Modeling Soil Temperature Timeseries Through a Bayesian Dynamic Linear Model for Structural Breaks

Manju M. Johny

December 11, 2017

1 Introduction

The trend of increasing temperatures over time has widely been used to signal evidence of climate change. In addition to an increase in mean surface temperatures, there has been evidence of increases in both daily minimum and maximum surface temperatures over the last 50 years [1]. Interestingly, the rate of warming of daily minimum and maximum surface temperatures have not been uniform. There has been a relatively larger increase in daily minimum surface temperatures compared to daily maximum surface temperatures [1]. Thus, increases in mean surface temperatures have coincided with decreases in diurnal surface temperatures, where diurnal temperature is defined as the difference between daily maximum and minimum temperatures. Since the soil is home to a diverse array of life, the detection of a similar pattern in soil temperature is of scientific interest as it could have profound implications for the soil ecosystem.

To mimic climate change, warmer temperatures and earlier snowmelt were experimentally simulated through the use of passive heating chambers and snow removal in the montane meadows. Soil temperature measurements were recorded for the aforementioned heating + snow removal treatment and for a control treatment [3]. In this paper, we focus on dissecting the patterns in soil temperature in the two treatments over time. Figure 1 shows the raw timeseries data over the course of 22 days, from July 23, 2011 to August 14, 2011. The timeseries exhibit trend, daily seasonality, and a severe structural break. The structural break, which are common for temperature data, was likely due to a sudden shift in weather patterns. In this project, we utilize a Bayesian dynamic linear model that is appropriate for structural breaks in order to extract the trend and seasonality in the data. While the trend will reflect the general trajectories of the two series, the seasonal component can elucidate differences in the diurnal soil temperatures between the control treatment and the heating + snow removal treatment.

2 Data

The data consist of soil temperature measurements taken from a sagebrush meadow at an elevation of 2100 m in Grand Teton National Park, WY. In this analysis, we focus on two treatments: (1) control, C; and (2) heating + snow removal,HS that were established in 8x8 plots. The heating + snow removal plot was established under a passive heating chamber and snow was removed from the surface of the chamber. The control plot lacked the passive heating chambers and snow was not removed from the soil's surface. Soil temperature was obtained hourly at a depth of 5cm over the course of 22 days, from July 23, 2011 to August 14, 2011 [3]. For computational reasons, we extracted every third hourly observation. Thus, our analysis is based on timeseries composed of 176 observations from the control treatment and heating + snow removal treatment.

3 Methods

3.1 Model Rationale

Figure 1, the plot of the two soil temperature timeseries, show evidence of trend, and seasonality. The cyclical pattern repeats daily with 8 observations in each cycle. Hence, a local linear trend plus seasonal factor DLM seems appropriate. The data does not show evidence of outliers, but there is a clear structural break. In order to account for the structural break in the states, the system evolution variances are given a heavy tailed, t-distribution with 2 degrees of freedom. The benefit of using a t-distribution is in its equivalent representation as a scale mixture of Normal distributions, which allows us to use the standard algorithms in DLM such as FFBS [2]. Suppose the i^{th} component of the system evolution variance is given by W_{it}, and $W_{it}|\lambda_i, \nu \sim T(0, \lambda_i; \nu)$ where ν is the degrees of freedom, λ_i is the scale parameter, and t are the time indices. Then, by introducing a latent variable ω_{it}, we can give the equivalent expression:

$$\begin{aligned} W_{it}|\lambda_i, \omega_{ti} &\sim N(0, \lambda_i \omega_{it}) \\ \omega_{it}|\nu &\sim IG\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \end{aligned} \tag{1}$$

Additionally, we allow these state evolution variances along with it's corresponding latent variable ω_{it} to be time varying. By doing so, we can use the posterior means of the ω_{it} over time to identify the time of the structural break [2]. Thus, outliers in the latent variable ω_{it} indicate that an abrupt change in the corresponding state occurred at that time point.

3.2 Model

The DLM for the heating + snow removal and control timeseries are both specified in the same way. For simplicity, we give the model for just one of the

timeseries.

Let Y_t be the soil temperature measurement at the t^{th} time. Let t, ranging from 1 through 176, be the time at which each soil temperature measurement was taken.

For t = {1,2,..,176},

Observation Equation

$$Y_t = F\theta_t + v_t \qquad\qquad v_t \sim N(0, \sigma^2) \tag{2}$$

State Equation

$$\begin{aligned} &\theta_t = G\theta_{t-1} + z_t \qquad\qquad z_t \sim N(0, W_t) \\ &\theta_0 \sim N(0, 10^7 \mathrm{I}_{9\times 9}) \end{aligned} \tag{3}$$

$$W_t = \begin{bmatrix} W_{1t} & 0 & 0 & 0 & \dots & 0 \\ 0 & W_{2t} & 0 & 0 & \dots & 0 \\ 0 & 0 & W_{3t} & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}_{9\times 9}$$

The i^{th} non-zero element of W_t is given by W_{it}, and is defined as below.

$$\begin{aligned} W_{it} &= \lambda_i \omega_{it} \\ \omega_{it} &\overset{ind}{\sim} IG(1,1) \end{aligned} \tag{4}$$

F, G, and θ_t are defined as below:

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 & \dots & 0 \end{bmatrix}_{1\times 9} \tag{5}$$

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & -1 & -1 & \dots & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \vdots & 0 & 0 & \ddots & 0 & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}_{9\times 9} \tag{6}$$

$$\theta_t = \begin{bmatrix} \mu_t & \beta_t & \alpha_j & \alpha_{j-1} & \dots & \alpha_1 & \alpha_s & \dots & \alpha_{j+2} \end{bmatrix} \tag{7}$$

where $j = t \bmod 8$

Priors

$$\begin{aligned} \sigma^2 &\sim IG\big(1,1\big) \\ \lambda_i &\overset{ind}{\sim} IG(1,1) \end{aligned} \tag{8}$$

The priors were chosen to be conjugate Inverse Gamma distributions. The shape and scale parameters were chosen to be 1, and seem to give sufficient density around plausible values of σ^2 and λ_i. In Figure 5 and 6, overlays of the posterior distributions of these parameters with the prior distributions show that the data are able to overwhelm the prior.

3.3 Model Fitting

We obtain samples from the posterior distributions of the parameters and states given in the model by implementing a Gibbs sampler. The initial value for σ^2 was 0.5, and the initial values for ω_{it} and λ_i were sequences of 1 of the appropriate dimension. The states are generated from their full conditional distribution given the parameters and observations using the FFBS algorithm in the dlm package. Then, the parameters are drawn from their full conditional distributions given the states and observations. This successive sampling is repeated for 35,000 iterations. Thus, a 1 chain MCMC was run for 35,000 iterations with 5,000 iterations discarded as burn-in. Ultimately, this culminated in 30,000 MCMC samples generated for the parameters and states. The trace plots indicated well mixed chains. Additionally, the geweke diagnostics testing equality of means between the first 10% and last 50% of the Markov chains were calculated, and gave insignificant z-scores. Based on the trace plots and geweke diagnostics, there is no indication of lack of convergence in the Markov chain.

4 Results

Posterior distributions for σ^2, ω_{ti}, and λ_i were obtained using the MCMC. The posterior distributions for σ^2, λ_1, λ_2 and λ_3 are shown in Figures 5 and 6. In the local linear trend + seasonal factor model that was fit, the state evolution variances were time-variant. Thus, the latent variable ω_{it} on the evolution variances are also time-variant. Figure 4 shows a plot of the posterior means over time for ω_{1t}, ω_{2t}, and ω_{3t}, corresponding to the latent variable on the state evolution variances for the local level, local growth rate, and seasonal component of the states. In Figure 2, we present the smoothing estimates of the trend and seasonal components for the control (C) and heating + snow removal (HS) timeseries along with 95% credible bands. Figure 3 shows the same estimates and 95% credible bands overlaied for comparison.

5 Discussion

The objective of this project was to fit a DLM that would account for a severe structural break, and subsequently extract the trend and seasonality of soil temperature timeseries from two treatments: control (C), and heating + snow removal (HS). After fitting the DLM, the plot of the posterior means of the latent variables ω_{it} on the system evolution variances can be used to identify the location of the structural breaks [2]. Particularly, the time points that have large posterior means for ω_{it} indicate that the corresponding system evolution variance needed to be large at that time point in order to make a "jump" in the structure of the state. Thus, high outliers in the ω_{1t}, ω_{2t}, and ω_{3t} posterior means will indicate structural breaks in the local level, local growth rate, and seasonal component of the states respectively. Figure 4 shows high outliers in

ω_{1t} and ω_{2t} for both treatments between August 2 and August 4, corresponding to a structural break in the local level and local growth rate. This is the time period at which the data experienced an unusual dip which is apparent in both the raw data (Figure 1) and the trend estimate (Figure 2). Additionally, Figure 4 shows high outliers in ω_{3t} for both treatments between August 4 and August 5, corresponding to a structural break in the seasonal component of the states. This is the time period at which the amplitude of the seasonal component, shown in Figure 2, slightly increases. It is worth noting that, although the amplitudes of the seasonal components increased for both treatments over time, the changes in amplitudes are not dramatic and a time-variant seasonal component may not have been necessary for this model.

Aside from fitting a DLM that accounts for structural breaks, we compare the estimated trend and seasonal component between the control and heating + snow removal treatments. Figure 2 shows estimates and 95% credible bands of both trend and seasonality for the two treatments. The trend of soil temperature for the two series indicate that the heating + snow removal treatment is associated with higher temperatures than the control treatment. This difference is apparent prior to the dip in the trend between August 2 and August 4, and less apparent after the dip where the credible bands overlap. The seasonal component estimates for the two series indicate a smaller amplitude for the heating + snow removal treatment compared to the control treatment. Thus, there is some visual evidence that the heating + snow removal treatment is associated with larger mean temperatures but smaller diurnal temperatures compared to the control treatment.

6 Improvements and Future Work

As noted in the discussion, the amplitudes of the seasonal components had no dramatic changes over time for either of the series. Thus, we could adapt the model in this paper to a simpler DLM with fixed seasonality. Additionally, for computational purposes, we limited the analysis to only 22 days out of the available 3 months of data. It would be interesting to extract the trend and seasonal component for the full 3 months of data to see if the same patterns persist. We may have also lost some information, particularly regarding the seasonal component, by looking at only every third observation. Since our main objective in looking at the seasonal component was to learn about diurnal temperatures, having the full 24 hourly observations in a day could give more insight regarding the maximum and minimum temperatures. Lastly, each treatment contained 2 additional soil temperature timeseries. For simplicity, we only looked at 1 series from each of the 2 treatments. We should extract the trend and seasonality for all of the series to see whether the finding of this paper hold, or whether they were due to an anomaly in the measurements.

7 Figures

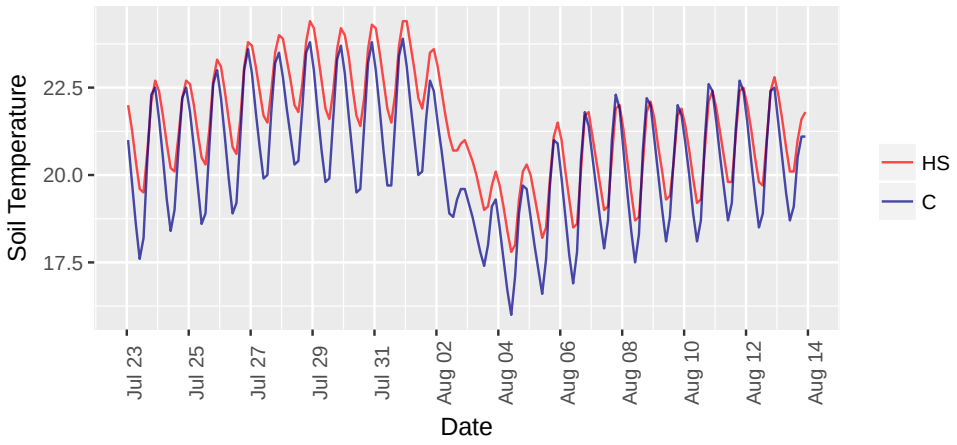

Figure 1: Plot of raw soil temperature measurements for control (C) and heating + snow removal (HS) treatments.

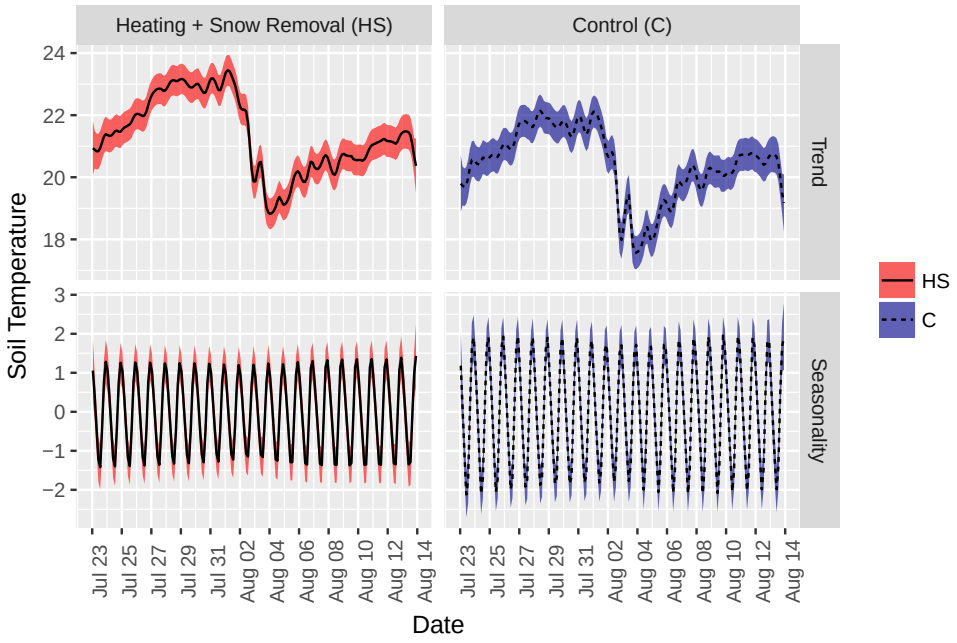

Figure 2: Estimated trend and seasonal component for the control (C) treatment is given by dashed line, and its 95% credible band is shown in blue. Estimated trend and seasonal component for the heating + snow removal (HS) treatment is given by solid line, and its 95% credible band is shown in red.

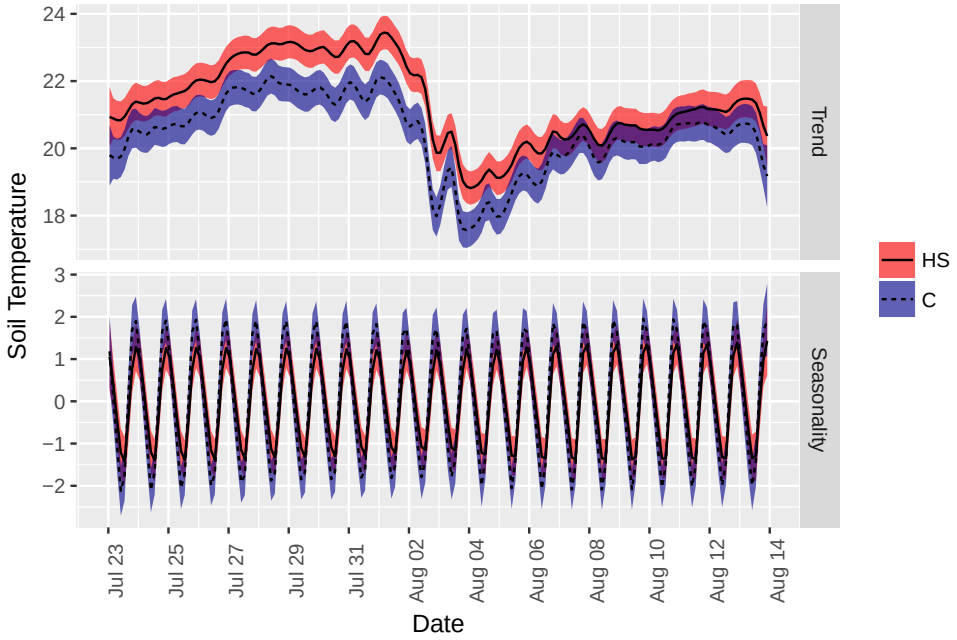

Figure 3: Estimated trend and seasonal component for the control (C) treatment is given by dashed line, and its 95% credible band is shown in blue. Estimated trend and seasonal component for the heating + snow removal (HS) treatment is given by solid line, and its 95% credible band is shown in red.

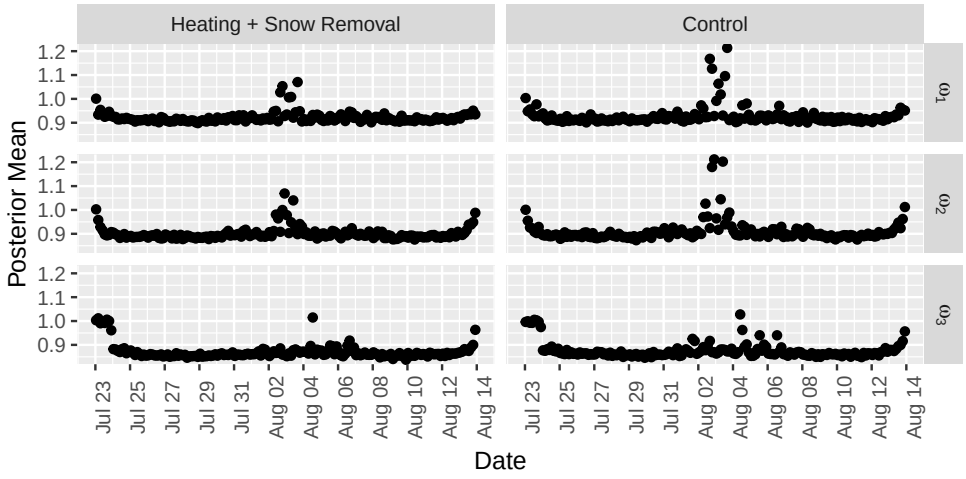

Figure 4: Posterior means for the latent variables ω_1, ω_2 and ω_3 on the system evolutions variances corresponding to local level, local growth rate, and seasonal component of states.

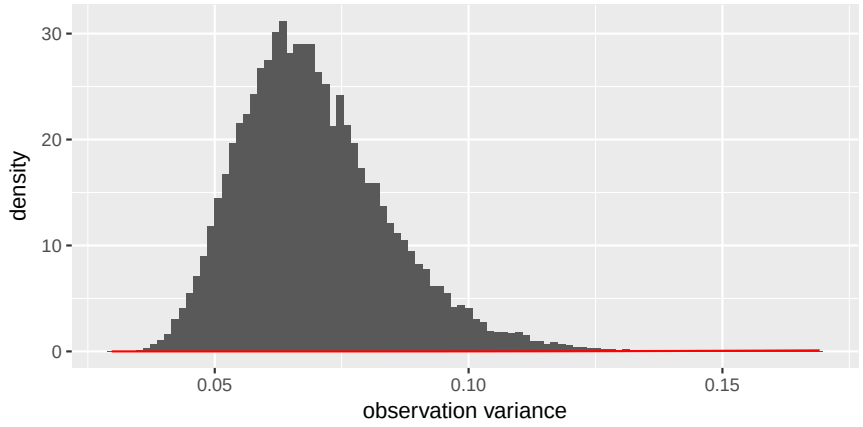

Figure 5: Posterior distribution of the observations variance (σ^2) shown in grey. Prior distribution shown in red

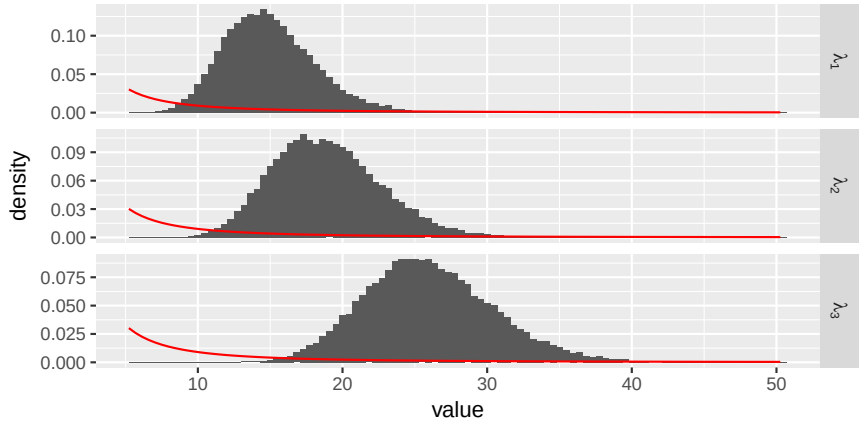

Figure 6: In gray are the posterior distributions of λ_1, λ_2, and λ_3 defined in the evolution variances that correspond to local level, local growth rate, and seasonal component respectively. Prior distribution shown in red

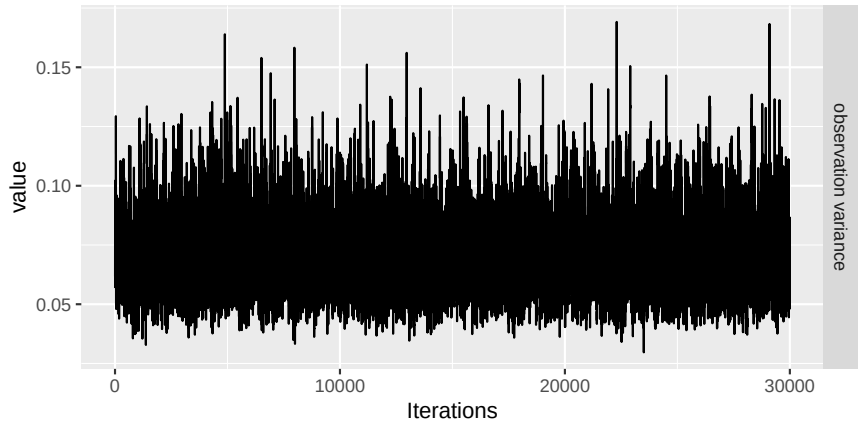

Figure 7: Observations variance (σ^2) traceplot

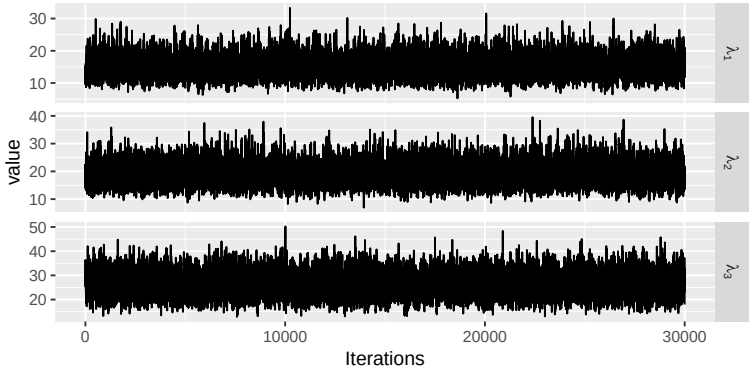

Figure 8: Trace plots of λ_1, λ_2, and λ_3 defined on the evolutions variances corresponding to local level, local growth rate, and seasonal component of states

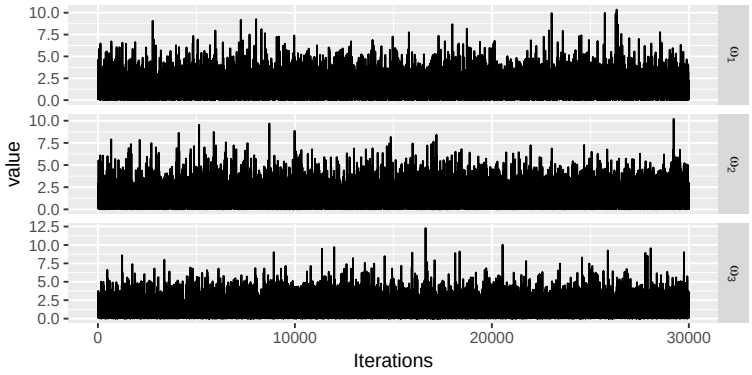

Figure 9: Trace plots of latent variable ω_{11}, ω_{21}, and ω_{31} on the evolutions variances corresponding to local level, local growth rate, and seasonal component of states at time point t=1

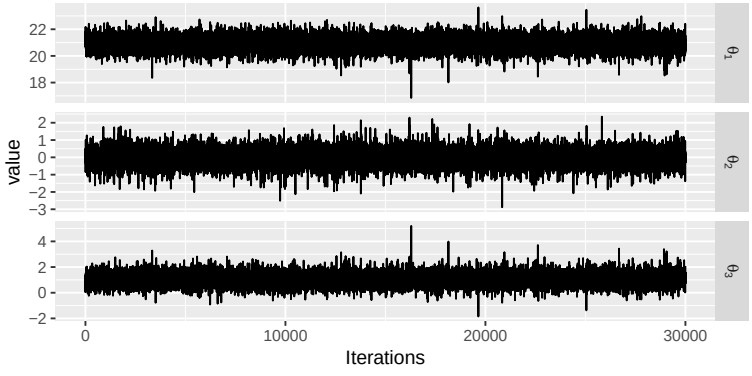

Figure 10: Trace plots of the states θ_1, θ_2, θ_3 corresponding to local level, local growth rate, and seasonal component at time point t=1

References

[1] K Braganza, D Karoly, and J.M. Arblaster. Diurnal temperature range as an index of global climate change during the twentieth century. *Geophysical Research Letters*, 31(13), 2004.

[2] G Petris, S Petrone, and P Campagnoli. *Dynamic Linear Models with R.* Springer, 2009.

[3] J. A. Sherwood, D. M. Debinski, P. C. Caragea, and M. J. Germino. Effects of experimentally reduced snowpack and passive warming on montane meadow plant phenology and floral resources. *Ecosphere*, 8(3), 2017.