Time-varying autoregressive models for the detection of seasonal trends in Alaska temperature data

Javier López-de-Lacalle

May 31, 2010

Abstract

We fit an autoregressive time series model with time-varying parameters to study changes in seasonal means of temperature data observed in Fairbanks (Alaska, USA). Other studies have found an increase in temperature using annual or seasonal averages. We develop a statistical model that allows us to have a further insight in this issue using monthly data. The results show that the overall increase in temperature is driven mainly by warmer winter months. The increase in temperature is more prominent in winter months. A decrease in temperature is found in October.

1 Introduction

Our planet has known periods of warm and cold mean temperatures throughout its millennial history. The tendency observed in climatological variables over the last decades indicates a more noticeable increase in temperature as well as a relatively sharper intensity of meteorological phenomena (Webster et al., 2005; Emanuel, 2005). There are a number of reasons that explain changes in climate variables: some of them are attached to natural phenomena (for instance, the Pacific Decadal Oscillation) while others, such as an increase in industrial activity, are related to anthropogenic causes.

There is a general agreement on the need for effective measures to cope with climate change and its impact both on Earth and human life. Yet, the analysis of climatological data is not only relevant in a long-term scenario. As weather forecasts have become more accurate, a variety of economic and social activities rely on weather related information for short-term decision making (Seater, 1993; Romilly, 2005).

The design of sensible economic and environmental policies requires scientific research and advice. Polar regions are critical areas which require a closer monitoring on climate variables. Those areas are more sensitive to changes in climate and disturbances on the environment, e.g. ice melting, may trigger consequences at a global scale.

Empirical investigations on sample data indicate an increase in temperature over the last decades. Wendler and Shulski (2009) fit a linear trend to temperature data observed in Fairbanks over the last century and estimate an overall increase in mean annual temperature of 1.4°C degrees. Taking five-year running means, they also notice a non-uniform evolution throughout the sample. In addition, a comparison of seasonal means for different sample periods suggests temperature evolves differently across seasons.

Most of the empirical analyses are based on annual data and, hence, a rich amount on information available on higher frequency data (monthly data). Hartmann and Wendler (2005) provide a detailed description of several climatological variables observed at different locations in Alaska considering both annual and seasonal means. In particular, they compare mean values observed in the periods 1951-75 and 1977-2001 (before and after the shift observed in the Pacific Decadal Oscillation Index). Contrary to the behaviour typical of non-linear structures (as it is the case of climate variables) the traditional assessment of changes in temperature assumes a constant rate of change over years.

Measuring stylized facts upon comparison of statistics in different subsamples –i.e. running means– provides some guidance about the overall evolution of climate variables. However, it does not reveal the latent structure of the data. The focus of this study is to have an insight into the stylized facts by means of a dynamic statistical model that allows for seasonal specific parameters and non-linear trends. We fit the model to the the series of average monthly temperatures observed in Fairbanks in the period December 1929 to April 2009. The results provide a dynamic view to changes in temperature at a monthly resolution.

The remaining of the paper is organized as follows. Section 2 gives the source of the data analysed in this study and describes the main features observed in a preliminary view to the data. Section 3 introduces the statistical model fitted to the data. Section 4 summarizes the results. Section 5 points out some concluding remarks.

2 Data

The Alaska Climate Research Center –`http://climate.gi.alaska.edu`– disseminates historical data on temperature and other climate variables recorded in several meteorological stations located in Alaska. In this document, we analyze the series of average monthly temperatures observed in Fairbanks in the period December 1929 to April 2009.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

3 Statistical model

We propose the following model:

$$y_t - \mu_{s,t} = \sum_{i=1}^{p_s} \phi_{s,i}\,(y_{t-i} - \mu_{s,t-i}) + \epsilon_t\,, \quad \epsilon_t \sim IID(0, \sigma_s^2)\,, \tag{1}$$

$$\mu_{s,t} = \sum_{i=0}^{b_s} \beta_{s,i}\,\psi(t)\,, \tag{2}$$

where $\psi(t)$ is a polynomial of degree b_s that captures changes in the seasonal means. Some of the parameters vary with the month $s = 1, 2, \ldots, 12$. The polynomials $\psi(t)$ that capture time-varying means are orthogonalized to avoid numerical instability. We also apply a tentative model which uses basic-splines polynomials (Pollock, 1999, Chapter 10).

The analysis of the series takes the following steps:

1. Choice of model (autoregressive order and polynomial orders) and fit the the model by maximum likelihood. The choice of the model is based on the significance of parameters. The parameters are estimated by maximizing the log-likelihood function concentrated with respect to the scale parameters.

2. Substract the monthly time-varying means from the series and compute the time-varying periodogram involved in the estimated autoregressive coefficients.

3. Extract components upon the decomposition of the AR model (West, 1997; Krystal et al., 1999).

4 Results

A preliminary analysis of the data was based on the analysis of the twelve series of observations separately for each month. After removing the monthly means, the demeaned series $y_t - \hat{\mu}_{s,t}$ exhibited periodic autocorrelation. This fact indicates that the individual analysis of the twelve series of monthly paths omits part of the dynamic in the data. The observation y_t is correlated with the values in previous months, thus, analysing the series as a whole is a more appropriate approach. Our model captures periodic correlation across months by means of lagged values of the series included as regressors. Lags up to order 2 and 3, depending on each season, were significant.

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

5 Concluding remarks

We have shown the usefulness of a time-varying autoregressive model to capture changes in trends of temperature data. By including month specific parameters, the analysis reveals different trend patterns across months.

The diagnostic of the model indicates that the residuals from the fitted model are in agreement with the disturbance term specified in the model. Nevertheless, a longer left-tail in the histogram of the residuals causes a departure from the Gaussian assumption. The same was found using the Laplace distribution. Using a common scaling parameter of the distribution, the histogram of residuals resembled the Lapace density. After rescaling the residuals using periodic standard deviations, the corresponding histogram was close to the Normal distribution. As we worked with the likelihood function concentrated with respect to the scale parameters, considering periodic scale parameters did not entail a greater computational burden in the optimization algorithm. The skew Normal distribution may be useful to capture asymmetry. However, as the degree of skewness is low, and considering that the skew Normal does not allow concentration of parameters, the Normal distribution is a sensible an efficient choice.

The results showed that the overall increase in temperature observed in annual data occurs over January, February, April and December. Non-linear trends are also captured by means of cubic polynomials.

[Figure 8 about here.]

[Figure 9 about here.]

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

References

Emanuel, K. (2005), 'Increasing Destructiveness of Tropical Cyclones over the Past 30 Years', *Nature* **436**(4), 686–688.

Hartmann, B. and Wendler, G. (2005), 'The Significance of the 1976 Pacific Climate Shift in the Climatology of Alaska', *Journal of Climate* **18**, 4824–4839.

Krystal, A. D., Prado, R. and West, M. (1999), 'New Methods of Time Series Analysis of Non-Stationary EEG Data: Eigenstructure Decompositions of Time-Varying Autoregressions', *Clinical Neurophysiology* **110**, 2197–2206.

Pollock, D. S. G. (1999), *A Handbook of Time-Series Analysis Signal Processing and Dynamics*, Academic Press.

Romilly, P. (2005), 'Time Series Modelling of Global Mean Temperature for Managerial Decision-Making', *Journal of Environmental Management* **76**, 61–70.

Seater, J. J. (1993), 'World Temperature-Trend Uncertainties and their Implications for Economic Policy', *Journal of Business and Economic Statistics* **11**(3), 265–277.

Webster, P., Holland, G., Curry, J. and Chang, H.-R. (2005), 'Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment', *Science* **309**, 1844–1846.

Wendler, G. and Shulski, M. (2009), 'A Century of Climate Change for Fairbanks, Alaska', *Arctic* **62**(3), 295–300.

West, M. (1997), 'Time series decomposition', *Biometrika* **84**(2), 489–494.

Figure 1: Data: Average monthly temperature recorded in Fairbanks (Alaska) in the period December 1929 to April 2009. The dots are the annual means. Source: Alaska Climate Research Center http://climate.gi.alaska.edu.

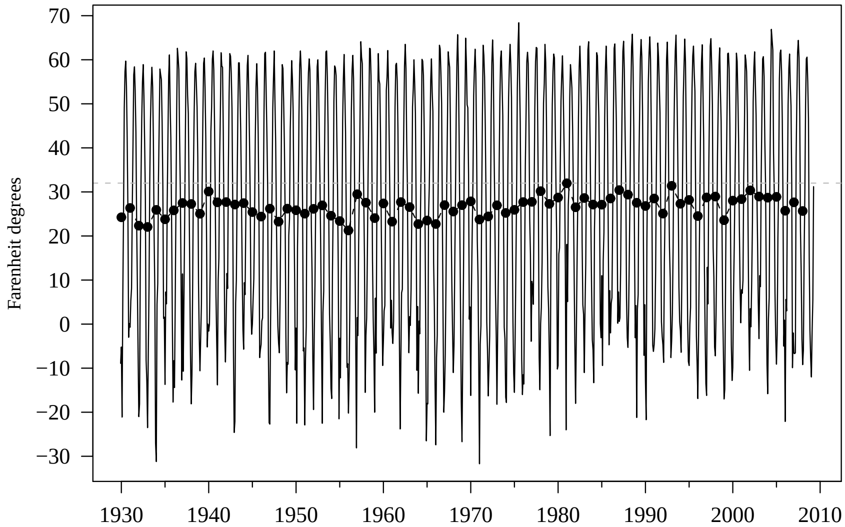

Figure 2: Annual means of temperature data. Coloured lines are trend estimates based on Gram polynomials of degrees 3, 4 and 5. Some geophysical research studies attach the shift in the level observed around 1976 to the Pacific decadal oscillation. This document investigates whether the observed increase –whatever the cause– is homogeneous across months.

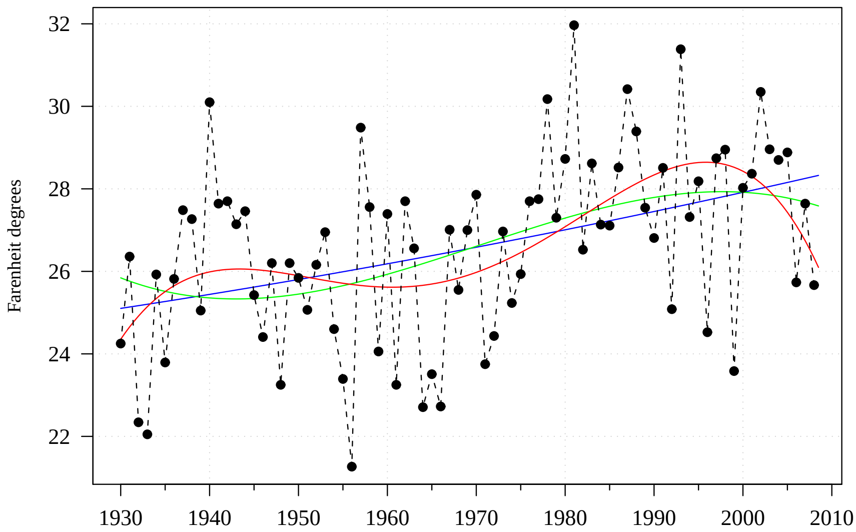

Figure 3: Monthly box plot of temperature data. Notice the higher variability in winter months compared to central months of the year.

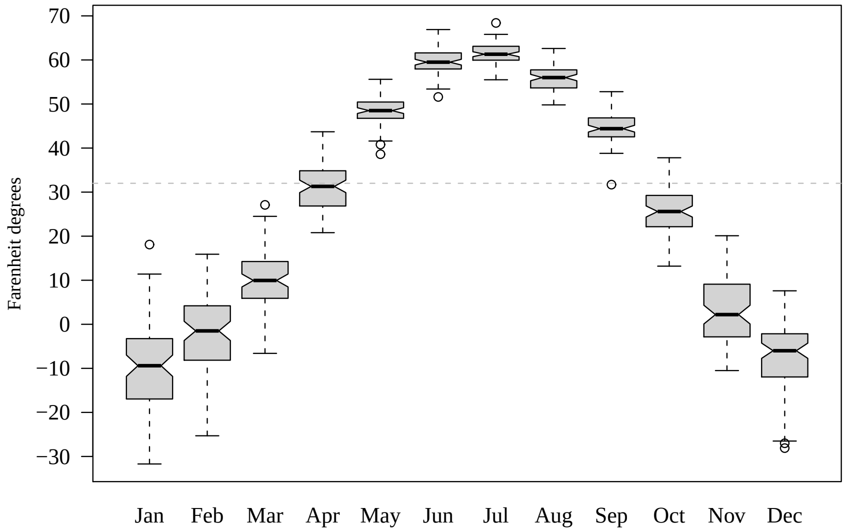

Figure 4: Fitted monthly time-varying means. Trends in temperature levels are not homogeneous across months: there is a substantial increase in temperature over January, February, April and December; the temperature remains relatively constant in November and September; the level decreases in October. Notice also non-linear patterns such as the level in March, where the temperature increases slightly in the beginning of the sample while, in the latest years, the level decreases.

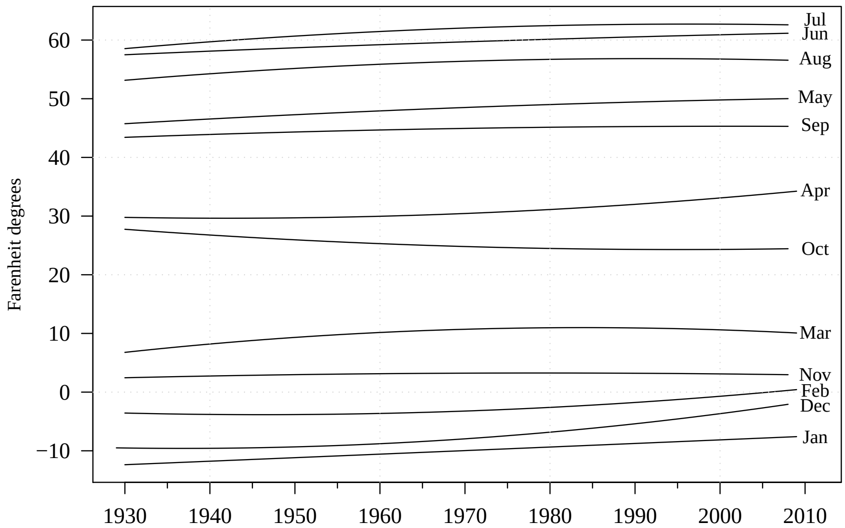

Figure 5: Monthly paths and fitted time-varying means. This graphic is based on the model given in equations (1-2) and provides a dynamic view to the boxplot shown in Figure 3. Comments given in Figure 4 applies to this plot.

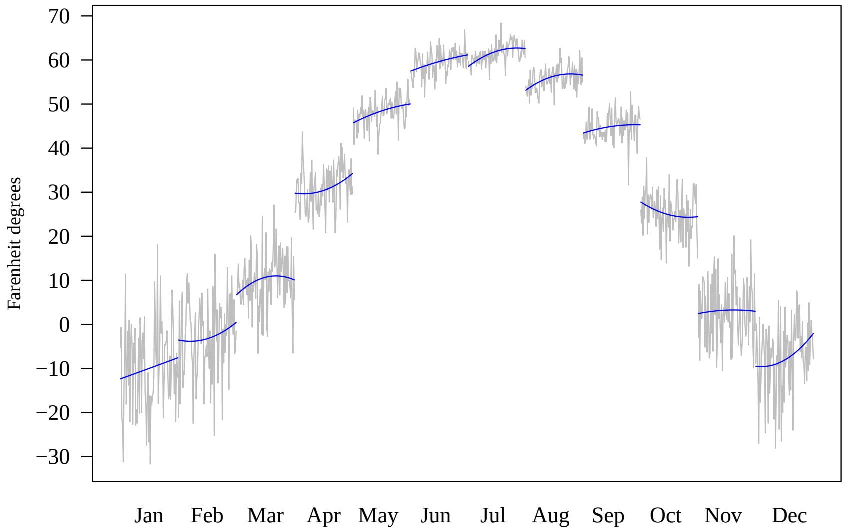

Figure 6: Time-varying spectrum. Spectral density related to the periodic autoregressive coefficients and variances. The presence of periodicities in the autocorrelations involves cycles of different frequency at each month of the year. We observe cycles related to trending patterns in December and January, which are related to a more noticeable increase in temperature in those months (as shown also in Figure 5). There are cycles of higher frequency in November. In the remaining months, there are no substantial fluctuations after removing the time-varying means.

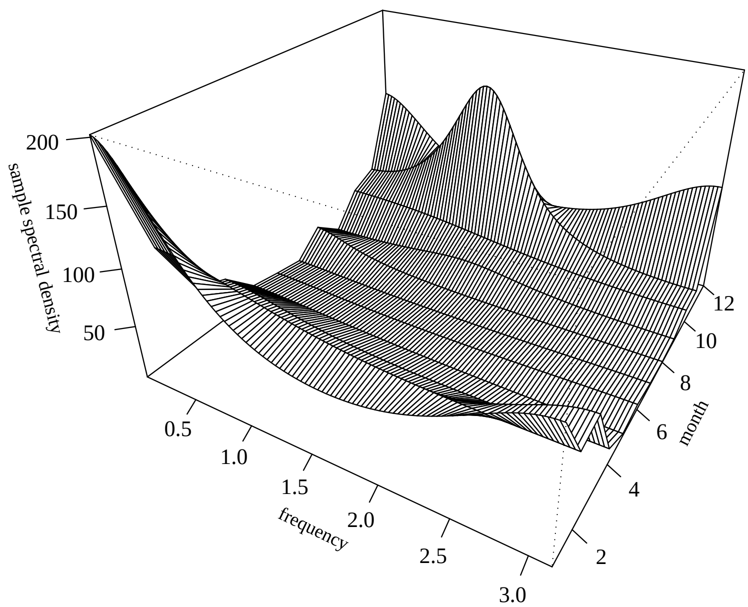

Figure 7: Model diagnostic. There is no structure remaining in the standardized residuals (sample autocorrelations lie within the 95% confidence bands for the null of no autocorrelation). The distribution of the standardized residuals is close to the standard Normal distribution. However, the Jarque-Bera test for normality revealed an excess of kurtosis; besides, the left-tail is longer than the right-tail.

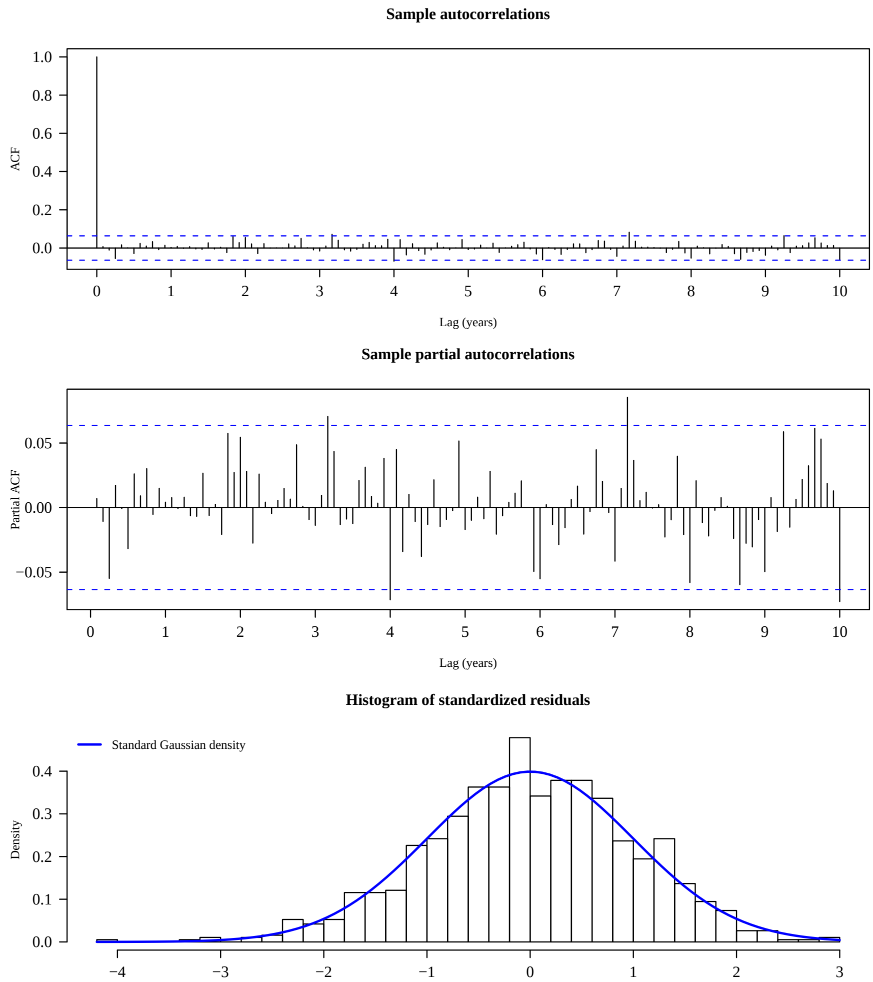

Figure 8: Components of the demeaned series. The components are based on the eigenstructure of the periodic autoregressive parameters.

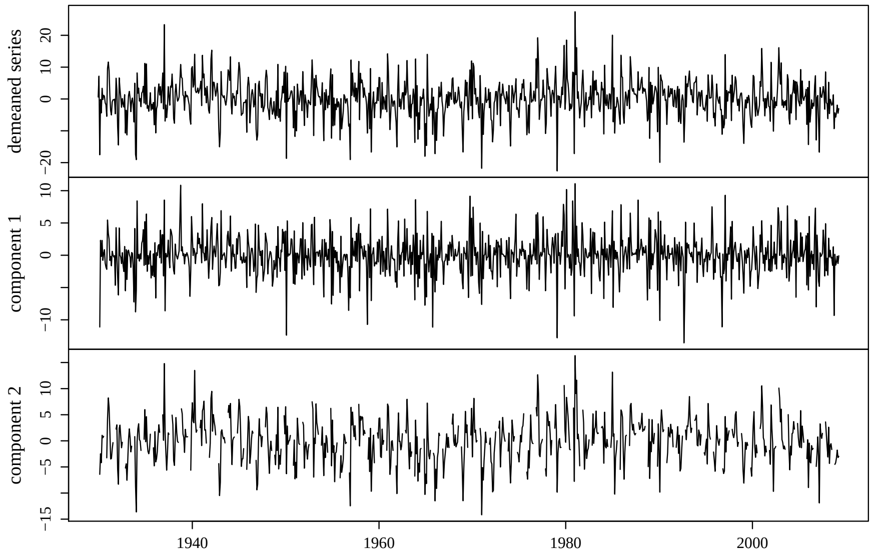

Figure 9: Components of the demeaned series (subsample).

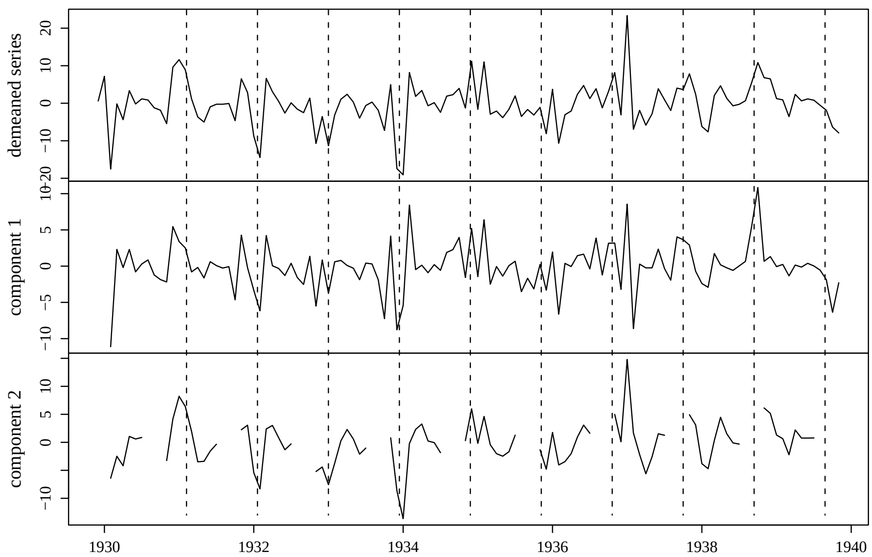

Figure 10: Contour lines of time varying sample spectra (two-dimensional version of Figure 6). Vertical dotted lines indicate the frequencies related to the components shown in Figures 8-9.

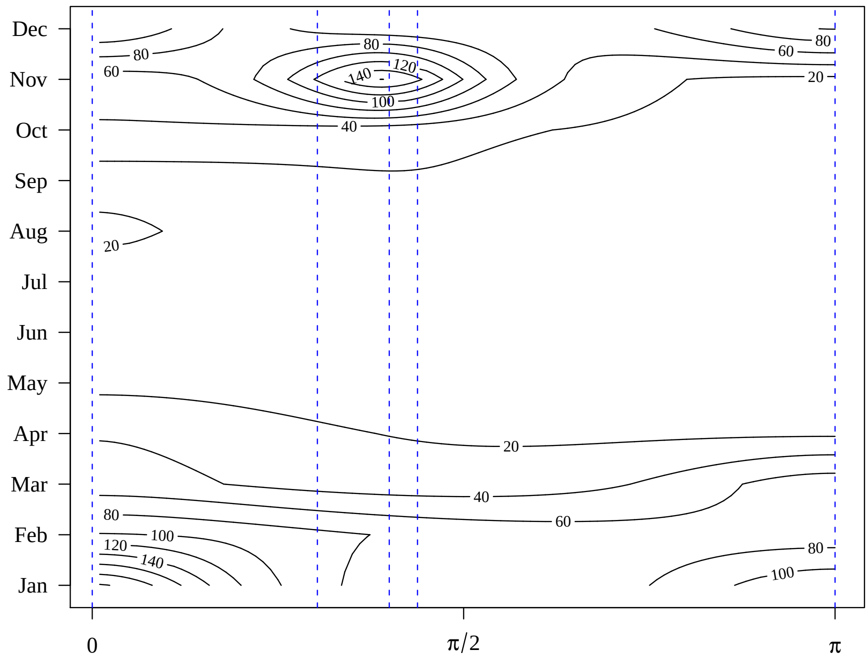

Figure 11: Fitted monthly time-varying means based on basic-spline functions. (The model is tentative and does not include autoregressive regressors.) See also Figure 12.

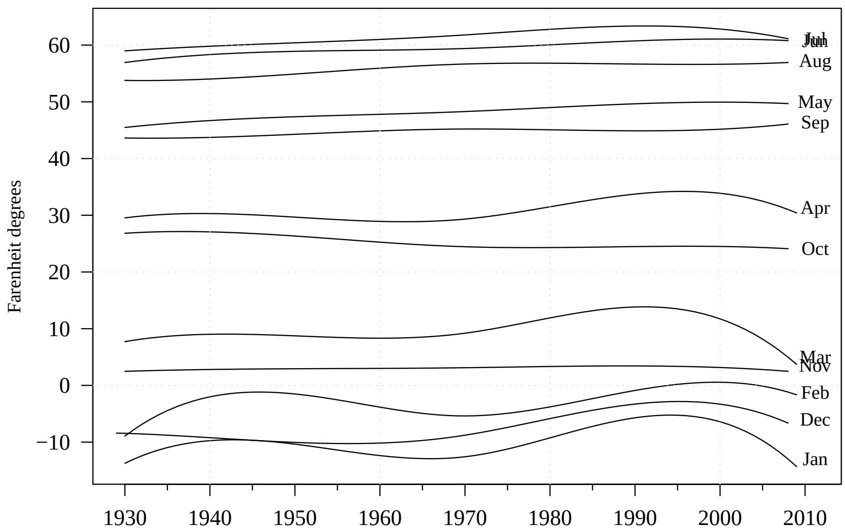

Figure 12: Monthly paths and fitted time-varying means based on basic-spline functions. (The model is tentative and does not include autoregressive regressors.) See also Figure 11.

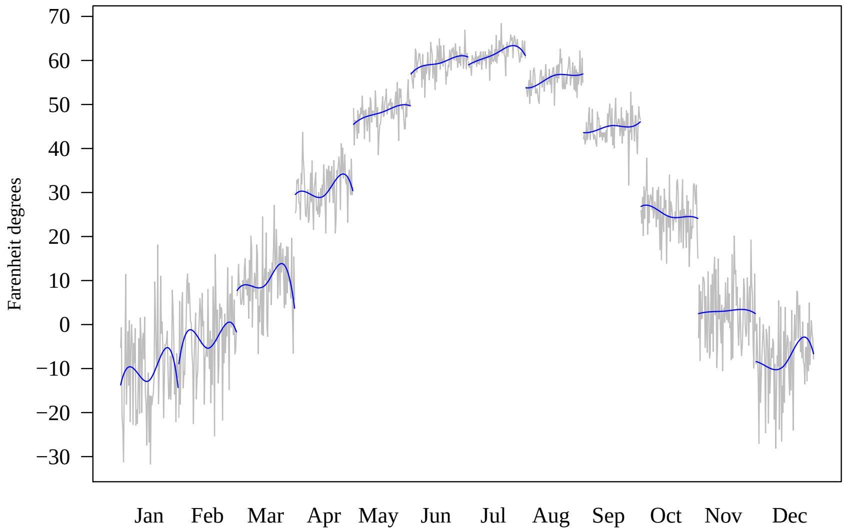

Figure 13: Density estimate of temperatures in January conditional on the PDO index. The conditional density is bimodal both when the PDO index is −1 and 1. When the PDO index is −1, the mode that reaches the highest value of the density is lower than the other mode. When the PDO index is 1, the density concentrates at higher values of temperature. This reflects the fact that a positive value of the PDO index leads to higher temperatures. The highest peak of the conditional density increases from −17 to −4 Farenheit degrees when the PDO switches from −1 to 1. The change is not that sharp across years in each circulation period, nonetheless. There are years with warmer temperatures when the PDO is −1 and cooler periods when the PDO is 1. The humps with lower density concentrate at temperatures of −5 and −15.5 Farenheit degrees, respectively for the PDO index equal to −1 and 1.

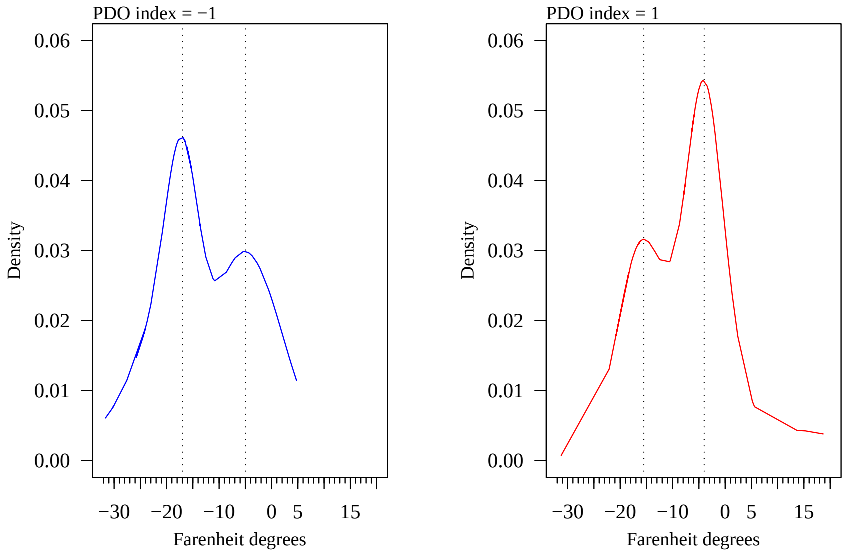

Figure 14: Estimated probabilities in an AR model with Markov switching mean for the sample of temperatures recorded in January. The shaded areas indicate the years in which the PDO index is over zero in January (warm regime). Despite the PDO involves a switch between two circulation patterns, we considered the presence of three regimes in the model fitted to the data. The traditional view states that, according to the PDO circulation patters, there are two regimes: a colder regime before the year 1976 and a warmer period in the second half of the sample after an abrupt increase in 1976. The statistical model fitted to the temperatures observed in January yields three regimes characterized as follows. The first regime is the coolest period, it refers mainly to the period 1966-1975. The second regime covers the years before and after the cooler period around 1970, except for some exceptional cold years in the first half of the sample that are attached to the first regime. The third regime stands for the highest peaks in temperature, it captures the warmest years 1937 and 1981. It also captures some other warmer years in the second half of the sample, in agreement with a warm period according to the PDO.