Comparion Using Interpolation

Ziang Zhang

16/09/2021

Simulation Setting

In this simulation study, we aim to compare the performance of RW2 method with ARIMA method, considering that there are some covariate values with missing observations.

We consider two different types of settings for simulations. In the first type of setting, we fix the region of interest to $[0, 100]$, and vary the number of observations n between $\{10, 50, 100\}$ in that fixed interval. In the second type of setting, we fix the number of observations to $n = 50$, but vary the length of the region of interest. For simplicity, we consider the spacing between locations to be equal in all the simulation study.

The simulated data set has the form of $\{(x_i, y_i) : i \in [n]\}$, where x_i denotes the i-th (observed) covariate value and y_i denotes its corresponding observation. For each simulated data set, we assume that there exists a set of covariate values $\{z_i : i \in [n]\}$ being disjoint of the observed covariate values. To compare the two Bayesian smoothing methods, we consider both the inference at the observed values (i.e. $g(\boldsymbol{x}) := \{g(x_i) : i \in [n]\}$, and its interpolation values $g(\boldsymbol{z}) := \{g(z_i) : i \in [n]\}$.

To obtain samples of the interpolation values $g(\boldsymbol{z})$, we first draw samples $\tilde{g}(\boldsymbol{x})$ from the posterior of $g(\boldsymbol{x})$, then sample from the conditional distribution of $g(\boldsymbol{z})|\tilde{g}(\boldsymbol{x})$ given by the prior distribution.

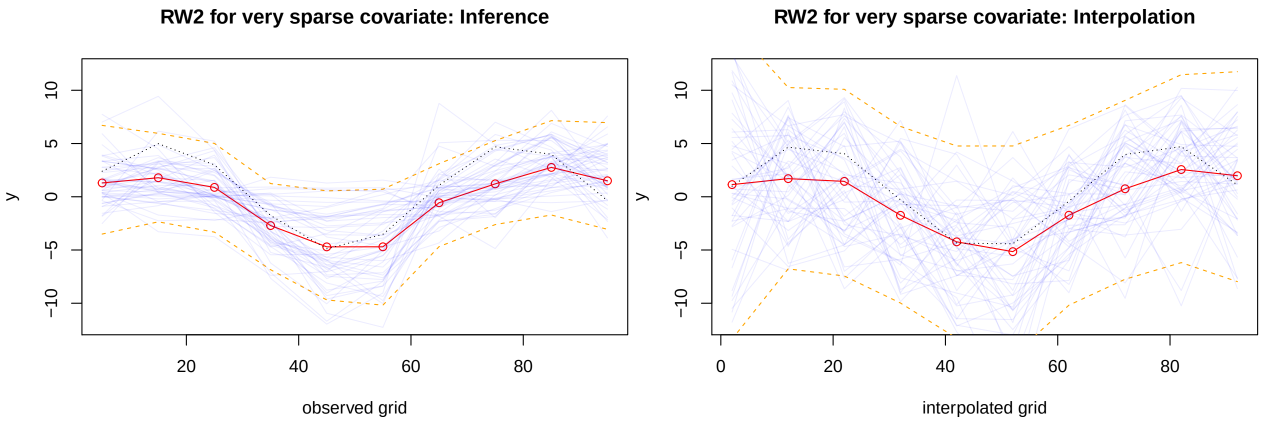
For the true function $g(.)$, we consider it being the function

$$g(x) = 5\sin(0.1x),$$

observed at $x \in [0, 100]$.

Sample Size being 10

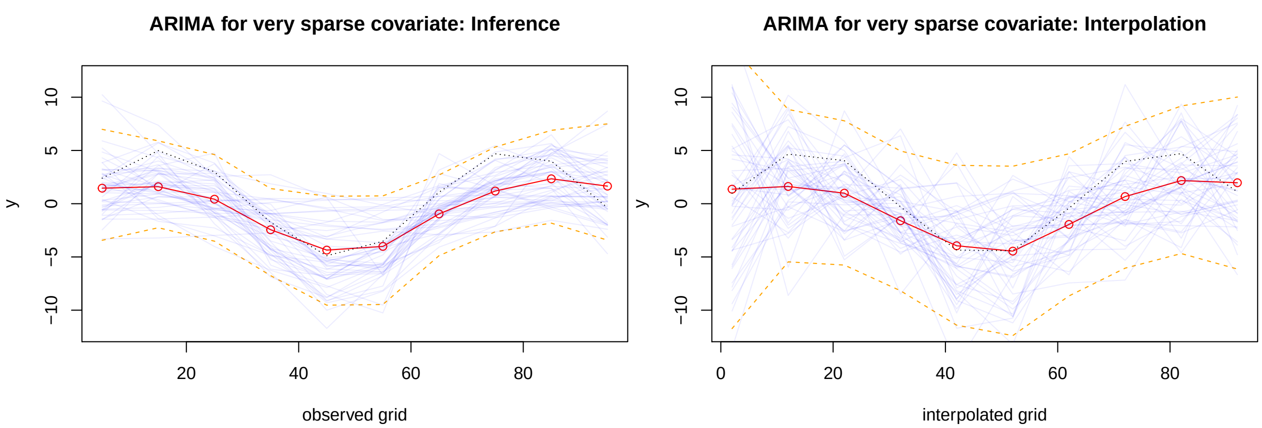
RW2 Method



The inferred results using RW2 method are summarized at above. The red line represents the posterior mean of $g(\boldsymbol{x})$ (or $g(\boldsymbol{z})$), the orange lines represent its 95 % credible interval, and the light blue lines are the sample paths simulated from the posterior distribution of $g(\boldsymbol{x})$ (or $g(\boldsymbol{z})$). The black dotted line is the true function.

For both inferences, we computed the mean squared error (MSE) using posterior mean and mean credible width (MCW) using their 95% intervals. For the observed locations, we have MSE being 3.805 and MCW being 9.062. For the interpolated locations, we have MSE being 3.563 and MCW being 18.942.

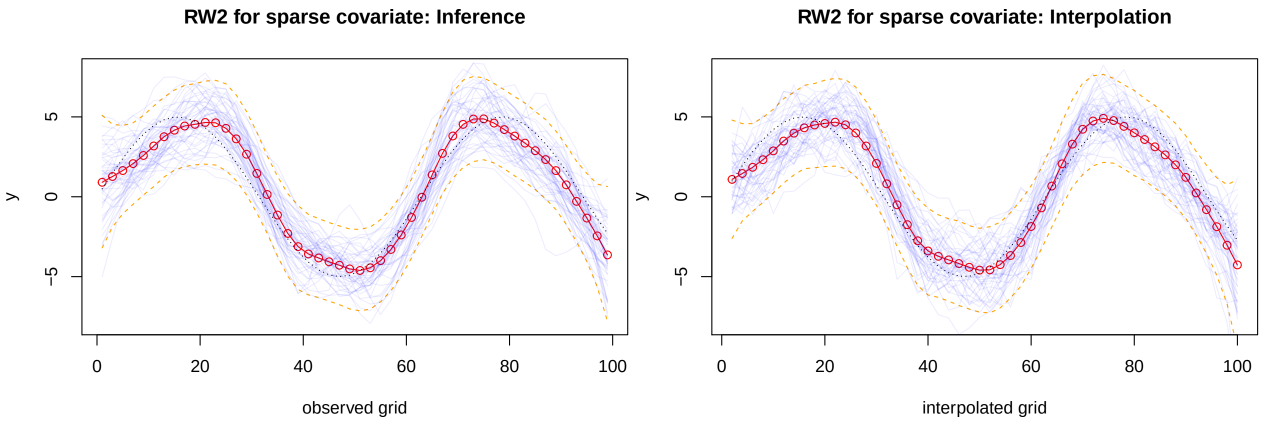
ARIMA Method



For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 4.319 and MCW being 9.05. For the interpolated locations, we have MSE being 4.094 and MCW being 15.501.

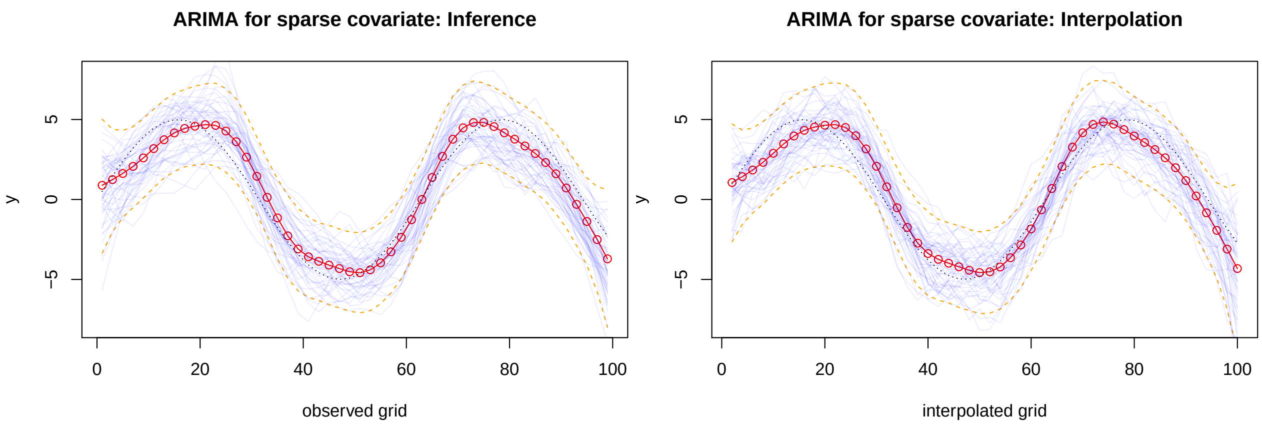
Sample Size being 50

RW2 Method



For the RW2 approach, inference of $g(\boldsymbol{x})$ has MSE being 0.7 and MCW being 5.277. For the interpolated locations, we have MSE being 0.724 and MCW being 5.541.

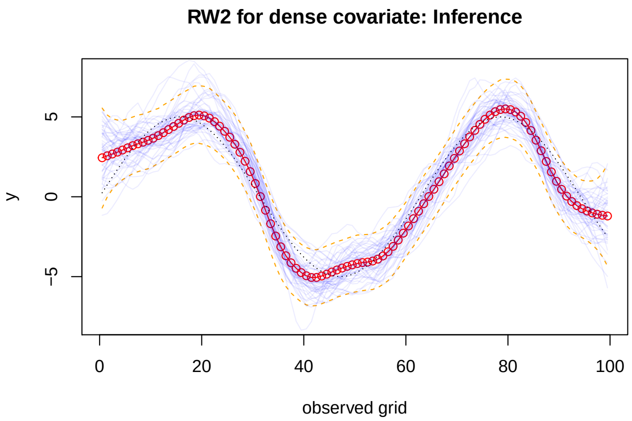
ARIMA Method

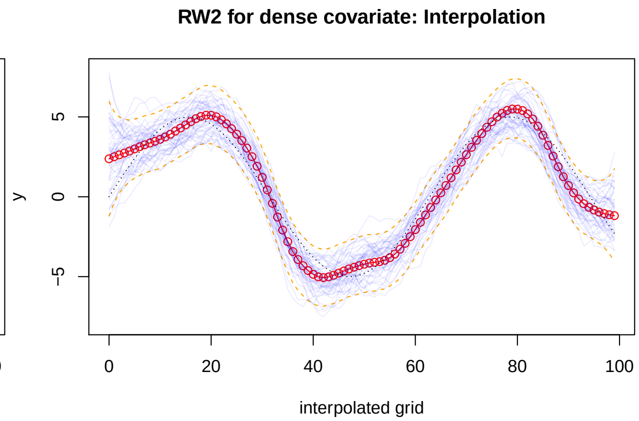


For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 0.703 and MCW being 5.226. For the interpolated locations, we have MSE being 0.728 and MCW being 5.369.

Sample Size being 100

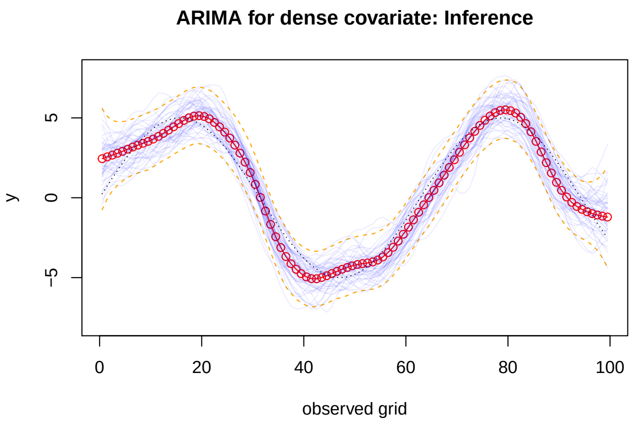
RW2

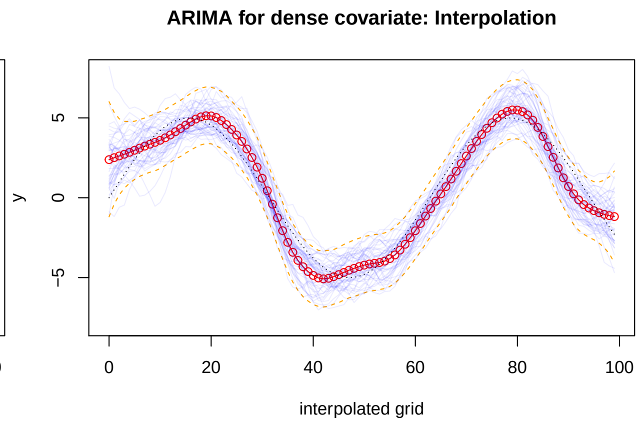




For the RW2 approach, inference of $g(\boldsymbol{x})$ has MSE being 0.579 and MCW being 3.664. For the interpolated locations, we have MSE being 0.599 and MCW being 3.714.

ARIMA

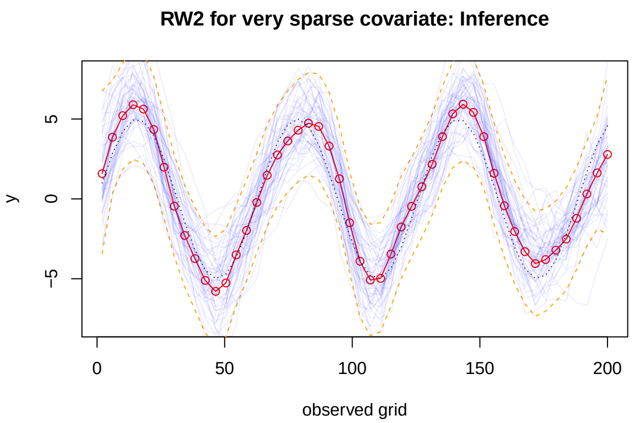


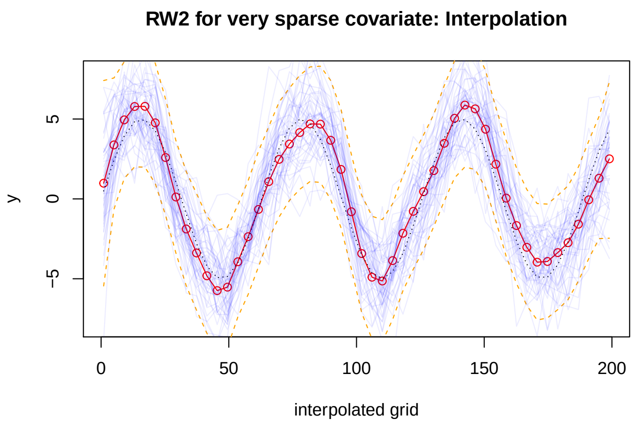


For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 0.58 and MCW being 3.653. For the interpolated locations, we have MSE being 0.601 and MCW being 3.68.

Study Region [0,200]

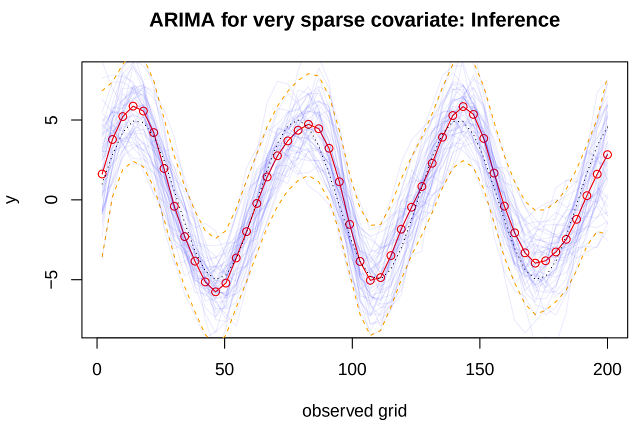
RW2

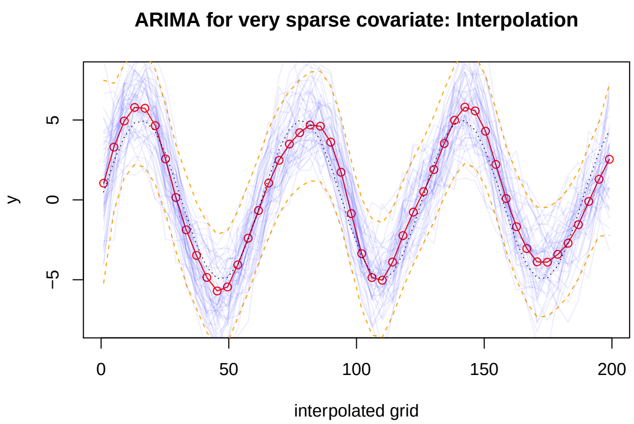




For the RW2 approach, inference of $g(\boldsymbol{x})$ has MSE being 0.8 and MCW being 6.678. For the interpolated locations, we have MSE being 0.785 and MCW being 7.469.

ARIMA

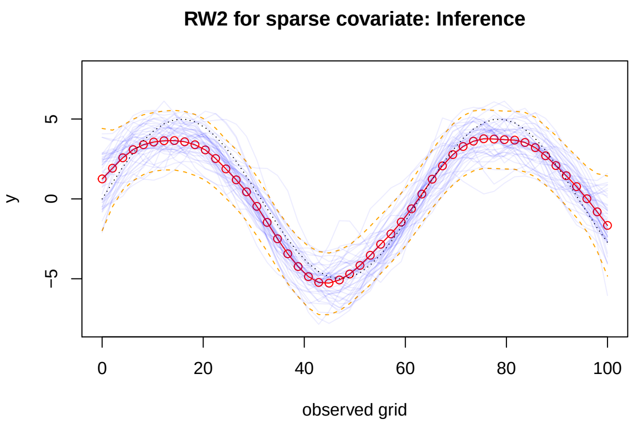


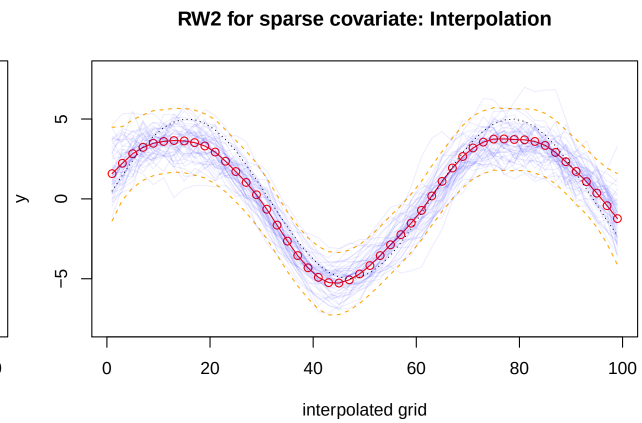


For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 0.757 and MCW being 6.592. For the interpolated locations, we have MSE being 0.746 and MCW being 7.032.

Study Region [0,100]

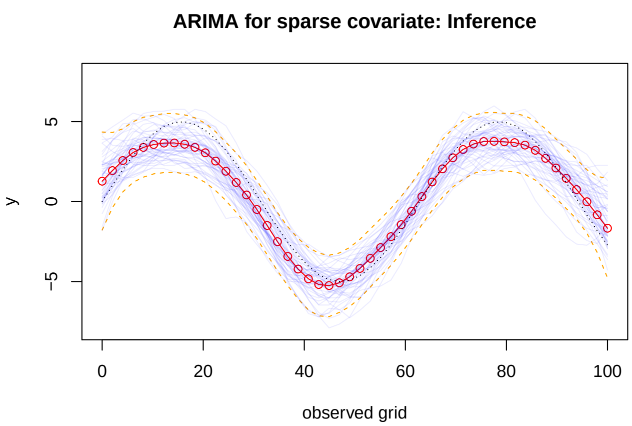
RW2

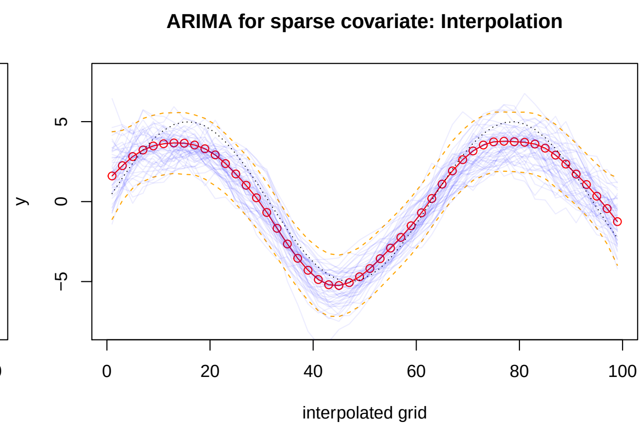




For the RW2 approach, inference of $g(\boldsymbol{x})$ has MSE being 0.704 and MCW being 3.88. For the interpolated locations, we have MSE being 0.689 and MCW being 3.995.

ARIMA

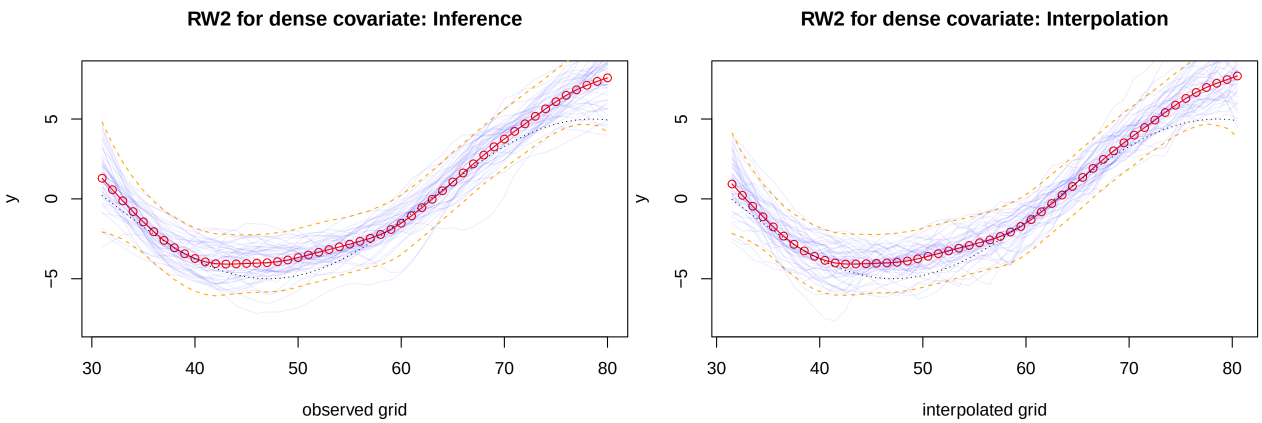




For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 0.697 and MCW being 3.832. For the interpolated locations, we have MSE being 0.681 and MCW being 3.859.

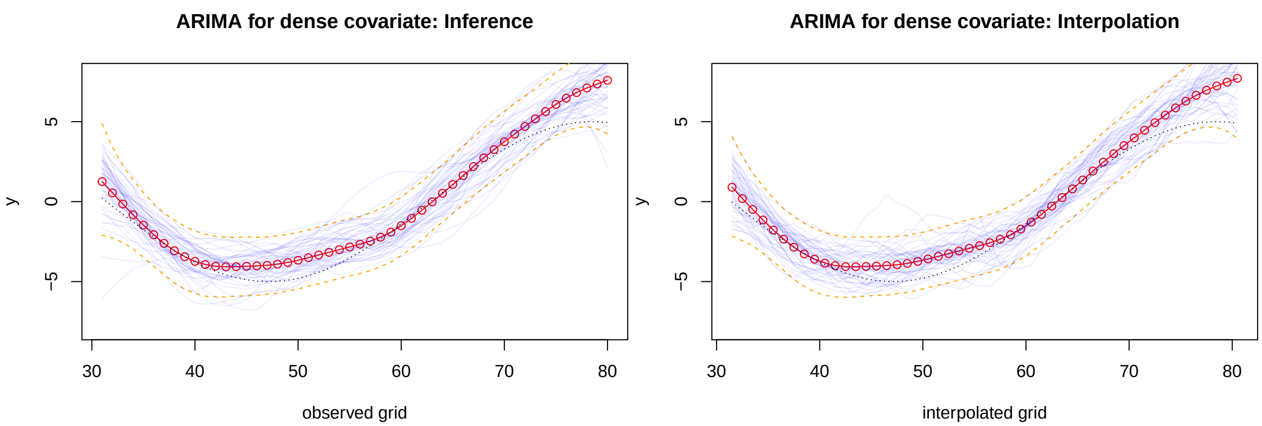
Study Region [30,80]

RW2



For the RW2 approach, inference of $g(\boldsymbol{x})$ has MSE being 0.868 and MCW being 3.996. For the interpolated locations, we have MSE being 0.939 and MCW being 4.045.

ARIMA



For the ARIMA approach, inference of $g(\boldsymbol{x})$ has MSE being 0.865 and MCW being 3.987. For the interpolated locations, we have MSE being 0.937 and MCW being 4.014.

Conclusion:

From the first three set of simulation studies (where n is changing from small to large), we can make the observation that, the two Bayesian smoothing methods perform similar in terms of MSE, but the ARIMA method gives smaller values of MCW. This difference gets larger as the sample size n declines.

The same conclusion is also supported by the second sets of simulation studies (where the region of interest is changing). Unless the sample size is large enough and the spacing between locations is small, ARIMA method yields more favorable inferential result than the RW2 method.