that the thresholds are underestimated when $\alpha > 0$, due to effective sample sizes smaller than actual sample size. The bias for the in-base period is very similar for different α. They are small when the expected total number of exceedance during the base period, the product of the nominal rate and the number of daily values being used for the estimation of threshold, is an integer number but negative if the number is an non-integer due to rounding error. The positive bias for the out-base period and tendency of negative bias for the in-base period would result in a sudden jump in the exceedance rate between the in-base and out-base periods. This jump would be bigger when higher percentiles are used to define extremes.

Figure 2 shows relative bias in the exceedance rates when α equals to 0.0, 0.4, and 0.8, respectively. Now the exceedance rates for the in-base period have been estimated using the bootstrap resampling procedure. Biases for both in-base and out-base periods are more or less the same. This suggests that resampling procedure could largely eliminate the jump in the exceedance rate between the two periods.

We also investigated the influence of seasonal cycle on the estimation of the exceedance rate. This time, we add a smooth annual cycle Y_t into the left side of equation (1). This annual cycle is the first harmonics of 30-year mean daily temperature at Ottawa Airport station. The upper panel of Fig. 3 displays relative bias in the exceedance rate, when $\alpha = 0.8$ and when the annual cycle is not removed prior to the computation of thresholds. The lower panel shows the bias when annual cycle is removed by subtracting 30-year mean value for every calendar day before the estimation of the thresholds. Clearly, removal of annual cycle makes the estimated exceedance rates more consistent for in-base and out-base periods.

The above analyses suggest that the existing method, that uses 5-consecutive day moving window along with empirical probability distribution for the computation of threshold, of exceedance estimation results in artificial jump in the the estimated rate between in-base and out-base periods. This is mainly caused by the underestimation of the thresholds due to effective sample size being smaller than actual sample size. Rounding error also contributes to the jump. Our resampling procedure effectively overcomes this problem.

3.2. Thresholds estimated by using different moving windows and different methods

The left panel of Fig. 4 shows biases in the exceedance rate when the thresholds are estimated by EPD or FGD with data from 5CD window. When thresholds are estimated by fitting a proper distribution to the data, negative bias for the in-base period caused by rounding error is almost eliminated. It would be useful to fit a distribution to the data if the form of a proper distribution is known. When exceedance rate for the in-base period is estimated using resampling procedure, its bias is very similar to that of out-base period, no matter if the thresholds are estimated by EPD or FGD. Thus, from the producing homogeneous exceedance rate point of view, the ways of which the thresholds are obtained are not important. It is in this sense that the use of EPD may be a preferred method because it is easier to use and also because we usually don't know the actual underlying probability distribution for the data. The middle panel of Fig. 4 displays biases in the exceedance rate when 5CD window is used. Two main features are important to note. 1) Bias for the in-base period remain, especially when EPD is used due to rounding error. 2) Bias for the out-base period, and that for in-base period with exceedance rate being computed by resampling procedure are much smaller than those when 5CD window was used, indicating the usefulness of using more information from the data. The right panel shows the results when 25CD window is used. Bias for the in-base period is much reduced compared with the 5CD and 5SD windows, due to much bigger sample size. Biases for out-base period and in-base period computed with resampling procedure are comparable with those from 5SD window. This suggests that the impact of smaller effective sample size on the estimated thresholds is smaller when sample size is large. Overall, it appears that the use of all available information within 25CD window provides better estimate than that when only partial information (5SD window) has been used. Additionally, the use of 25CD window would also result in much smoother estimate of the thresholds among different calendar days than those from 5SD window. This would be advantageous when thresholds are used to compute other indices such as the runs of extreme days (e.g. spell length of extreme events). Therefore, we recommend to use 25CD window. It should also be noted that the average of variance of exceedance rate for the out-base period is also greater than that for the in-base period (Fig. 5). Our resampling procedure yields in-base variance very close to that of out-base period. (*Francis: is the difference too small to mention?*)

4. Conclusions and Discussion

We have compared the performances of different methods in producing temporally homogeneous time series of exceedance rates. We use empirical probability distribution and also fit a Gaussian distribution to estimate thresholds from data covered by 5-consecutive-days moving window, 5-day moving window spaced by 5-day, and 25-consecutive-days moving window. Our comparison was made with the aid of Monte Carlo simulation experiments. We found that the time series of exceedance rate is discontinues at the boarder of in- and out-base periods if the rate is estimated using existing methods, due mainly to sampling uncertainty. Our bootstrap resampling procedure overcomes this problem and produces much homogeneous estimates of the exceedance rate across the two periods. The 5-consecutive-days moving window approach produces largest bias in the estimated exceedance rate because the effective sample size is smaller than actual sample size that impacted the estimate of variance which in turn results in underestimation of thresholds. The 5-days moving window spaced by 5 days that used in Jones et al. (1999) offers remarkable improvement. However, this method may not have used the available information as efficiently as a 25-consecutive-days moving window. In addition, the 25-days window yields much smoother estimation of threshold for the year, and thus could provide better estimate for the length of spells of extreme events. The difference in using empirical probability distribution or fitting a proper probability distribution to estimate the thresholds is very small. We therefore recommend the use of empirical probability distribution for its simplicity along with the 25-consecutive-days moving window to estimate thresholds. We also suggest that the in-base exceedance rate being estimated with bootstrap resampling procedure.

The inhomogeneity of exceedance time series estimated by existing methods could have profound impacts if those series are used for climate change monitoring. For example, we show in Fig. 6 the average exceedance rates of daily temperature greater than its 90th and 95th percentiles during 1961-1990 at 210 stations in Canada. The daily temperature data at those stations have been homogenized: step changes caused by change in station location and/or measurement programs have been minimized (Vincent et al. 2002). They have been used for analyzing trends in daily and extreme temperatures (Bonsal et al. 2001). We computed exceedance rates at each station with thresholds being estimated by using 5CD window and EPD. The rates are then averaged to obtain extreme indices for Canada. The dashed lines indicate base-period exceedance rates being computed by resampling procedure. Clearly, exceedance rate would have been underestimated if resampling procedure was not used. As a results, two artificial jumps would exist in the time series, one at the beginning and the other at the end of the base period. Note that the jumps are greater when higher percentile is used to define the threshold. Trend in the extreme indices would also be distorted if existing method is used to estimate in-base exceedance rate. The distortion would be greater if the base period is at the beginning or at the end of the time series.

More importantly, misleading conclusion could be reached if inhomogeneous indices series are used in climate change detection studies. The idea of climate change detection is to regress observed series to model simulated climate change signal series and to test if the regression coefficient is significantly great than zero. If extreme indices for both observed and model simulated data are computed using existing methods with the same base period, there would be artificial jumps in the series obtained from observed and model simulated data at the same time. This could become a main feature in the series that regression analysis would pick up. At the end, it is possible that the regression coefficient is significantly greater than zero even if climate change signal could not be detected otherwise.

Acknowledgments.

References

Bonsal, B.R., X. Zhang, L.A. Vincent, and W.D. Hogg, 2001: Characteristics of daily and extreme temperatures over Canada. *J. Climate*, **14**, 1959-1976.

Folland, C.K., C.Miller, D.Barder, M. Crowe, P. Jones, N. Plumner, M. Richman, D.E. Parker, J. Rogers, and P. Scholefield, 1999: Workshop on indices and indicators for climate extremes, Asheville, NC, USA, 3-6 June 1997, Breakout Group C: Temperature indices for climate extremes. *Climatic Change*, **42**, 31-43.

Frich, P., L.V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A.M.G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Research*, **19**, 193-212.

Jones, P.D., E.B. Horton, C.K. Folland, M. Hulme, D.E. Parker, and T.A. Basnett, 1999: The use of indices to identify changes in climatic extremes. *Climatic Change*, **42**, 131-149.

Karl, T.R., N. Nicholls, and A. Ghazi, 1999: CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes: Workshop summary. *Climatic Change*, **42**, 3-7.

Kiktev, D., D.M.H. Sexton, L. Alexander, and C.K. Folland, 2003: Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate*, **16**, 3560-3571.

Klein Tank, A.M.G., and co-authors, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment. *Interna. Climatato.*, **22**, 1441-1453.

Klein Tank, A.M.G., and G.P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946-99. *J. Climate*, **16**, 3665-3680.

Peterson, T.C., C. Folland, G. Gruza, W. Hogg, A. Mokssit, and N. Plummer, 2001: Report on the activities of the Working Groupo on Climate Change Detection and Related Rapporteurs 1998-2001. World Meteorological Organization Rep. WCDMP-47, WMO-TD 1071, Geneva, Switzerland, 143 pp.

von Storch, H., and. F.W. Zwiers, 1999: *Statistical Analysis in Climate Research*, Camberidge University Press, Cambridge, UK, 484 pp.

Vincent, L.A., X. Zhang, B.R. Bonsal, and W.D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322-1344.

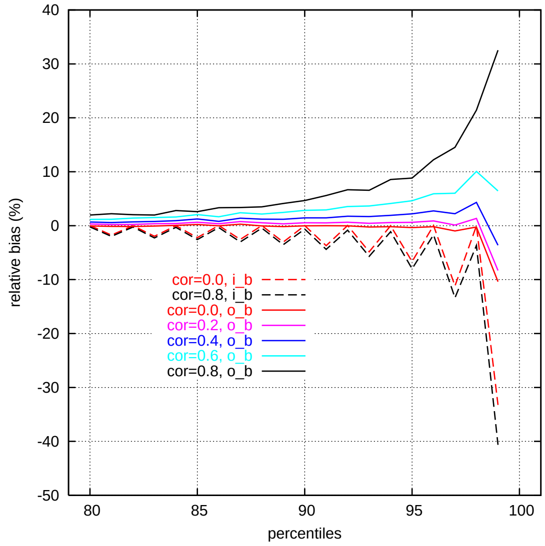

Fig. 1. Relative bias in the exceedance rate, when thresholds are estimated by using empirical probability distribution and data from 5-consecutive-days moving window, as a function of percentiles for in-base (i_b) and out-base (o_b) periods. cor=0.0, 0.2, ..., 0.8 indicate the first-order autocorrelation coefficients (α) of 0.0, 0.2, ... , 0.8 have been used, respectively.

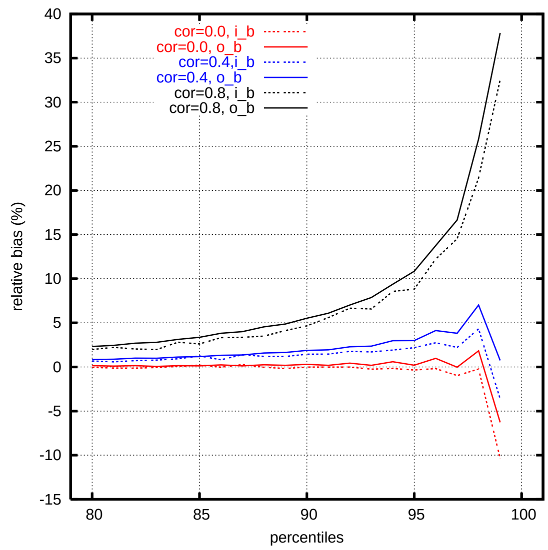

Fig. 2. Same as Figure 1, except the exceedance rates for in-base period are estimated by bootstrap resampling procedure.

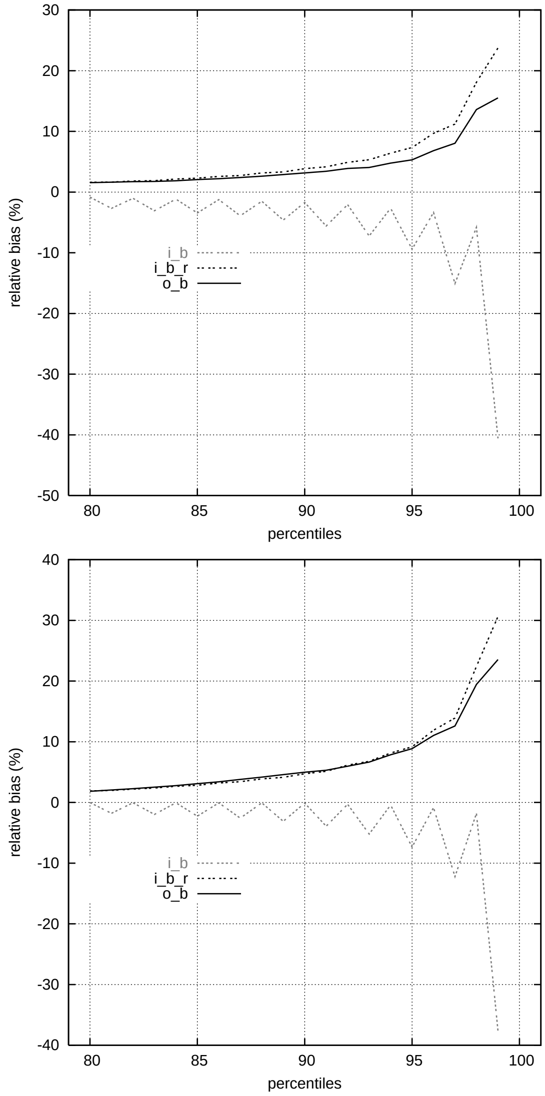

Fig. 3. Relative bias in the exceedance rate estimated by 5-consecutive-days moving window as a function of percentiles for in-base (i_b) and out-base (o_b) periods. The first-order autocorrelation coefficient $\alpha = 0.8$ has been used when simulating the data. i_b_r indicates the rate for the in-base period estimated by using bootstrap resampling procedure. Upper panel shows result when annual cycle is added to the data, while lower panel displays results when annual cycle is removed from the data.

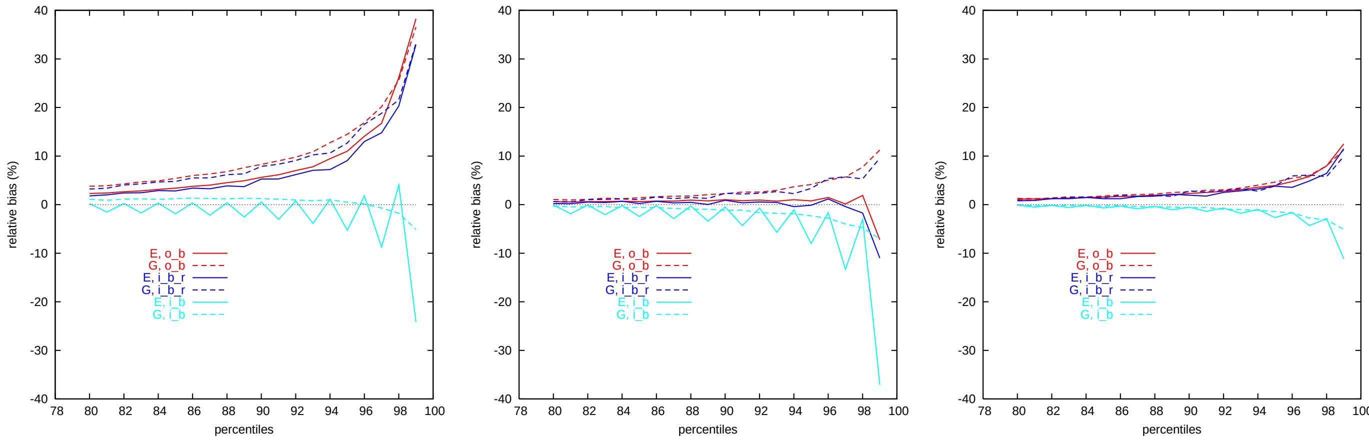

Fig. 4. Relative bias in the exceedance rate when the thresholds are estimated by using empirical probability distribution (E) or by fitting a Gaussian distribution (G) to the data from 5-consecutive-days moving window (left panel), 5-days moving window spaced by 5 days (middle panel), and 25-consecutive-days moving window (right panel) as a function of percentiles for in-base (i_b) and out-base (o_b) periods. The first-order autocorrelation coefficient $\alpha = 0.8$ has been used when simulating the data. i_b_r indicates the rate for the in-base period estimated by using bootstrap resampling procedure.

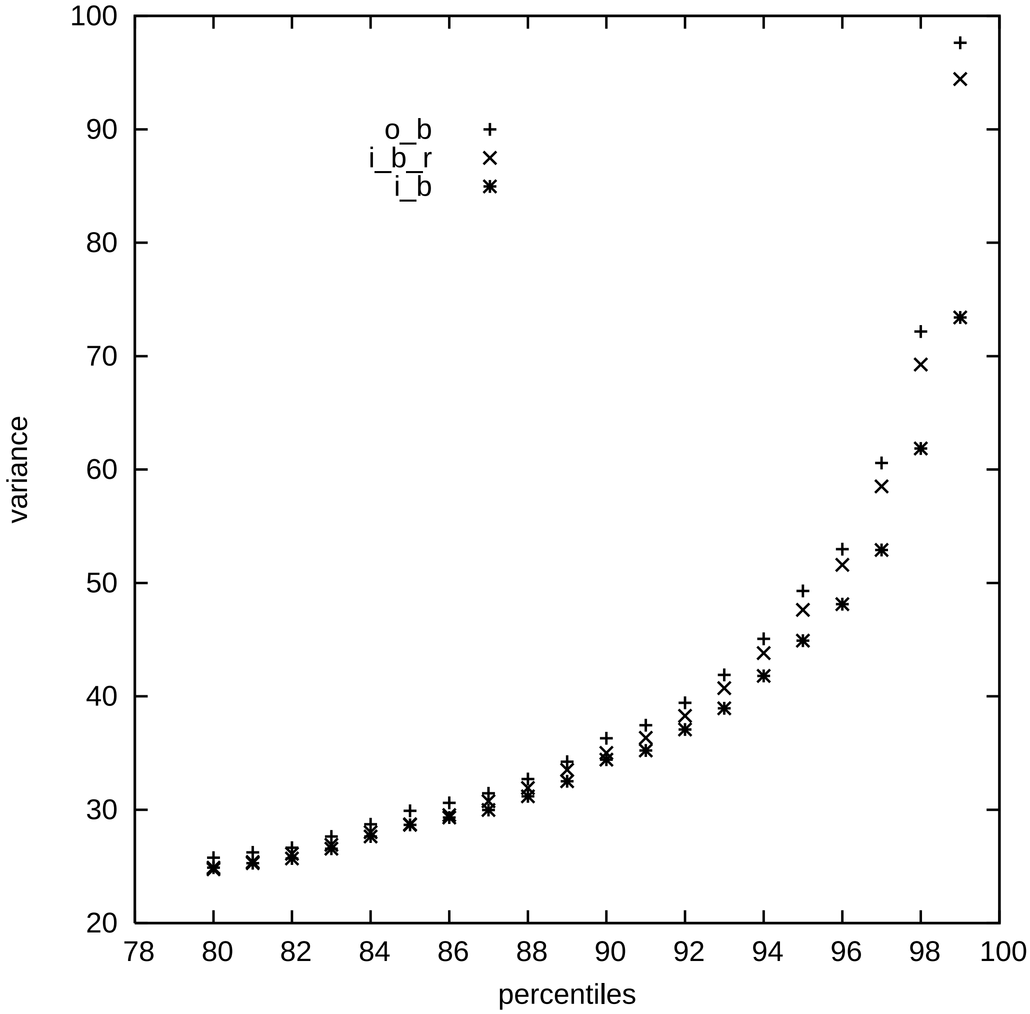

Fig. 5. Standard deviation of exceedance rates in 30-years periods expressed as the percentage of the mean values for the in-base (i_b) and out-base (o_b) periods. i_b_r indicates in-base rates being estimated by resampling procedure.

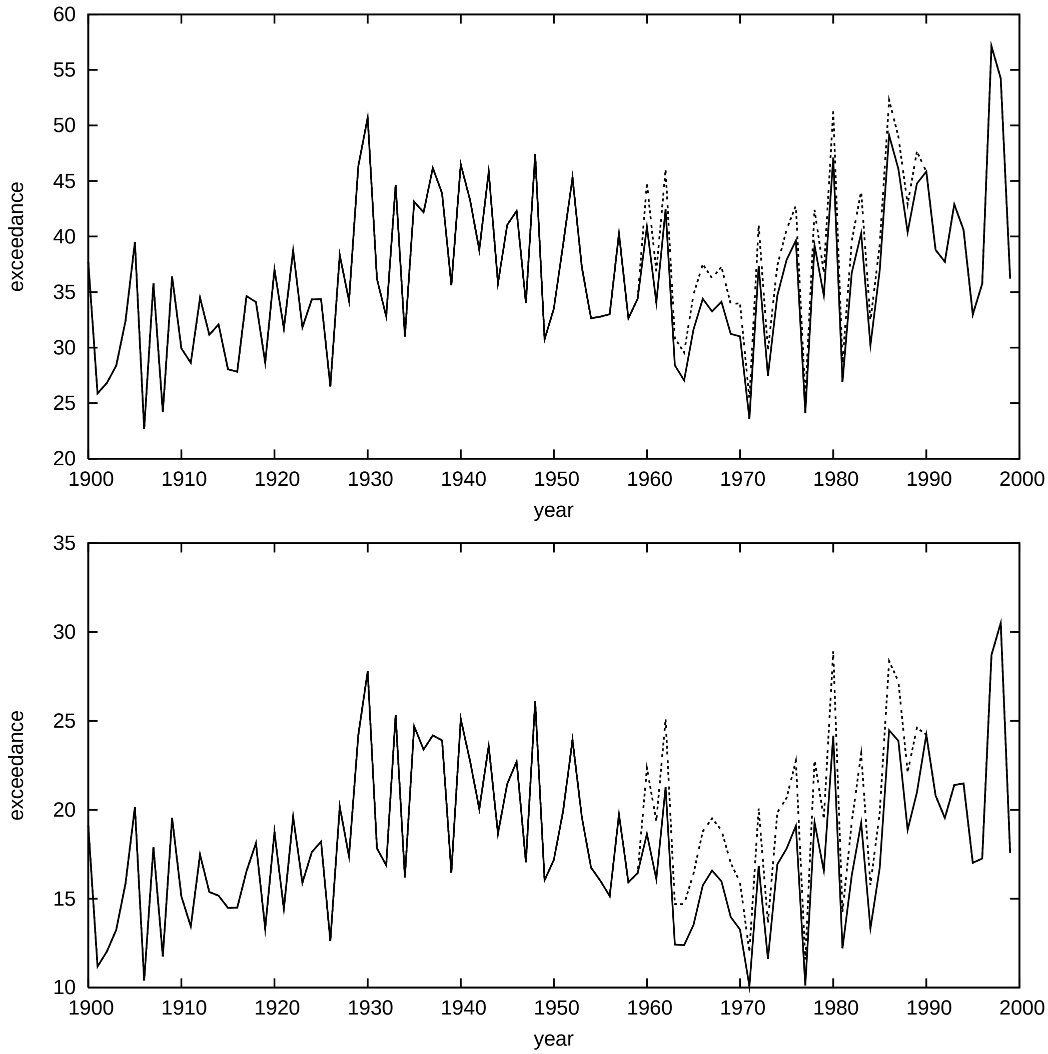

Fig. 6. Average number of days when daily mean temperature exceeds its 90th (upper panel) and 95th (lower panel) percentiles over Canada. Rates for the in-base period computed by bootstrap resampling procedure are plotted dash.