Forecasting Texas Energy Grid Demand with Dynamic Linear Models

May 1, 2017

Professor James Scott

Jennifer Starling

1. Introduction

2. Data and Methods

2.1 Energy grid load data

Our data set consists of hourly temperature, business hour and energy grid load data for Texas, measured from Dec. 31, 2009 to Aug. 15, 2016. Energy grid load is measured in separately for eight zones in Texas, which are then aggregated to calculate total load for the entire Texas grid. We fit separate models for each zone to accomodate each zone's unique matrix of covariates.

Temperature is measured following the current methodology used by ERCOT to forecast Texas energy demand. This model calculates each zone temperature at time t as the weighted average of selected weather stations. Missing weather station observations are linearly interpolated.

Table 1: Temperature Weighting by Zone

Zone	Weather Station	Weight
North	KSPS	0.50
North	KPRX	0.50
North Central	KDFW	0.50
North Central	KACT	0.25
North Central	KMWL	0.25
East	KTYR	0.50
East	KLFK	0.50
Far West	KINK	0.50
Far West	KMAF	0.50
West	KABI	0.40
West	KSJT	0.40
West	KJCT	0.20
South Central	KAUS	0.50
South Central	KSAT	0.50
Coast	KLVJ	0.50
Coast	KGLS	0.30
Coast	KVCT	0.20
Southern	KCRP	0.40
Southern	KBRO	0.40
Southern	KLRD	0.20

2.2 Dynamic Linear Models

We fit a univariate dynamic linear model as described by West & Harrison [1]. These models are frequently used for forecasting time-series data. Under the Gaussian setting, these models are computationally tractable via the well-known Kalman Filter. Taking advantage of the Gibbs Sampler allows us to leverage the Gaussian updates while accounting for uncertainty regarding the unknown variance and covariance terms. *(Add info about other methods which have been used to do this - a bit of general background? Perhaps not in this section, but in intro?)*

2.3 Model-Fitting Details

We fit a univariate dynamic linear model as described by West & Harrison [1], with y_t describing energy grid load at discrete hourly intervals t. The system and state equations can be written as

$$\begin{aligned} y_t &= F_t\theta_t + \epsilon_t, \text{ with } \epsilon_t \sim N_1(0, V_t) \\ \theta_t &= G_t\theta_{t-1} + \eta_t, \text{ with } \eta_t \sim N_p(0, W_t) \end{aligned}$$

We set G_t to the identity matrix, and assume error variances V_t and W_t are time-invariant. The simplified model can be written as follow. Note that V is scalar and W is a $p \times p$ matrix.

$$\begin{aligned} y_t &= F_t\theta_t + \epsilon_t, \text{ with } \epsilon_t \sim N_1(0, V) \\ \theta_t &= \theta_{t-1} + \eta_t, \text{ with } \eta_t \sim N_p(0, W) \end{aligned}$$

The matrix F_t describes intercept and time-varying covariates, where γ_t represents temperature at time t, β_t is an indicator for business hour at time t. Specifically, $\beta_t = 1$ indicates a business hour, while $\beta_t = 0$ indicates nights, weekends and holidays. An autoregressive term y_{t-1} is the previous hour's grid load, and $h_{1t}, \ldots, h_{24t}$ are indicators representing the hours of the day.

$$F_t = \begin{bmatrix} 1 & \gamma_t & \gamma_t^2 & \beta_t & y_{t-1} & h_{1t}, & ..., & h_{24t} \end{bmatrix}$$

There are a variety of methods available for estimating the unknown variances V and W. Discount factor methods are popular, as is the *d-inverse-gamma* prior. We adopt the latter method, which is described as the most popular method by Petris [2]. The *d-inverse-gamma* prior is easily written in terms of precisions, such that

$$\begin{aligned} V &= \phi_y^{-1} \\ W &= diag\left(\phi_{\theta,1}^{-1}, \ldots, \phi_{\theta,p}^{-1}\right) \end{aligned}$$

We call the vector of precisions $\psi = \left(\phi_y^{-1}, \phi_{\theta,1}^{-1}, \ldots, \phi_{\theta,p}^{-1}\right)$ and assign the $d = (p+1)$ terms independent Gamma priors. Then the prior on the vector of variances is the product of d inverse Gamma densities, hence the name of the method. We parameterize the hyperparameters in terms of guesses of the means and variances of the unknown precisions, as described by Petris[2]. Let $E(\phi_y) = a_y$, and $E(\phi_{\theta,i}) = a_{\theta_i}$. Let prior uncertainty by expressed as $Var(\phi_y) = b_y$ and $Var(\phi_{\theta,i}) = b_{\theta,i}$ for $i = 1, \ldots, p$. We can then parameterize our priors as

$$\begin{aligned} \phi_y &\sim Ga(\alpha_y, \beta_y), \text{ with } \alpha_y = \frac{a_y^2}{b_y}, \beta_y = \frac{a_y}{b_y} \\ \phi_{\theta,i} &\sim Ga(\alpha_{\theta,i}, \beta_{\theta,i}), \text{ with } \alpha_{\theta,i} = \frac{a_{\theta,i}^2}{b_{\theta,i}}, \beta_{\theta,i} = \frac{a_{\theta,i}}{b_{\theta,i}}, i = 1, \ldots, p \end{aligned}$$

Given observations $y_{t:T}$ we can write the joint posterior of $\theta_{0:T}$ and $\psi = \left(\phi_y^{-1}, \phi_{\theta,1}^{-1}, \ldots, \phi_{\theta,p}^{-1}\right)$ as

$$\begin{aligned} p\left(\theta_{0:T}, \psi | y_{1:T}\right) &= p\left(y_{1:T}|\theta_{0:T}, \psi\right) \cdot p\left(\theta_{0:T}|\psi\right) \cdot p\left(\psi\right) \\ &= \prod_{t=1}^{T} p\left(y_t|\theta_t, \phi_y\right) \cdot \prod_{t=1}^{T} p\left(\theta_t|\theta_{t-1}, \phi_{\theta,1}, \ldots, \phi_{\theta,p}\right) \cdot p\left(\theta_0\right) \cdot p\left(\phi_y\right) \cdot \prod_{t=1}^{T} p(\phi_{\theta,i}) \end{aligned}$$

We then derive the full conditional of ϕ_y.

$$\begin{aligned} p\left(\phi_y | \ldots\right) &\propto \prod_{t=1}^{T} p\left(y_t|\theta_t, \phi_y\right) \cdot p\left(\phi_y\right) \\ &\propto \theta_y^{\frac{T}{2}} \exp\left[-\frac{\phi_y}{2}\sum_{t=1}^{T}\left(y_t - F_t\theta_t\right)^2\right] \cdot \phi_y^{\alpha_y - 1} \exp\left[-\phi_y\beta_y\right] \\ &= \phi_y^{\alpha_y + \frac{T}{2} - 1} \exp\left[-\frac{\phi_y}{2}\sum_{t=1}^{T}\left(y_t - F_t\theta_t\right)^2 + \beta_y\right] \end{aligned}$$

which we recognize as the gamma kernel. The full conditional of ϕ_y is

$$\phi_y \sim Ga\left(\alpha_y + \frac{T}{2}, \frac{1}{2}\sum_{t=1}^{T}\left(y_t - F_t\theta_t\right)^2 + \beta_y\right)$$

The full conditionals of the $\phi_{\theta,i}$ for $i = 1, \ldots, p$ are derived similarly. Since W is a diagonal matrix, the full conditionals can be written in terms of a sum over the p θ_t terms, so that each $\phi_{\theta,i}$ depends only on its corresponding $\theta_{i,t}$.

$$\begin{aligned} p\left(\phi_{\theta,i}|\ldots\right) &\propto p\left(\phi_{\theta,i}\right) \cdot \prod_{t=1}^{T} p\left(\theta_t|\theta_{t-1}, \phi_{\theta,1}, \ldots, \phi_{\theta,p}\right) \\ &\propto \phi_{\theta_1}^{\frac{T}{2}} \cdots \phi_{\theta_p}^{\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{p}\sum_{t=1}^{T} \phi_{\theta,i}\left(\theta_{i,t} - \left(G_t\theta_{t-1}\right)_i\right)^2\right] \\ &= \exp\left[-\frac{1}{2}\sum_{i=1}^{p}\sum_{t=1}^{T} \phi_{\theta,i}\left(\theta_{i,t} - \theta_{i,t-1}\right)^2\right] \end{aligned}$$

Then for the single $\phi_{\theta,i}$, all other $\phi_{\theta,j}$ terms become part of the constant of proportionality, leaving

$$p\left(\phi_{\theta,i}|\ldots\right) \propto \phi_{\theta,i}^{\left(\alpha_{\theta,i}-1\right)} \exp\left[-\phi_{\theta,i}\beta_{\theta,i}\right] \cdot \phi_{\theta,i}^{\frac{T}{2}} \exp\left[-\frac{\phi_{\theta,i}}{2}\sum_{t=1}^{T}\left(\theta_{i,t} - \theta_{i,t-1}\right)^2\right]$$

which we again recognize as the gamma kernel. The full conditional of $\phi_{\theta,i}$ is

$$p\left(\phi_{\theta,i}|\ldots\right) \sim Ga\left(\alpha_{\theta,i} + \frac{T}{2}, \beta_{\theta,i} + \frac{1}{2}\sum_{t=1}^{T}\left(\theta_{i,t} - \theta_{i,t-1}\right)^2\right)$$

We implement this model using a Gibbs Sampler, where the update for the unknown states $\theta_{0:T}$ is performed using the Forward Filtering Backward Sampling algorithm (FFBS), as detailed in Carter & Kohn [3]. Borrowing notation from Petris [2], the FFBS algorithm for sampling the states is

(1) Run the Kalman Filter.

(2) Draw $\theta_T \sim N_p\left(m_t, C_t\right)$

(3) For $t = T-1, \ldots, 0$: Draw $\theta_t \sim N_p\left(h_t, H_t\right)$ where

$$\begin{aligned} h_t &= m_t + C_t G_{t+1}' R_{t+1}^{-1}\left(\theta_{t+1} - a_{t+1}\right) \\ H_t &= C_t - C_t G_{t+1}' R_{t+1}^{-1} G_{t+1} C_t \end{aligned}$$

The Kalman Filter step is as follows.

Begin with $\theta_0 \sim N_p\left(m_0, C_0\right)$. For $t \geq 1$, let $\left(\theta_{t-1}|y_{1:(t-1)}\right) \sim N_p\left(m_{t-1}, C_{t-1}\right)$.

(i) The 1-step-ahead predictive distribution of $\left(\theta_t|y_{1:(t-1)}\right) \sim N_p\left(a_t, R_t\right)$ with

$$\begin{aligned} a_t &= E\left(\theta_t|y_{1:(t-1)}\right) = G_t m_{t-1} \\ R_t &= Var\left(\theta_t|y_{1:(t-1)}\right) = G_t C_{t-1} G_t' + W \end{aligned}$$

(ii) The 1-step-ahead predictive distribution of $\left(y_t|y_{1:(t-1)}\right) \sim N_1\left(f_t, Q_t\right)$ with

$$\begin{aligned} f_t &= E\left(y_t|y_{1:(t-1)}\right) = F_t a_t \\ Q_t &= Var\left(y_t|y_{1:(t-1)}\right) = F_t R_t F_t' + V \end{aligned}$$

(iii) The filtering distribution of $\left(\theta_t|y_{1:T}\right) \sim N_p\left(m_t, C_t\right)$ with

$$\begin{aligned} m_t &= E\left(\theta_t|y_{1:T}\right) = a_t + R_t F_t' Q_t^{-1} e_t \\ C_t &= Var\left(\theta_t|y_{1:T}\right) = R_t - R_t F_t' Q_t^{-1} F_t R_T \\ e_t &= y_t - f_t \text{ (forecast error)} \end{aligned}$$

Notice that because we implement a Gibbs Sampler, we sample V and W, and so the step to sample the states via FFBS treats V and W as known. Therefore our standard Gaussian Kalman Filter updates apply.

After estimating V, W, and the states $\theta_{1:T}$, we forecast future obserations at times $\{t+1, \ldots, t+K\}$. We set $K = 100$, to estimate approximately the next four days of energy demand. Begin by letting $a_t(0) = m_t$ and $R_t(0) = C_t$, where zero represents the current observed time. Then for $k \geq 1$, the following recursions hold.

$$
\begin{aligned}
(1)\ \theta_{t+k}|y_{1:T} &\sim N_p\left(a_t(k), R_t(k)\right), \text{with} \\
a_t(k) &= G_{t+k}at, k-1 \\
R_t(k) &= G_{t+k}Rt, k-1G'_{t+k} + W
\end{aligned}
$$

$$
\begin{aligned}
(1)\ y_{t+k}|y_{1:T} &\sim N_1\left(f_t(k), Q_t(k)\right), \text{with} \\
f_t(k) &= F_{t+k}a_t(k) \\
Q_t(k) &= F_{t+k}R_t(k)F'_{t+k} + V
\end{aligned}
$$

3. Results

The following prior distributions were used.

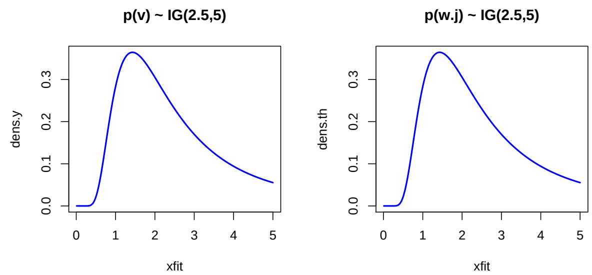

Figure 1: Prior densities for v^{-1} and w_j^{-1}

In-sample fit testing for a single coast still shows overfitting. Varying the hyperparameters did not significantly address overfitting; this is an action item for next steps.

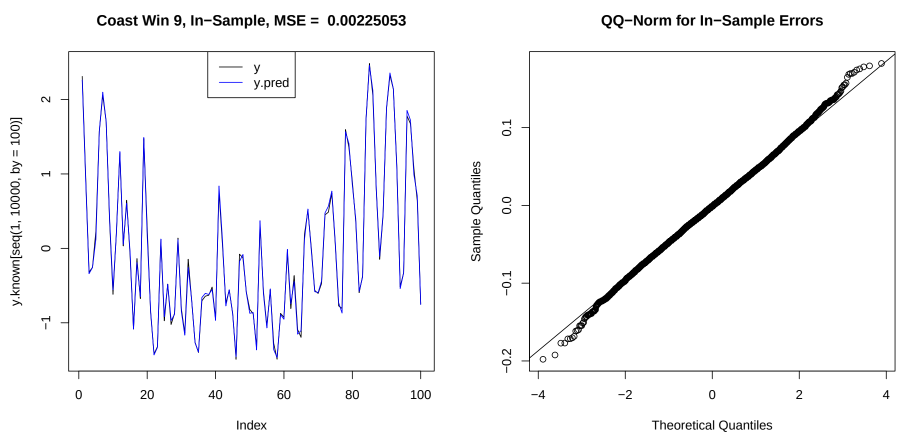

Figure 2: In-sample fit and residuals

Predictions were performed on each zone, with moving 'observed data' windows of 10,000 hours, predicting the next 100 hours. Results for each of these zones are not unreasonable, but tend to underestimate the load at peak times, which is when accuracy matters most. There is room for improvement here.
Forecasting detail with residual qq plot for a single window for the Coast zone.

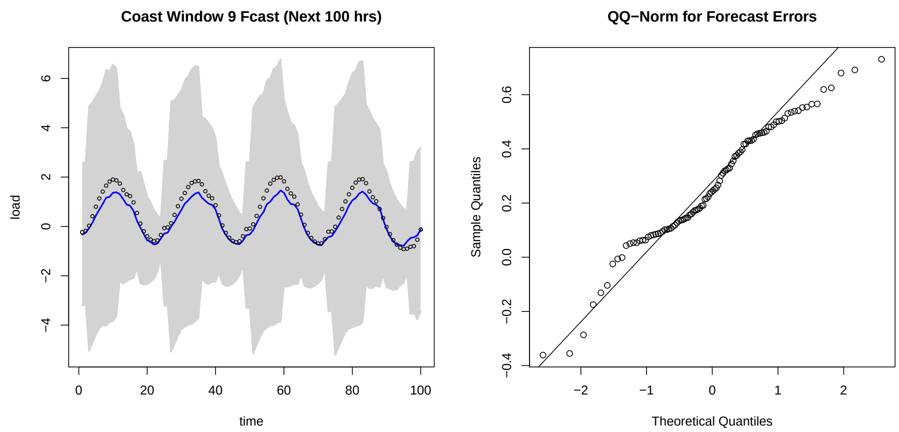

Figure 3: Coast forecast and residuals

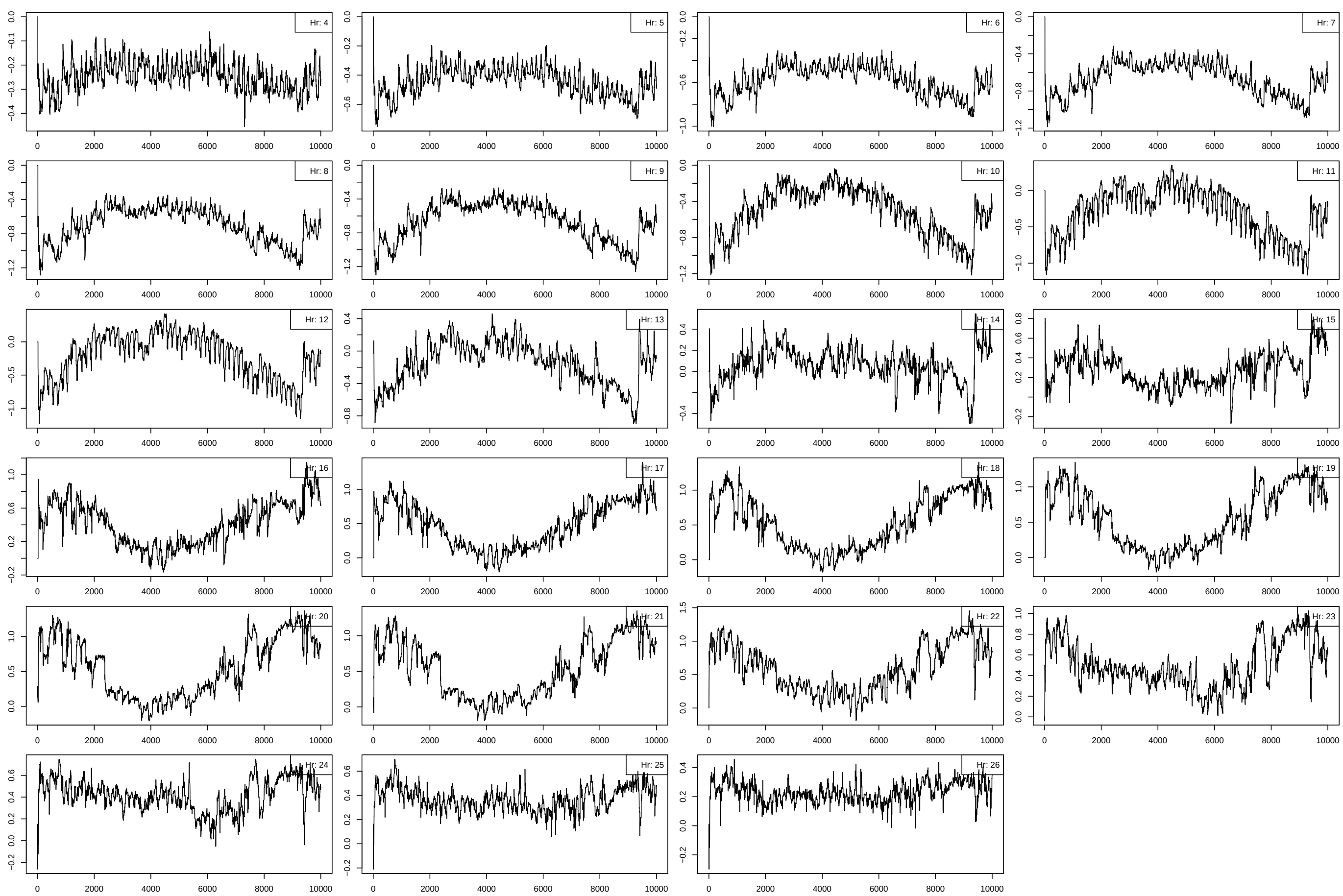

Figure 4: Coast hourly dummy variable trace plots

The results for all zones are below.

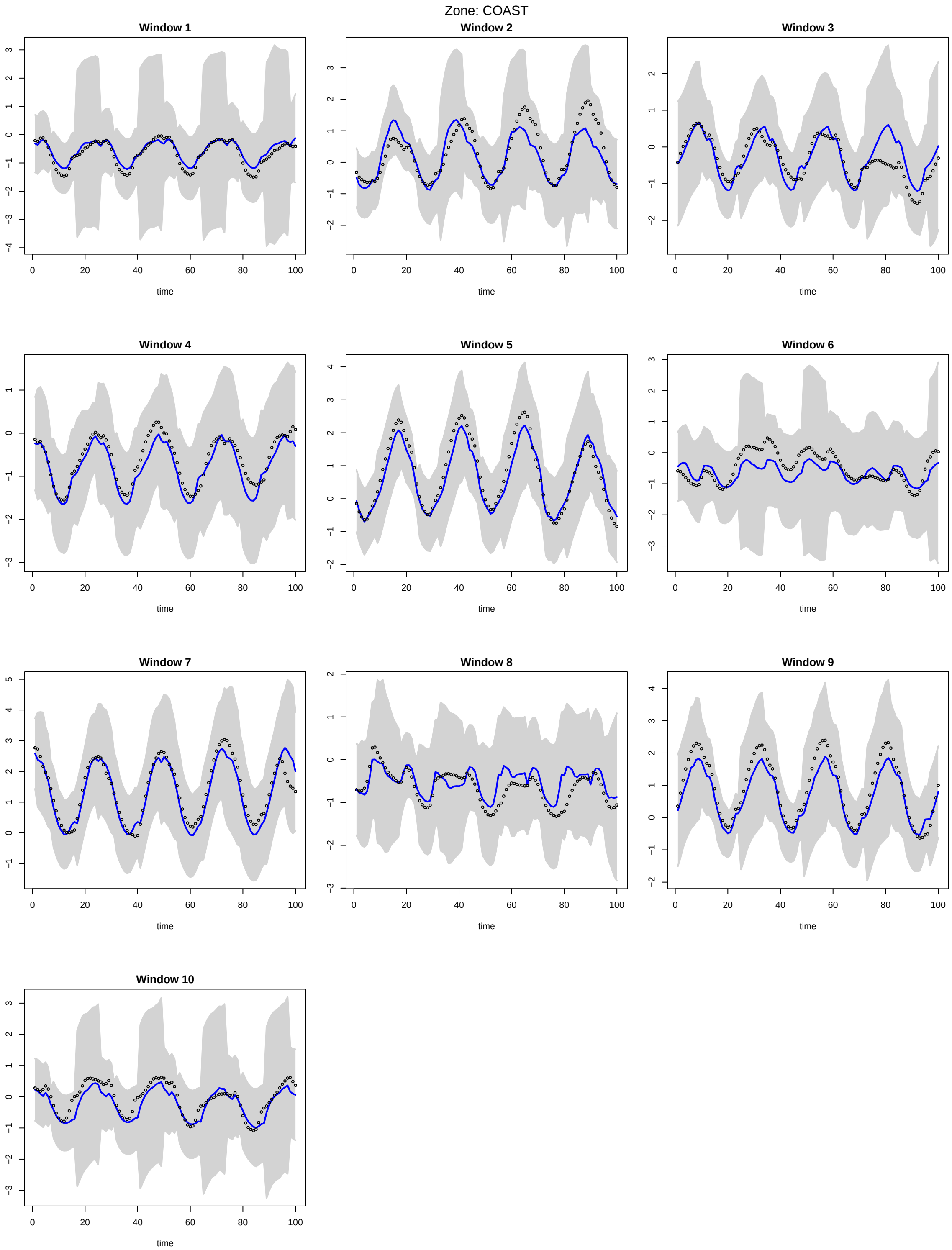

Figure 5: Moving window forecasts for Zone Coast

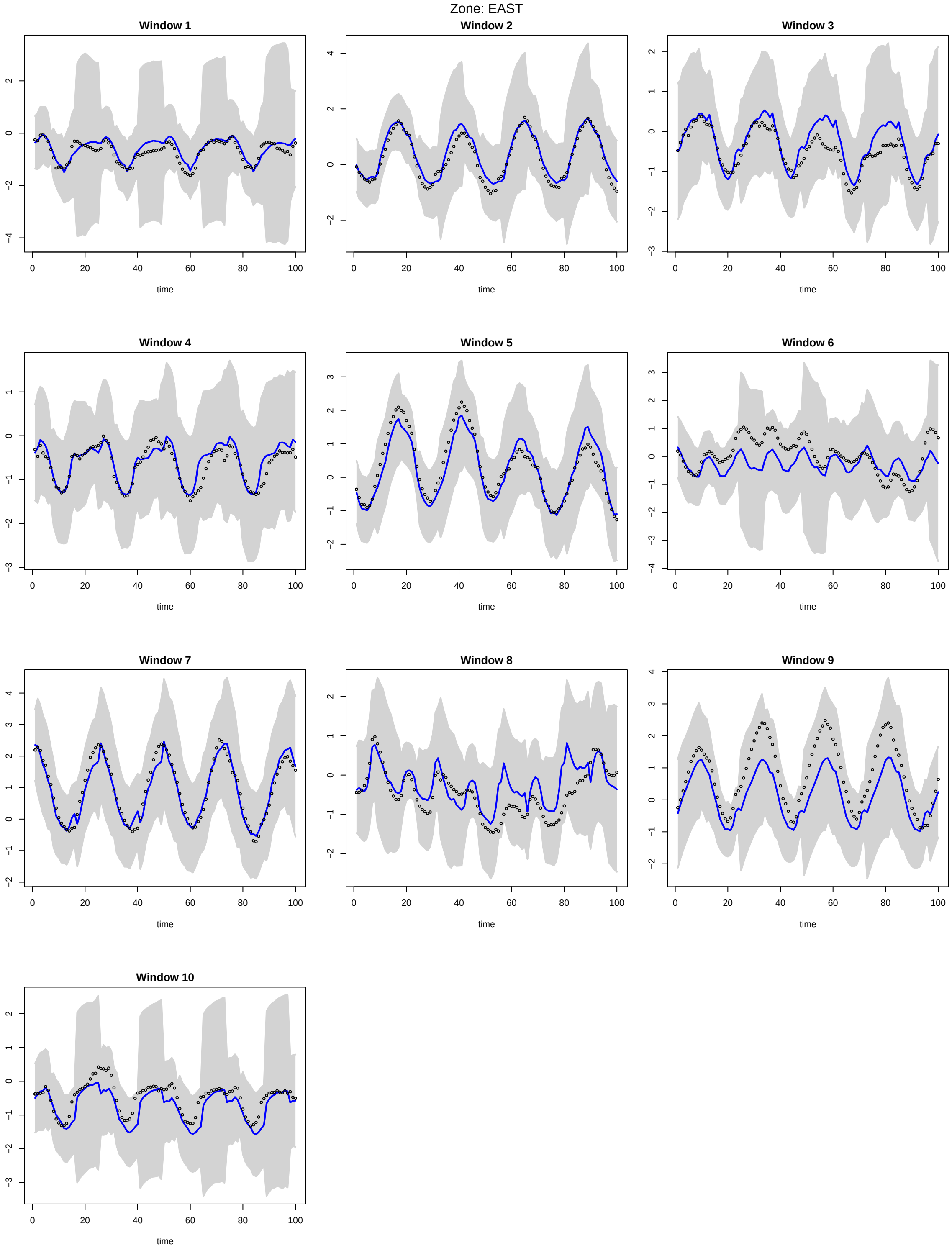

Figure 6: Moving window forecasts for Zone East

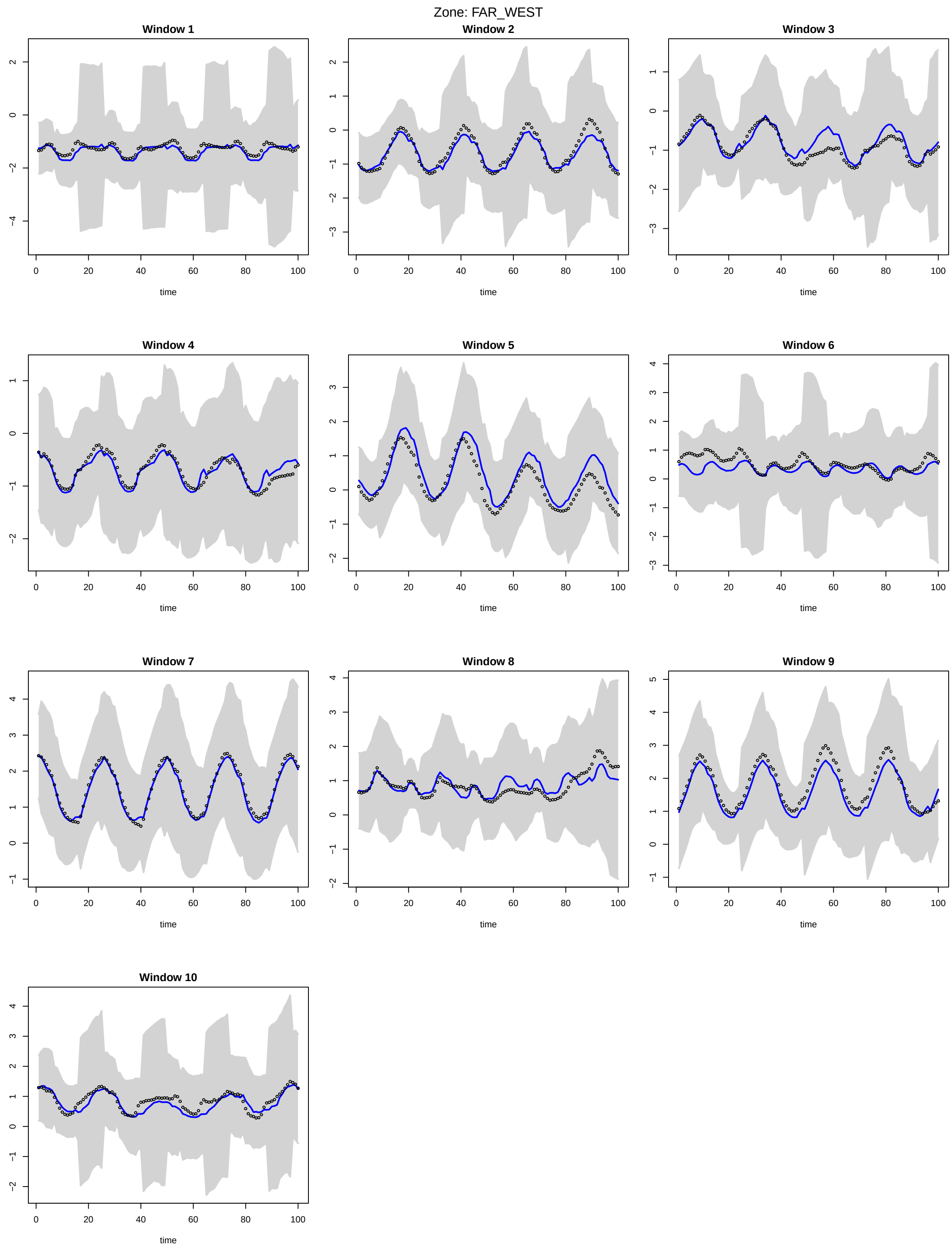

Figure 7: Moving window forecasts for Zone Far West

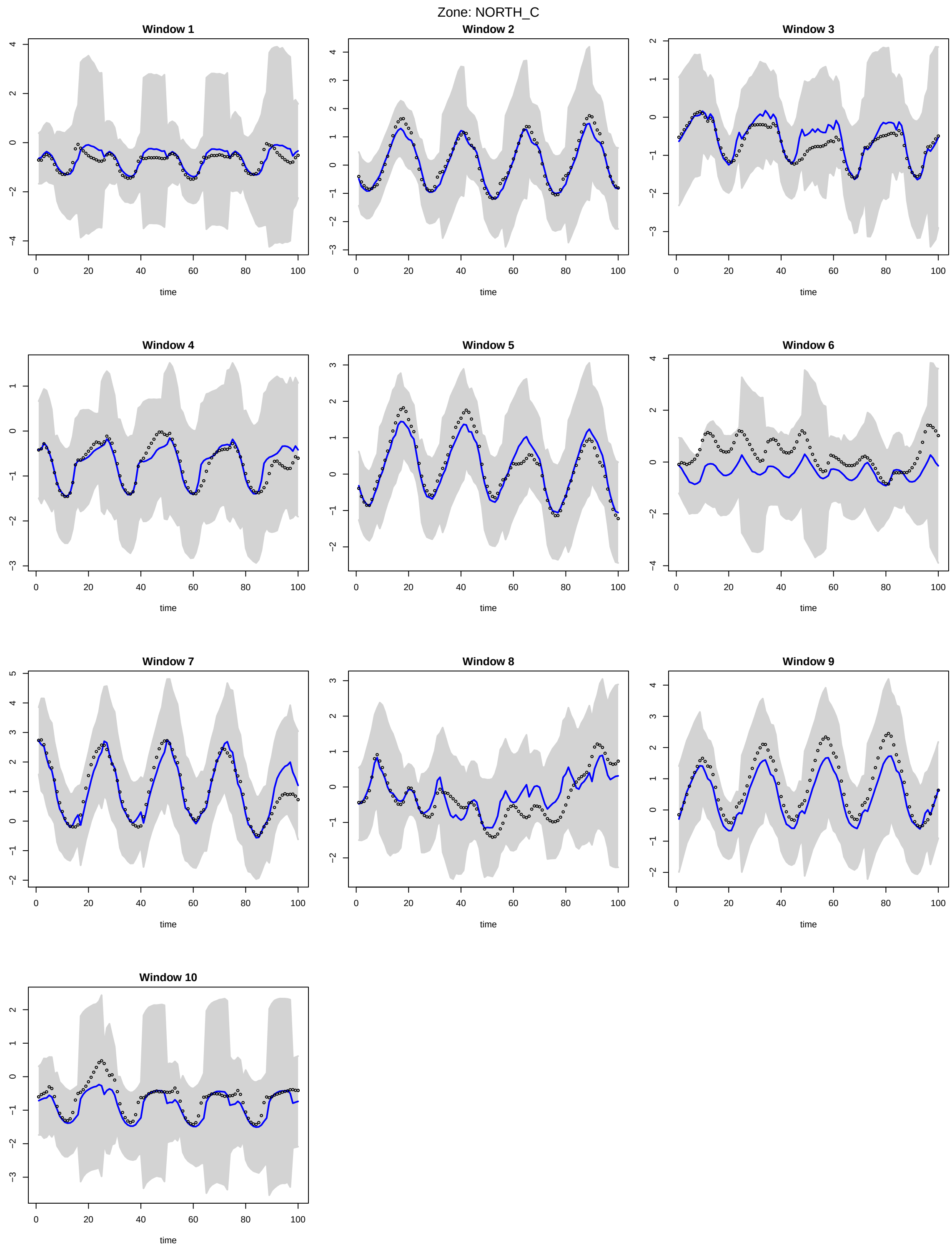

Figure 8: Moving window forecasts for Zone North Central

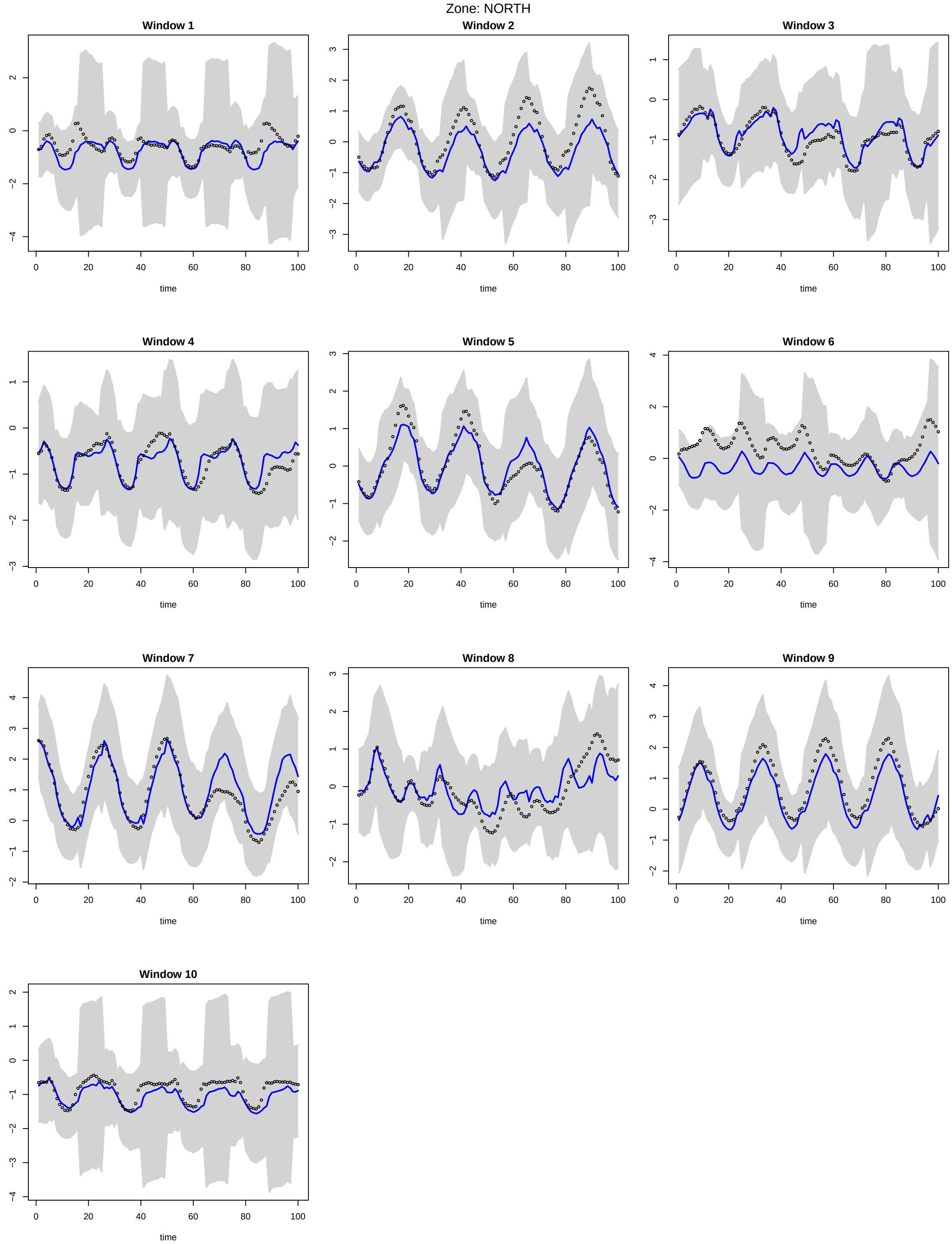

Figure 9: Moving window forecasts for Zone North

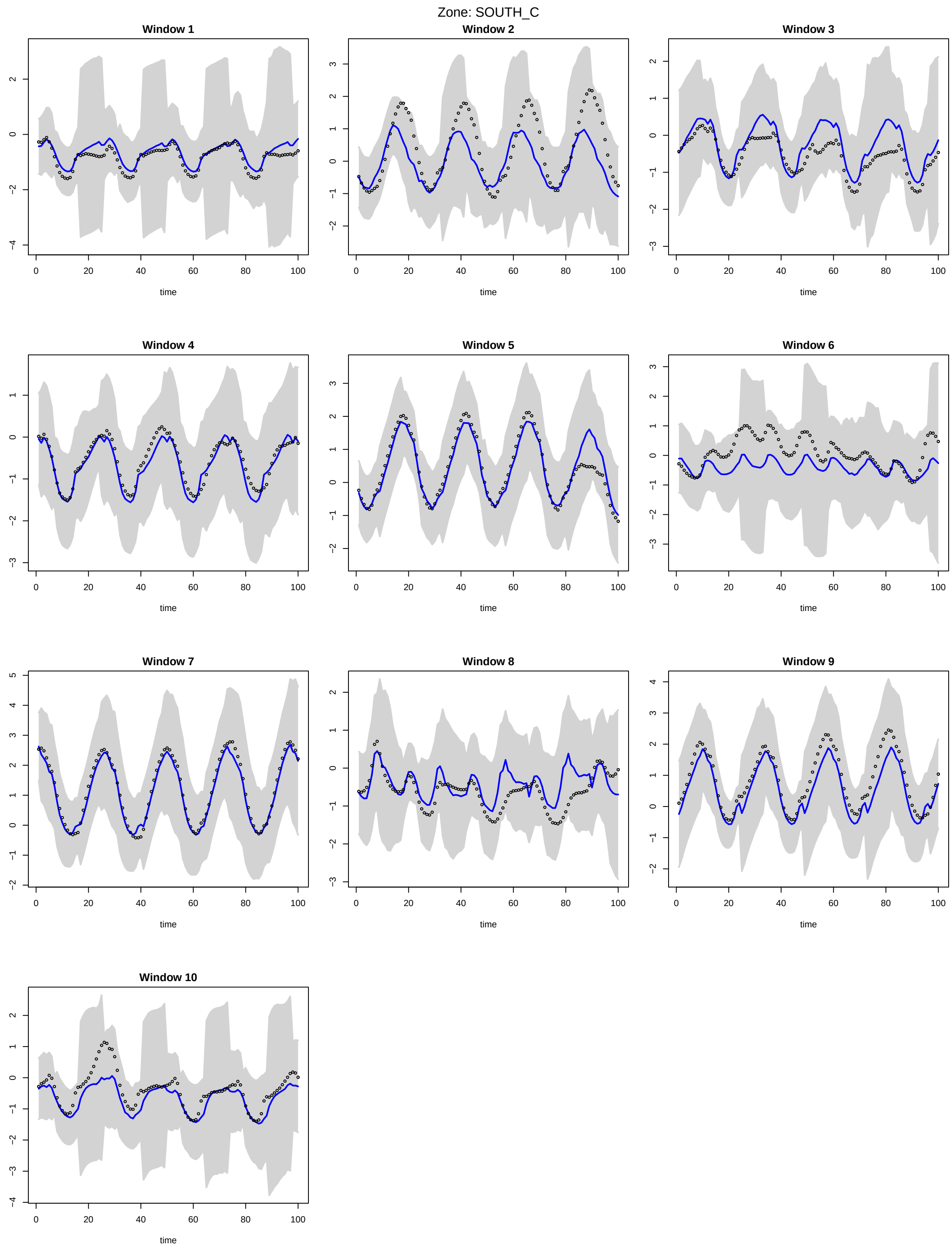

Figure 10: Moving window forecasts for Zone South Central

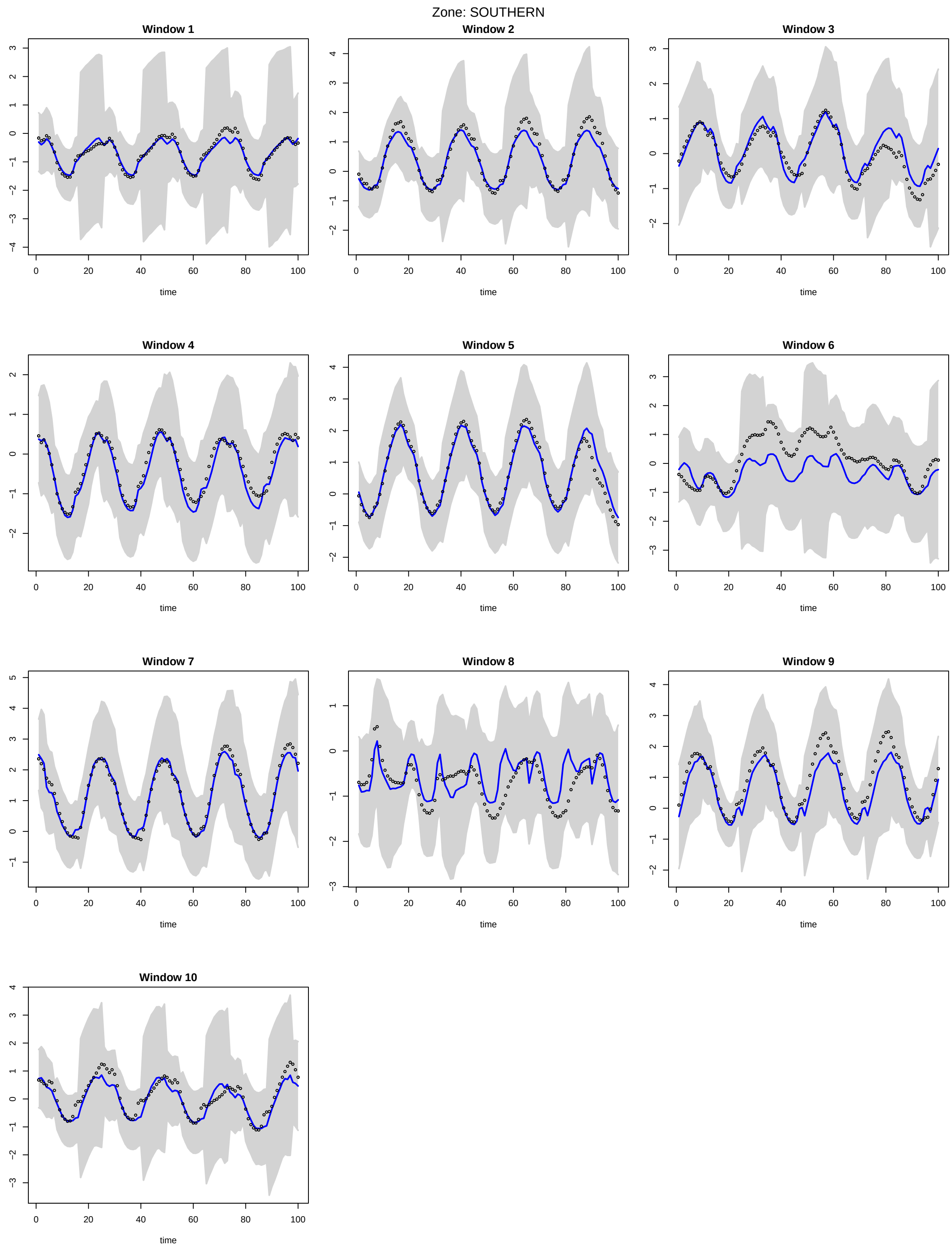

Figure 11: Moving window forecasts for Zone Southern

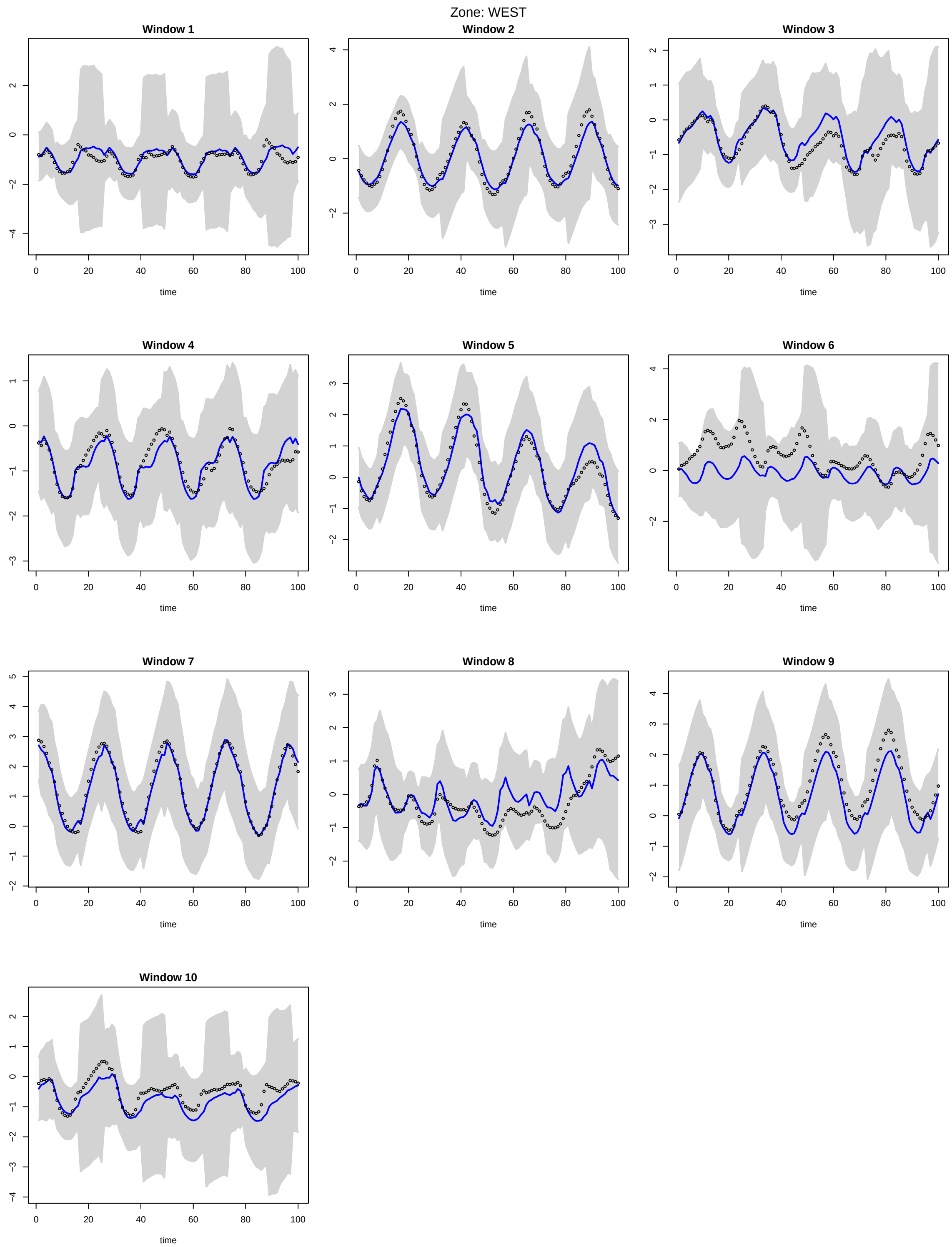

Figure 12: Moving window forecasts for Zone West

4. Discussion

Forecasts are within realm of reasonable, but require further fine-tuning. To do list:

[1] Address in-sample overfitting.

[2] Address under-estimation of peaks.

[3] Credible intervals are wider than desired. (Possibly tuning hyperparameters will help address?)

5. References

[1] West M, Harrison J. Baeysian Forecasting and Dynamic Models. 2nd ed. New York: Springer, 1997.

[2] Petris G, Petrone S, Campagnoli P. Dynamic Linear Models with R. New York: Springer Science + Business Media, 2009.

[3] Carter C, Robert K. 1994. On Gibbs sampling for state space models. Biometrika, 81, 541- 553.