Master's Paper of the Department of Statistics, the University of Chicago
(Internal document only, not for circulation)

Statistical Analysis of the United States' Summer Temperatures

YE TIAN

Advisor(s): Michael Stein
(at least one statistics faculty member)

Approved _____

Date _____

February-15, 2018

Abstract

In this paper, we discuss the method of modeling the simulated temperature data. We first introduce our methods of stationarizing the raw data. Then we explain how to use the artificial neuron network estimate the conditional mean function of the normalized data and the visualization of the function. We then reduced the model to AR model and evaluate the loss of the reduction. We also introduce the long-range dependence into our model, build the FAR model, compare their performances on the normalized data and the raw data. Finally, we talk about the lengths of the predictability of the two models.

Contents

1 Introduction

1.1 Focus and motivation

In public imagination, heat waves may remain a B-list natural disaster. But, considering 700 dead in Chicago in 1995, 50,000 across Europe in 2003 and 11,000 in Russia in 2010, heat waves have become the deadliest weather events, and they can have big impacts on farming, energy use, and other critical aspects of society. However, what if we could foresee the heat coming ahead of time? If forecasters could predict heat waves much more in advance, giving emergency planners more time to prepare, would more people survive, fewer treasures lost? That's the motion for my research. In this paper, we aim to build a model to describe the dynamic of the temperatures in summer, through which we may be able to have an early alarm of heat waves.

1.2 Brief introduction to the dataset

We apply our methods to an ensemble of 50 historical/future simulations from the Community Earth System Model (CESM). The atmospheric component is the low-resolution Community Atmosphere Model version 4, with T31 spectral resolution ($\sim 3.75^{\circ} \times 3.75^{\circ}$) and 26 vertical levels. The ensemble is based on a ~10,000 year pre-industrial control simulation. After a ~ 4000 year spin-up using constant preindustrial conditions, they initialize 50 historical hindcasts (1850-2005) from snapshots of the coupled model state taken every 100 years. So, the last hindcast is initialized after approximately 9000 years of the control simulation. They then extend each hindcast to 2100 using the Representative Concentration Pathway (RCP) 8.5 scenario. (The RCPs are consistent with a wide range of possible changes in future anthropogenic (i.e., human) greenhouse gas (GHG) emissions, and aim to represent their atmospheric concentrations. RCP8.5 is a so-called 'baseline' scenario that does not include any specific climate mitigation target. The greenhouse gas emissions and concentrations in this scenario increase considerably over time, leading to a radiative forcing of 8.5 W/m2 at the end of the century. More information can be found in ref. [1].) The 100-year gap between each new initialization ensures nearly independent ensemble members that fully capture internal variability within the coupled system. There is no difference of the common year and the leap year in the model, so, each year has 365 days. More information about the model and ensemble design can be found in ref. [2].

In this paper, we focus on some typical points for three representative parts of the United States:

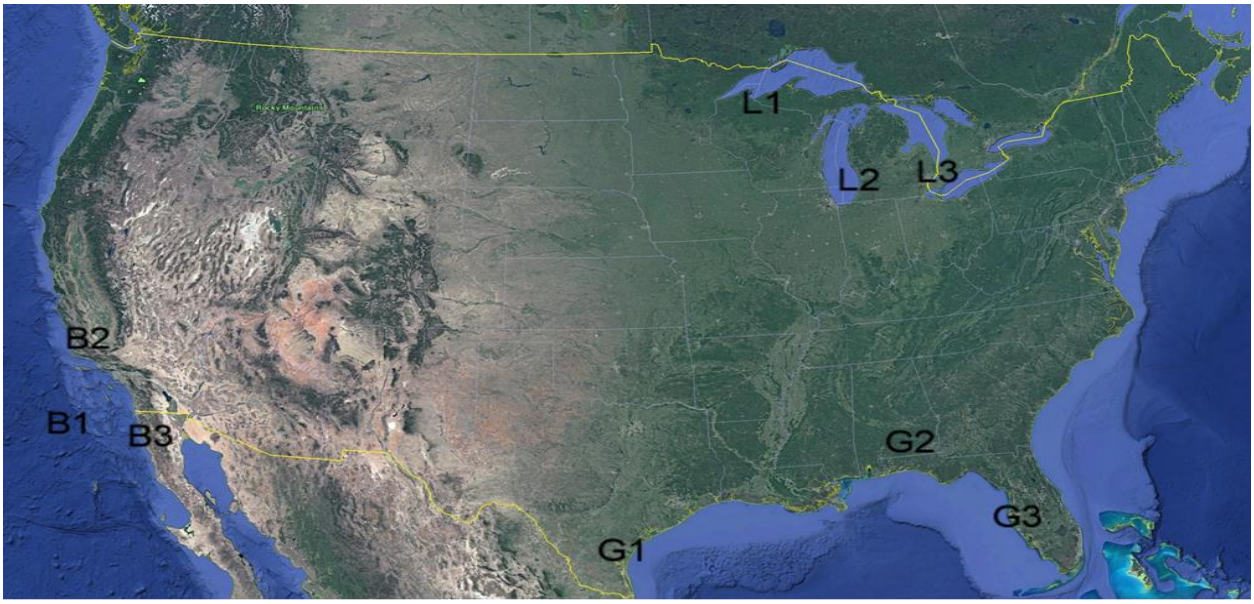


Figure 1.2.1 The map shows the locations

B_1(240˚E,31.54˚N), B_2(240˚E,35.26˚N), B_3(243.75˚E,31.54˚N) for the California Coast;
L_1(270˚E,46.39˚N), L_2(273.75˚E,42.68˚N) ,L_3(277.5˚E,42.68˚N) for the Great Lakes; and
G_1(262.5˚E,27.83˚N), G_2(273.75˚E,31.54˚N), G_3(277.5˚E,27.83˚N) for the Gulf of Mexico.

As time goes by, the side effects of human activities, like greenhouse effects, on the climate would be increasingly significant so that the dynamic of the climate could change a lot. So, we only use the data from 2050 to 2099, the last 50 years of the model runs, during which the human effects on climate will be largest.

1.3 the structure of the paper

In Chapter 2, we introduce the processing of the data. In Chapter 3, we describe how we decide the lag order of the conditional mean function. In Chapter 4, we reduce the conditional mean function to linear form and build *AR* models. In Chapter 5, we introduce long-range dependence, building *FAR* models to see whether including this dependence improves model fit and predictions. In the last chapter, we briefly summary our conclusions.

2 Stationarization of the data

The temperature data are non-stationary. We need to normalize the data first.

2.1 Model formula

We introduce some notations first and then give the model formula. We use a pair of variables $t=(y, d)$ to represent the time index, where d is the day of the year, spanning from 1 to 365 and $y=year+\frac{d}{365}$, year is the real year of the time, from 2050 to 2099. To make the variable y more continuous, we add the term $\frac{d}{365}$. We suppose the mean, m, and variance, v (We use log$v$ to represent the logarithm of v.), of the raw temperatures $\overline{\overline{T}}(t)$ are functions of both the trend and periodic components. We assume the trend of the mean function can be represented as a set of basis functions of y, $\{D_i^{m_y}(y)\}$, a set of basis functions of d, $\{D_j^{m_d}(d)\}$ and their interactions $\{D_i^{m_y}(y)D_j^{m_d}(d)\}$. We consider periodic components as a function of the variable d. The set of basis functions $\{P_i^m(d)\}$ we use are harmonic functions with the following form:

$$P_{2i-1}^m(d)=\sin(\frac{2\pi\cdot d\cdot i}{365}),\ P_{2i}^m(d)=\cos(\frac{2\pi\cdot d\cdot i}{365}),\qquad i=1, 2, 3, \cdots \tag{2.1.1}$$

Thus, the mean function $m(y, d)$ has the following forms:

$$m(y,d)= \alpha_m+ \sum_i \beta_i^m(y)D_i^{m_y}(y) + \sum_j \gamma_j^m(d)D_j^{m_d}(d) + \sum_{i,j} \rho_{ij}^m(d)D_i^{m_y}(y)D_j^{m_d}(d) + \sum_k [a_k^m(d)P_{2k-1}^m(d)+b_k^m(d)P_{2k}^m(d)] \tag{2.1.2}$$

where a_k^m and b_k^m are all linear combinations of periodic B-splines of d. For the logarithm of variance function $logv(t,d)$, we use a similar form:

$$logv(y,d)= \alpha_v+ \sum_i \beta_i^v(y)D_i^{v_y}(y) + \sum_j \gamma_j^v(d)D_j^{v_d}(d) + \sum_{i,j} \rho_{ij}^v(d)D_i^{v_y}(y)D_j^{v_d}(d) + \sum_k [a_k^v(d)P_{2k-1}^v(d)+b_k^v(d)P_{2k}^v(d)] \tag{2.1.3}$$

2.2 Model estimation

We suppose the random parts are independent and identically distributed normal random variables with mean 0, variance 1. So, $\overline{\overline{T}}(t) \sim N(m(y,d), v(y,d))$; denote by M_m and M_{lv} the design matrices of m and log$v$, separately; define logv_s as the logarithm of sample variances of 50 different runs. Then we use the following procedure to normalize the data:

(1) Use ordinary least square approach to estimate the linear model $m=M_m\beta_m+\varepsilon_m$, and use the estimated coefficients as the initial values for β_m.

(2) Use least square approach to estimate the linear model $\log v_s = M_{lv}\beta_{lv} + \varepsilon_{lv}$, and use the estimated coefficients as the initial values for β_{lv}.

(3) Get the MLE $\hat{m}$ and $\hat{v}$, and normalize the data

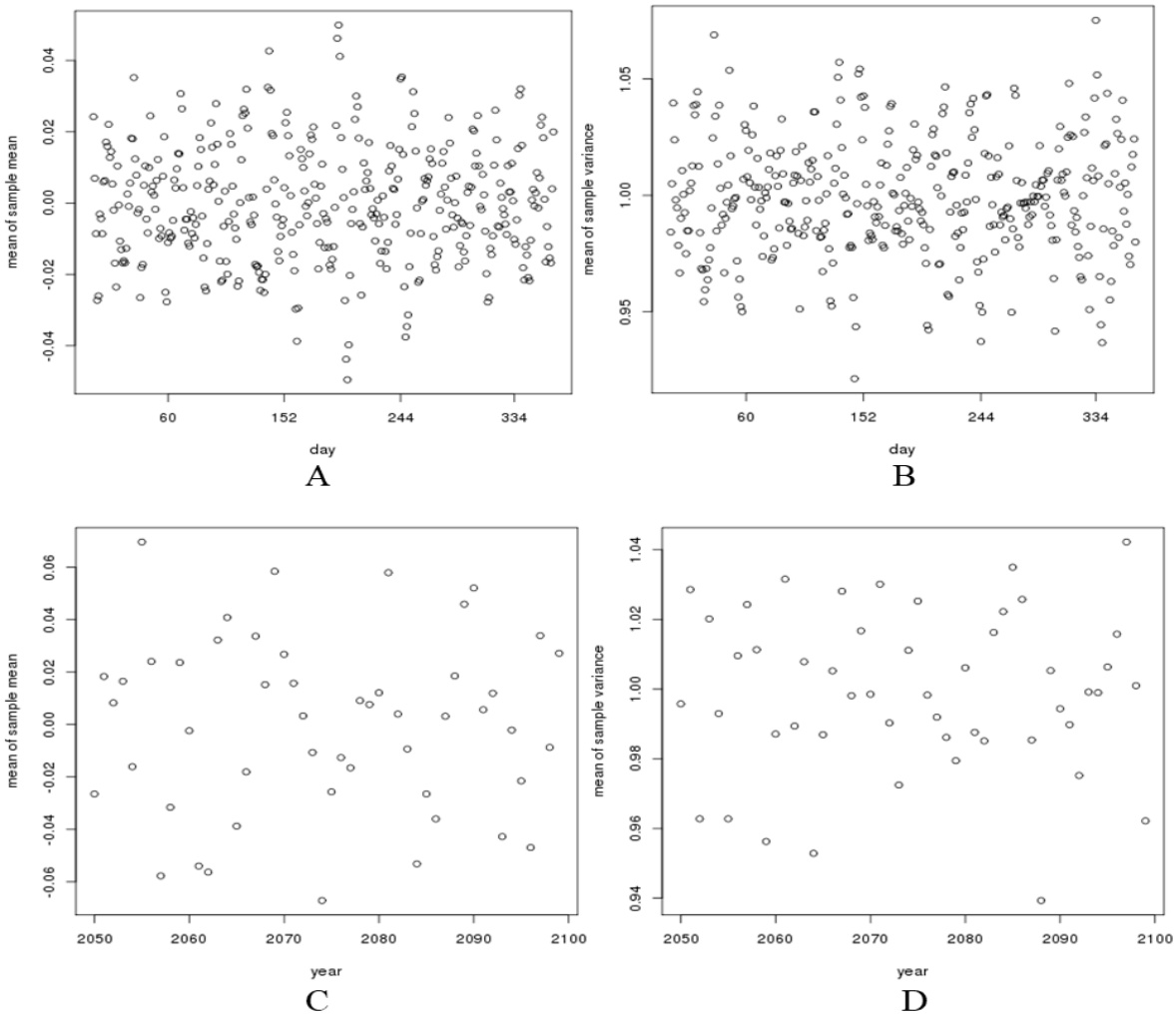
$$T(t) = \frac{\bar{\bar{T}}(t) - \hat{m}(t)}{\sqrt{\hat{v}(t)}} \tag{2.2.1}$$

for the following procedures.

In the following text, we denote $T(t)$ by T_t if it is not ambiguous.

2.3 Model selection

We use B-splines for $\{D_i^{m_y}(y)\}$, $\{D_j^{m_d}(d)\}$, $\{D_i^{v_y}(y)\}$, and $\{D_j^{v_d}(d)\}$ with no internal breakpoint, and add additional breakpoints if it is necessary. For the basis functions of a_k^m, b_k^m, a_k^v and b_k^v, we use periodic B-splines, start with no breakpoint, and add additional breakpoints according to the fitting. For the harmonic functions, we start from order 1, i.e., 2 harmonic functions and gradually increase the order until there is no apparent periodicity.



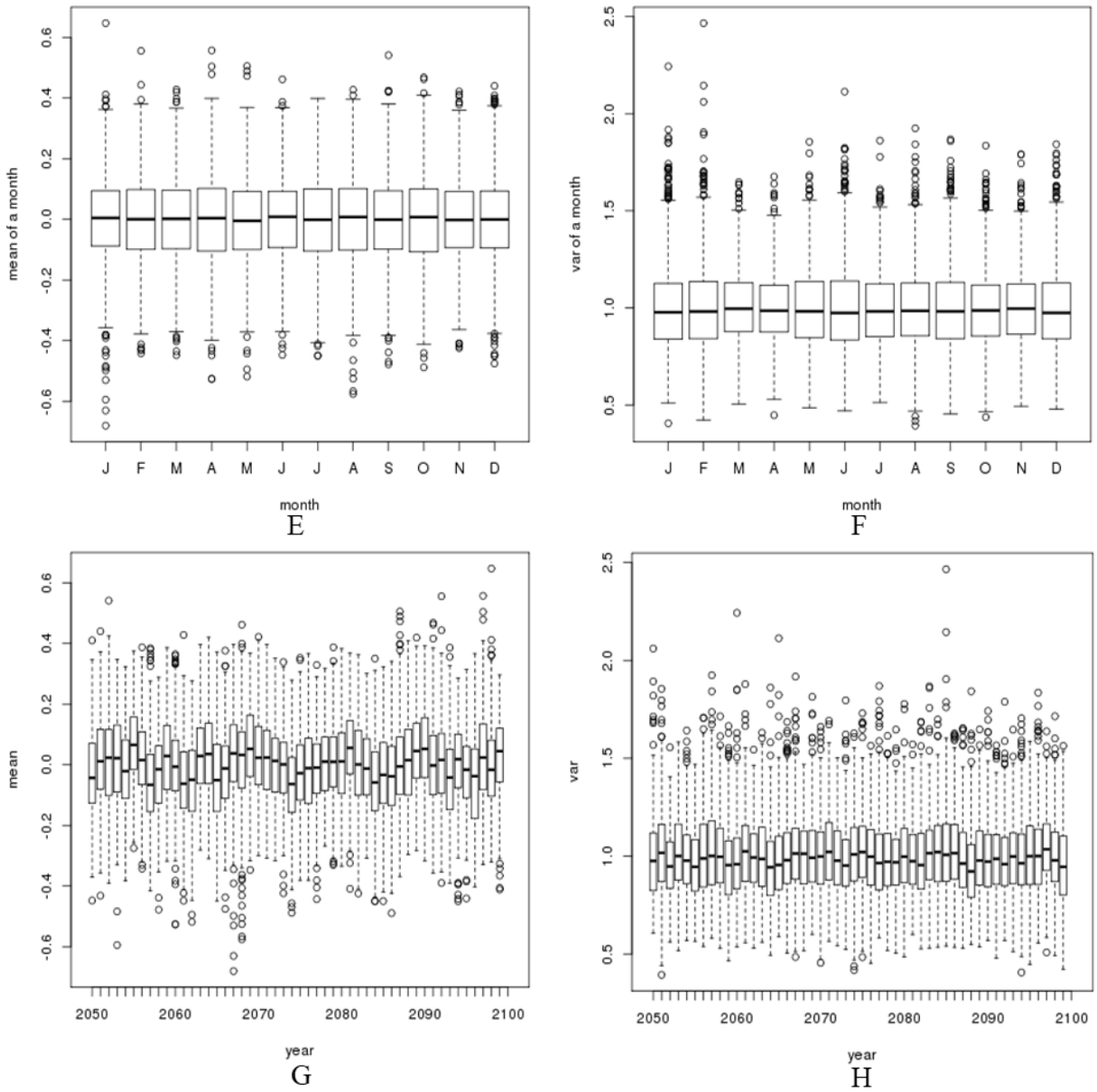


Figure 2.3.1

Plots for T_t of the point B_1. For each day on the horizontal axis in A, the corresponding vertical coordinate is the average of the temperatures through 50 years and 50 runs. For each point in C, we first calculate the sample average of 50 runs for all $t(y, d)$, then for each fixed year we compute the average of the 365 averaged temperatures of that year. In E, we consider all the points (50 runs multiple 50 years) of a month as a group and make the boxplots for 12 months. In G, we consider all the points (50 runs multiple 365 days) of a year as a group and make the boxplots for 50 years. The subfigures B, D, F, H are those for the variance corresponding to A, C, E and G, respectively. All the means should be around 0 and variances 1.

We select the model based on 8 plots of T_t (Figure 2.3.1). Based on subfigures A and B, we can justify whether we need to adjust the basis functions $\{D_j^{m_d}(d)\}$, $\{D_j^{v_d}(d)\}$, $\{P_i^m(d)\}$ and $\{P_i^v(d)\}$. We increase the number of basis functions in $\{D_j^{m_d}(d)\}$ and $\{D_j^{v_d}(d)\}$ until there is no apparent trend within the 365 days

and increase the number of basis functions in $\{P_i^m(d)\}$ and $\{P_i^v(d)\}$ until there is no periodicity. C and D are used for selecting $\{D_i^{m_y}(y)\}$ and $\{D_i^{v_y}(y)\}$, we increase the number of basis functions until there is no apparent trend within the 50 years. E, F, G and H are used to determine the order of interaction $\{D_i^{m_y}(y)D_j^{m_d}(d)\}$ and $\{D_i^{v_y}(y)D_j^{v_d}(d)\}$. For E, there should be no trend among the 12 months, if there is certain trend in E, it means the if we remove the effect of y (or we consider the effect of y are the same), there is still certain trend in d, so, we should increase the order of $D_j^{m_d}(d)$. Similar for F, if there is certain trend in F, we should increase the order of $D_j^{v_d}(d)$. For G, there should be no trend among the 50 years, if there is certain trend in G, it means the if we remove the effect of d (or we consider the effect of d are the same), there is still certain trend in y, so, we should increase the order $D_i^{m_y}(y)$. Similar for H, if there is certain trend in H, we should increase the order of $D_i^{v_y}(y)$. Plots of other locations are in Appendix 2.3.

We find that for all the locations we pick up, for $\{D_j^{m_d}(d)\}$, $\{D_j^{v_d}(d)\}$, B-splines with one internal breakpoint, for $\{D_i^{m_y}(y)\}$, $\{D_i^{v_y}(y)\}$, $\{D_i^{m_y}(y)D_j^{m_d}(d)\}$ and $\{D_i^{v_y}(y)D_j^{v_d}(d)\}$, B-splines with no internal breakpoint, for $\{P_i^m(d)\}$ and $\{P_i^v(d)\}$, 6 harmonic functions, for the basis of a_k^m, b_k^m, a_k^v and b_k^v, periodic B-splines with no internal breakpoints are enough to make the data roughly stationary (in terms of marginal distribution) and increase model complexity cannot significantly improve the model performance and even overfit the data. Thus, the final model formula for the mean function is:

$$\mathrm{m}(y,d)= \alpha_m+ \sum_{i=1}^{3} \beta_i^m(y)D_i^{m_y}(y) + \sum_{j=1}^{4} \gamma_j^m(d)D_j^{m_d}(d) + \sum_{i,j=1}^{3} \rho_{ij}^m(d)D_i^{m_y}(y)D_j^{m_d}(d) + \sum_{k=1}^{3} [a_k^m(d)P_{2k-1}^m(d)+b_k^m(d)P_{2k}^m(d)] \tag{2.3.1}$$

The logarithm of variance function $logv(t,d)$ has a similar form:

$$\mathrm{logv}(y,d)= \alpha_v+ \sum_{i=1}^{3} \beta_i^v(y)D_i^{v_y}(y) + \sum_{j=1}^{4} \gamma_j^v(d)D_j^{v_d}(d) + \sum_{i,j=1}^{3} \rho_{ij}^v(d)D_i^{v_y}(y)D_j^{v_d}(d) + \sum_{k=1}^{3} [a_k^v(d)P_{2k-1}^v(d)+b_k^v(d)P_{2k}^v(d)] \tag{2.3.2}$$

where a_k^m, b_k^m, a_k^v and b_k^v are all linear combinations of periodic B-splines of d with an intercept and no internal breakpoints.

2.3 Data for cross-validation

For model estimate and evaluation, we use 5-fold cross-validation. To process the data, for the data of each location:

(1) Treat the 1st to 40th simulations as the training set and 41th to 50th simulations to be the testing set, build models (2.3.1) and (2.3.2) on the training set and normalize the data of both training set and testing set.

(2) Treat the 31th to 40th simulations to be the testing set and other simulations to be the training set, repeat step (1).

(3) Treat the 21th to 30th simulations to be the testing set and other simulations to be the training set, ⋯

…

We then get 5 different normalized temperatures used for the cross-validation in the following parts, and mark the whole sets as S_i, the testing sets $test_i$ and the training sets $train_i$, $i=1, 2, 3, 4, 5$.

3 Deciding the past information to be used

We suppose T_t in (2.2.1) satisfies the following process:

$$T_t = f(T_{t-1}, T_{t-2}, \cdots, T_{t-n}, \cdots) + \varepsilon_t \tag{3.1}$$

where f is an infinite-dimensional vector function of all the past temperatures, and ε_t are independent normal variables with mean 0 and variance σ^2. So, the conditional mean function of T_t, has the following formula:

$$E(T_t | T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-n}, \cdots) = f(T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-n}, \cdots) \tag{3.2}$$

where T^*_{t-i} represents fixed past temperatures. We aim to estimate the function f with reasonable complexity. So, we need to decide the lag order of $E(T_t|\cdot)$. Because we focus on the dynamic of the temperatures in summer spanning from June 1 to August 31(for which $152 \le d \le 243$), in the following part of the chapter and Chapter 4, we restrict the time index t on $\{t=(y, d) | 152 \le d \le 243\}$.

3.1 Deciding the lag order

Based on our i. i. d. assumption of ε_t, we use ANN (Artificial Neuron Network) to estimate the function E_i with different lag order i and select parameter i based on *MSE*.

To be specific:

(1) Take T^*_{t-1} as the predictor and use ANN to estimate the model $T_t = f_1(T^*_{t-1}) + \varepsilon_t$, and extract the mean squared error $MSE^{ANN}(T^*_{t-1})$.

(2) Add T^*_{t-2} as the second predictor and use ANN to estimate the model $T_t = f_2(T^*_{t-1}, T^*_{t-2}) + \varepsilon_t$ and get the mean squared error $MSE^{ANN}(T^*_{t-1}, T^*_{t-2})$.

(3) Add T^*_{t-3} and so on so for, stop the procedure based on the decreasing of the *MSE*.

One problem need to be considered is that because ε_t has mean 0, if $\widehat{E}(T_t|\cdot)$ is precise enough, the mean of the residuals should be independent of $\widehat{E}(T_t|\cdot)$. When use ANN for regression, we only aim at minimizing the *MSE*. Because $MSE = Var + Bias^2$, it is not guaranteed that we get an unbiased estimate, so we need to check the behavior of $\widehat{E}(T_t|\cdot)$.

From Figure 3.1.1, we can see that on plots means against predicted values, they are randomly distributed around 0 and they show no apparent patterns, which means there is no serious bias. More details about other locations and how we build ANN model are in Appendix 3.1, which shows that in most cases, the model has no serious bias problem. For all the locations we pick up, over 90% data distributed within [-2, 2], if we consider those data with predicted value below 0.05 and above 0.95 quantile as "extreme values", for the locations near the Great Lakes, the mean of residuals is controlled better (by which I mean it is close to 0) than other locations. From table 3.1.1, we can see near the California Coast and the Gulf of Mexico, $MSE_{|e} > MSE_{|u}$, while near the Great Lakes, $MSE_{|e} < MSE_{|u}$. The reason could be that near the California Coast and the Gulf of Mexico the extreme values are not so "extreme" influencing the total loss so that the estimate function at the extreme values is not precise enough, while near the Great Lakes extreme values are extreme enough to influence the total loss.

Location \ MSE		$MSE_{\|e}$	$MSE_{\|u}$
the California Coast	B_1	0.20	0.27
	B_2	0.15	0.18

	B_3	0.14	0.19
the Great Lakes	L_1	0.44	0.32
	L_2	0.40	0.31
	L_3	0.41	0.33
the Gulf of Mexico	G_1	0.08	0.11
	G_2	0.25	0.35
	G_3	0.23	0.36

Table 3.1.1

This table compare $MSE_{|u}$ and $MSE_{|e}$.We denote the 0.05 quantile and 0.95 quantile of the predictions by a and b. Then we denote MSE on the subset with predicted values less than a and larger than b as $MSE_{|e}$ and on the subset with predicted values between a and b as $MSE_{|u}$.

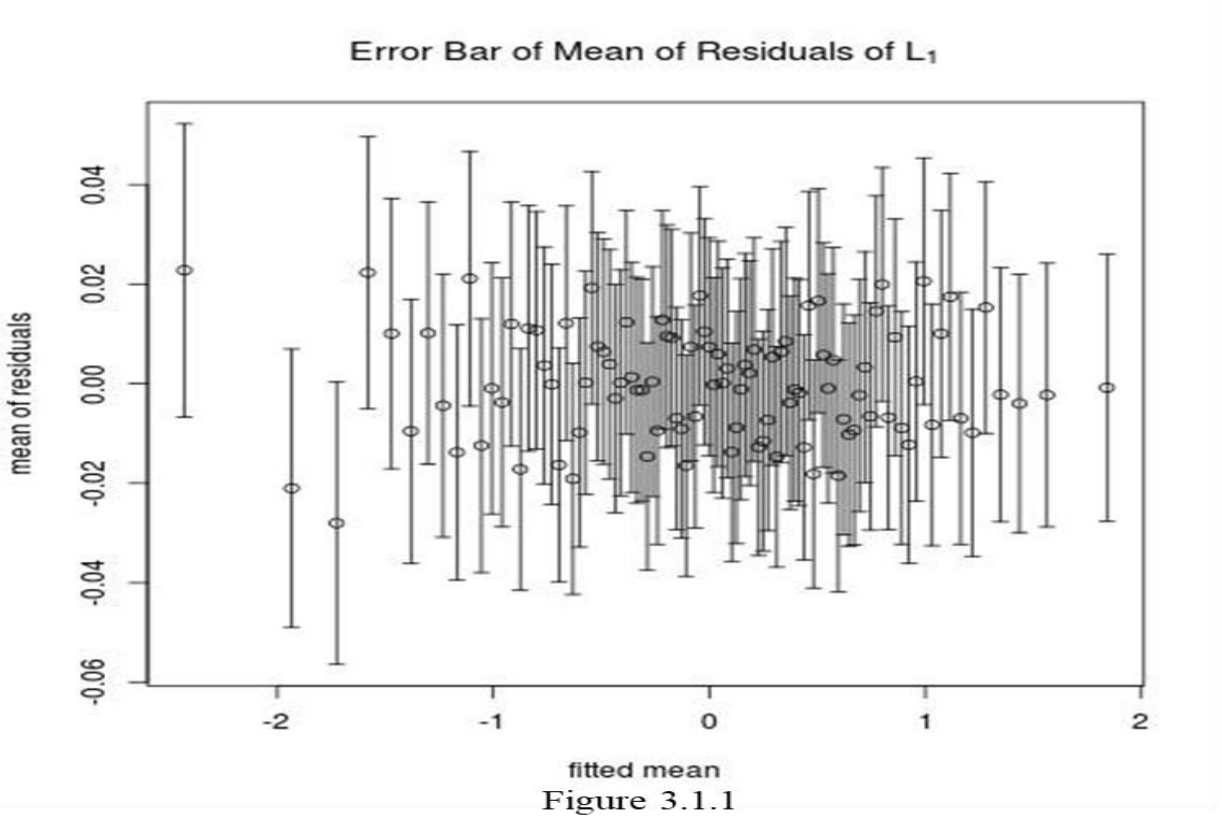


Figure 3.1.1

The error bar of the mean of residuals versus the estimated mean of L_1. We take all the percentiles as break points and divide the residuals into 100 bins, we calculate the sample mean and variance within each bin to make the error bar.

Table 3.1.2 show how the $MSE^{ANN}(\cdot)$ changes as the lag order increases. In general, for locations in the same area, *MSE* are close, while for different locations the *MSE* can be very different. Places near the California Coast and the Gulf of Mexico have smaller *MSE* while those near the Great lakes have larger ones.

Location \ Lag order		1	2	3	4
the California Coast	B_1	0.283349	0.221680	0.208040	0.207586
	B_2	0.228325	0.163500	0.152118	0.151449
	B_3	0.244155	0.163967	0.149077	0.148525
the Great Lakes	L_1	0.398441	0.345474	0.334851	0.333291
	L_2	0.385976	0.332253	0.318518	0.316962
	L_3	0.385976	0.332253	0.318518	0.316962
the Gulf of Mexico	G_1	0.141886	0.097207	0.086690	0.086353
	G_2	0.286078	0.262122	0.259775	0.259547
	G_3	0.289000	0.258410	0.248865	0.248578

Table 3.1.2
This table compares the $MSE^{ANN}(\cdot)$ of model with different lag order from 1 to 4.

One point needs to be considered is that the *MSE* of G_1 is much smaller than those of other locations. G_1 is on the northwest coast of the Gulf of Mexico, while G_2 on the north and G_3 on the northeast, and G_1 is a bit far from others. There may be few factors affecting local summer temperatures, so the *MSE* is small.

Table 3.1.2 shows in most cases, until we increase i from 2 to 3, MSE can reduce by more than 3%, it is a significant improvement. But if we increase i from 3 to 4, *MSE* only decreases a little bit, all less than 1%, so we finally choose i=3. So, our estimated conditional mean function is:

$$\widehat{E}(T_t|T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-n}, \cdots)=\hat{f}_3\,(T^*_{t-1}, T^*_{t-2}, T^*_{t-3}) \qquad (3.1.1)$$

Location \ Lag order		C_1	C_2	C_3
the California Coast	B_1	21.76	6.15	0.22
	B_2	28.39	6.96	0.44
	B_3	32.84	9.08	0.37

the Great Lakes	L_1	13.29	3.07	0.47
	L_2	13.92	4.13	0.49
	L_3	13.92	4.13	0.49
the Gulf of Mexico	G_1	31.49	10.82	0.39
	G_2	8.37	0.90	0.09

Table 3.1.3

Here, we define C_i as the percentile reduction of *MSE* when we increase the lag order from i to $i+1$, i.e.,

$$C_i=\frac{MSE^{ANN}(T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-i-1})-MSE^{ANN}(T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-i})}{MSE^{ANN}(T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-i})}\times 100\%.$$

The table also shows an interesting pattern: in general, if C_1 is larger, C_2 is also larger, i.e., if T_{t-2} can, comparatively, explain more of $\widehat{E}(T_t|\cdot)$, then T_{t-3} can explain more of $\widehat{E}(T_t|\cdot)$. More quantitatively, C_1 is about 3 to 4 times of C_2.

Another point is that C_2 of G_2 is very small, and C_1 of G_2 is also smaller than others, while *MSE* of G_2 is not strange, which means, T_{t-1} could explain most variance of local temperatures. On the contrary, while *MSE* of G_1 is quite small, but its C_1 and C_2 are large. Besides, the *MSE* of B_3 is not special, but its C_1 and C_2 are both quite large. So, there is no clear relationship between *MSE* and C_1 (and C_2).

It is worth to notice that because of the 50 ensembles available, we can apply a black box nonparametric method, which is impossible with only one ensemble.

3.2 Decomposition of the conditional mean function

We can often get a good fitting predictor through Black box methods like ANN, but it s difficult to attain an intuitive impression of how the model looks like. We could visualize the estimated function to see what it looks like. Basically, we want to project the function onto the space expanded by the predictors. In this part we will introduce our method of visualizing the function $\hat{f}_3(T^*_{t-1}, T^*_{t-2}, T^*_{t-3})$.

We suppose $\hat{f}_3$ is additive with the formula:

$$\hat{f}_3(x_1, x_2, x_3)=h_1(x_1)+h_2(x_2)+h_3(x_3)+h_{12}(x_1, x_2)+h_{23}(x_2, x_3)+h_{123}(x_1, x_2, x_3) \quad (3.2.1)$$

Then we decompose $\hat{f}_3$ the following way:

(1) Pick up a set of 60 points $P=\{p_i\}_{i=1}^{60}$ equally spaces on the interval [-6, 5], which contains most T_t in the samples. Construct set $P^3=\{(x_1, x_2, x_3)|x_1 \epsilon P, i=1, 2, 3\}$. Then we calculate the values of $\hat{f}_3$ on P^3 and get the set $\tilde{T}=\{\hat{f}_3(\vec{x})|\vec{x}\,\epsilon P^3\}$.

(2) Consider $\tilde{T}$ as the response and the corresponding x_1 as the predictor and use ANN to estimate the model $h_1^1(x_1)$. Extract the residuals r_1 and regress r_1 on x_2, get $\hat{h}_2^1(x_2)$, extract r_2 and regress it on x_3, extract r_3, regress it on x_1 and x_2, and so on. Finally, we get r_{13}. Then we need to treat r_{13} as the "new $\tilde{T}$", $\tilde{T}$", $\tilde{T}'$.

(3) Repeat the procedures (1) and (2) until the residuals converges. Then, $\hat{h}_i$should be the sum of all $\hat{h}_i^j$.

Luckily, for all the points we picked up, it seems that using the procedure above one time is enough.

Appendix 3.2 (1) shows how the *MSE* change in the second loop, the *MSE* only changes slightly and there is no guarantee that it decreases, so we can just consider that the residuals "converge" after the first loop.

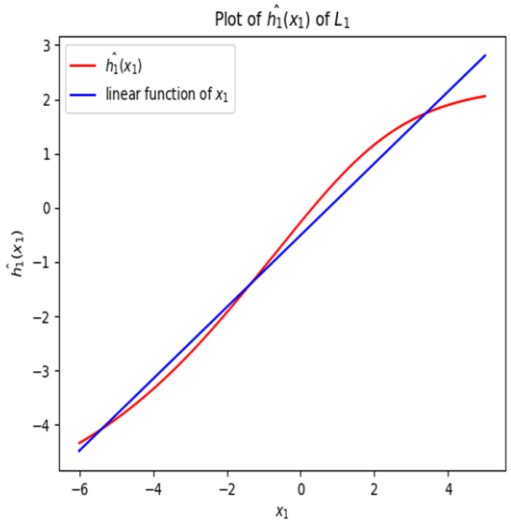


Figure 3.2.1 plot of $\hat{h}_1(x_1)$ and linear function of x_1 of B_1

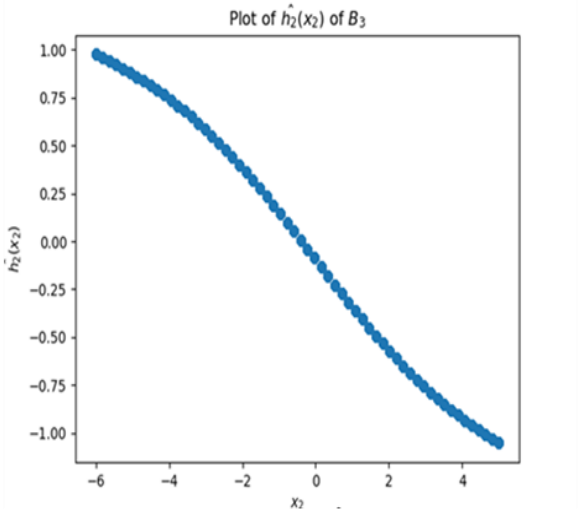


Figure 3.2.2figure of $\hat{h}_2(x_2)$ of B_3

Figures of interactions of B_1

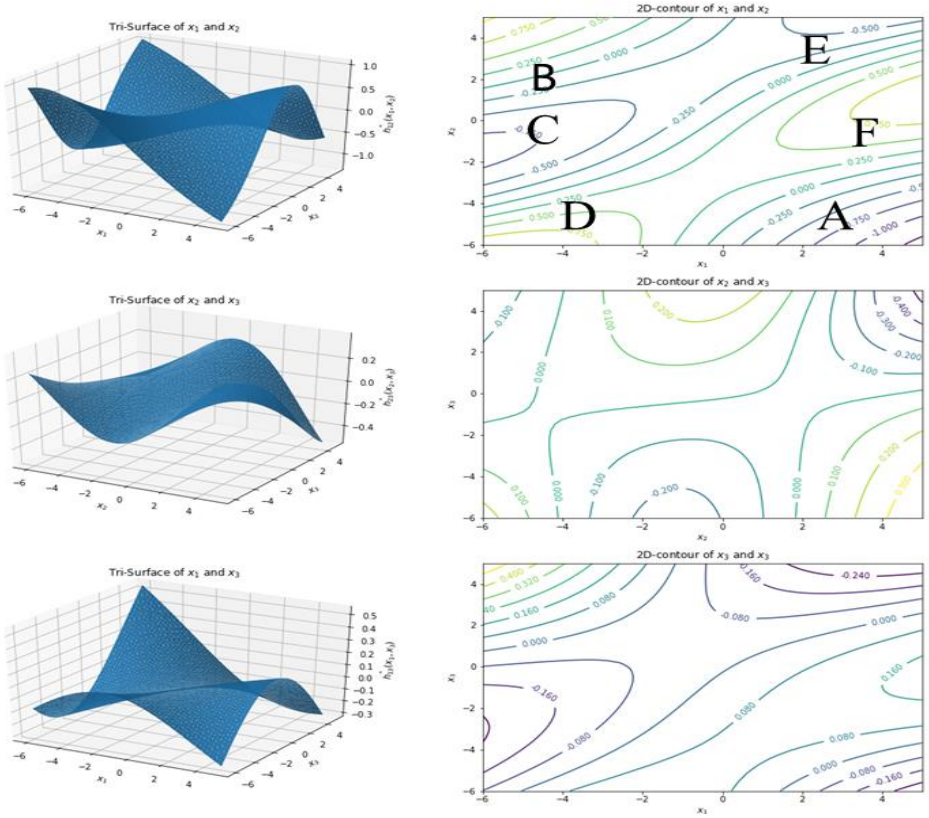


Figure 3.2.3 interactions of B_1

Under the assumption of additive model, the *MSE* represents the interaction of x_1, x_2 and x_3. We can see the *MSE* are not large, which means the interaction is small. But if we neglect this assumption, the *MSE* represents the combination of the interaction of x_1, x_2, x_3 and the "distance" between the addictive model and the function $\hat{f}_3(x_1, x_2, x_3)$, so, we can see the functions are not far from additive functions.

We provide some figures for some estimated functions, from which we can get the intuition of what is the function $\hat{f}_3(x_1, x_2, x_3)$ looks like. Figure 3.2.1 shows $\hat{h}_1(x_1)$ and linear function of x_1 of B_1, we can see that $\hat{h}_1(x_1)$ is an increasing function on [-6, 5] and very close to a linear one. Compared with the linear function, the absolute value of slope of the curve is larger when $|x_1|$ is small, and smaller when $|x_1|$ is large. The $\hat{h}_1(x_1)$ for all the locations we pick up show the similar pattern. Figure 3.2.2 shows the figure of $\hat{h}_2(x_2)$ of B_3. Figures for other locations are in Appendix 3.2 (2). In most cases, $\hat{h}_2(x_2)$ is a decreasing function like Figure 3.2.2, but from Appendix 3.2 (2) we can see $\hat{h}_2(x_2)$ is quite complicated on the interval [-6, 5]. But for all the locations we pick up, over 90% data distributed within [-2, 2], and for most locations, $\hat{h}_2(x_2)|_{[-2,\,2]}$ is decreasing. So, $\hat{h}_2(x_2)$ is only complicated on those extreme value points, the reason for which may be that the extreme values and "usual values" obey different criterions, and $h_2(x_2)|_{x_2\ extreme}$ be more complicated. However, because there are too few extreme values to estimate the model, we cannot expect that $\hat{h}_2(x_2)|_{x_2\ \text{extreme}}$ is close enough to $h_2(x_2)|_{x_2\ \text{extreme}}$, but for the "usual" values, considering the large dataset, we consider that $\hat{h}_2(x_2)|_{x_2\ \text{usual}}$ close enough to $h_2(x_2)|_{x_2\ \text{usual}}$. Cases for $\hat{h}_3(x_3)$ are even more complicated. Even though we constrain $\hat{h}_3(x_3)$ on the "usual" values, there is no guarantee of the monotonicity even within the same area. Figures can be found in Appendix 3.2 (3). Figure 3.2.3 shows the Tri-surfaces and 2D contours of all two-way interactions of B_1. Comparatively, the interaction of x_1 and x_2 is more regular. As we marked in the 2D contour of the interaction of x_1 and x_2 of B_1 in Figure 3.2.4, we can divide the figure in to A to F 6 Areas, and the sign of the values of $\hat{h}_{12}(x_1, x_2)$ in the 6 parts are all the same and they change the same way. In Area A and B, $\hat{h}_{12}(x_1, x_2)$ decreases as x_1 increases and x_2 decreases, while $\hat{h}_{12}(x_1, x_2) < 0$ in B and $\hat{h}_{12}(x_1, x_2) > 0$ in A. In Area E and D, $\hat{h}_{12}(x_1, x_2)$ decreases as x_1 increases, while $\hat{h}_{12}(x_1, x_2) < 0$ in E and $\hat{h}_{12}(x_1, x_2) > 0$ in D. In Area C and F, $\hat{h}_{12}(x_1, x_2)$ increases as x_1 increases, while $\hat{h}_{12}(x_1, x_2) < 0$ in C and $\hat{h}_{12}(x_1, x_2) > 0$ in F. Things are more complex for other interaction terms, of which the surface can be very complicated. More details of relative figures can be seen in Appendix 3.2 (4).

We infer that, in general, functions like $\hat{h}_1(x_1)$, $\hat{h}_{12}(x_1, x_2)$ shows more global properties, while functions like $\hat{h}_2(x_2)$ capture more local structures. Those functions together describe how the mean temperatures changes given the past information.

3.3 Model evaluation by cross-validation

We use 5-fold cross-validation to evaluate the model performance. For each normalized temperature set S_i, we estimate an ANN on $train_i$, calculate the *MSE* on $test_i$ to get MSE_{i1}, and calculate the *MSE* on $train_i$ to get MSE_{i2} then we calculate the average of MSE_{i1}, $i=1, 2, 3, 4, 5$ to get MSE_{t1} and MSE_{i2} to get MSE_{t2}. In table 3.3.1, we show MSE_{i1} and MSE_{t1} for all the locations. The model performance is similar with that on the whole data. We also provide MSE_{i2} and MSE_{i2} in Appendix 3.3. The *MSE* on training sets and testing sets are very close to each other but there's no guarantee which is smaller. This is in a bit of conflict with our expectation that the *MSE* on the training set should be smaller, but perhaps is not so surprising given the large training sets. However, MSE_{t2} is always a bit smaller than MSE_{t1}.

In the following passage, we will note MSE_{t1} of the ANN model as MSE_1^{ANN} We will compare the performance of different models based on cross-validation in following parts.

MSE / Location	1	2	3	4	5	t

the California Coast	B_1	0.207158	0.210714	0.205653	0.208005	0.209151	0.208136
	B_2	0.150394	0.154177	0.150888	0.152270	0.153564	0.152259
	B_3	0.149692	0.149594	0.149922	0.148115	0.148459	0.149156
the Great Lakes	L_1	0.333375	0.332919	0.337700	0.333540	0.339094	0.335326
	L_2	0.321724	0.317261	0.318710	0.320456	0.316174	0.318865
	L_3	0.342445	0.337410	0.337237	0.339807	0.336421	0.338664
the Gulf of Mexico	G_1	0.087318	0.085693	0.089534	0.086561	0.084907	0.086803
	G_2	0.261097	0.260972	0.263922	0.257481	0.256694	0.260033
	G_3	0.252508	0.245386	0.247433	0.252198	0.247665	0.249038

Table 3.3.1

This table shows MSE_{*1}. We estimate an ANN on $train_i$, calculate the *MSE* on the corresponding $test_i$ to get MSE_{i1}, $i=1, 2, 3, 4, 5$, and calculate the average of MSE_{i1} to get MSE_{t1}.

4 Reducing to AR models

From Part 3, we see that we can use function $\hat{f}_3(x_1, x_2, x_3)$ to approximate the conditional mean function $E(T_t|T^*_{t-1}, T^*_{t-2}, \cdots, T^*_{t-n}, \cdots)$. $\hat{f}_3(x_1, x_2, x_3)$is non-linear and close to additive, but the non-linearities vary from place to place and are complicated, so it is difficult to build a uniform model to explain the temperatures for different locations. The *AR* model is a simple but widely used linear time series model, and we want to explore how much we would lose if we just use linear *AR* model to estimate f_3. The proof of theorems in this chapter can be found in ref. [3].

4.1 Brief review of AR model

Definition 4.1.1: A time series $\{y_t; t=0, \pm1, \pm2, \cdots\}$ is $ARMA(p, q)$ if it is stationary and

$$y_t=\phi_1 y_{t-1}+\phi_2 y_{t-2}+ \cdots +\phi_p y_{t-p}+\varepsilon_t+\theta_1\varepsilon_{t-1}+ \theta_2\varepsilon_{t-2}+ \cdots + \theta_q\varepsilon_{t-q} \quad , (4.1.1)$$

where $\{\varepsilon_t\}$ is a white noise sequence with variance $\sigma^2>0$, and $\phi_p\neq0$, $\theta_q\neq0$. when $q = 0$, the model is called an autoregressive model of order p, $AR(p)$.

Definition 4.1.2: The AR polynomials are defined as

$$\phi(z)=1-\phi_1 z-\phi_2 z^2-\cdots-\phi_p z^p, (\phi_p\neq0), \quad (4.1.2)$$

where z is a complex number.

Definition 4.1.3: An $ARMA(p, q)$ model is said to be causal, if the time series $\{y_t; t=0, \pm1, \pm2, \cdots\}$ can be written as a one-sided linear process:

$$y_t=\sum_{j=0}^{+\infty}\psi_j\varepsilon_{t-j}=\psi(B)\,\varepsilon_t, \quad (4.1.3)$$

where $\psi(B)=\sum_{j=0}^{+\infty}\psi_j B^j$, and $\sum_{j=0}^{+\infty}|\psi_j|<\infty$; $\psi_0=1$.

Theorem 4.1.1: An AR model is causal only when all roots of $\phi(z)$ are out of the unit disk $\{z;|z|\leq 1\}$.

Basically, the concept causality means the model is future-independent.

4.2 Model estimate and forecasting

We use Ordinary Least Squares method to estimate the model.

We suppose T_t satisfies:

$$T_t=\phi_1 T^*_{t-1}+\phi_2 T^*_{t-2}+\phi_3 T^*_{t-3}+\varepsilon_t, \tag{4.2.1}$$

where ε_t are independent, identically distributed random variables with mean 0 and variance $\sigma^2>0$.

On each training set $triain_i$, we use ordinary least square method to estimate the model Mod_i, and get the *MSE* on the training set MSE_{i2}, and calculate the average of MSE_{i2} to get MSE_{t2}.

Appendix 4.2 (1) shows the estimated coefficients of the *AR* models. One problem need to be considered is that from Appendix 3.2 (2), $\hat{h}_2(x_2)|_{[-2,\,2]}$ is not monotone for both G_1 and B_2, while for both locations $\hat{\phi}_2$ are negative.

Appendix 4.2 (2) shows the corresponding solutions of the *AR* polynomials. The roots of the polynomials are all out of the unit disk, i.e., all the estimated models are causal.

If the time series is a causal $AR(p)$ process, then, for $t\geq p$, the best linear predictor for one-step ahead prediction is

$$\tilde{y}_{t+1}=\phi_1 y_t+\phi_2 y_{t-1}+\cdots+\phi_p y_{t-p+1}. \tag{4.2.2}$$

Based on (4.2.2), we use Mod_i to make one-step ahead prediction on $test_i$, and get MSE_{i1} on each testing set. We then calculate the mean of MSE_{i1} to get MSE_{t1}.

Table 4.2.1 shows MSE_{i1} and MSE_{t1} of the of the $FAR(3)$ model. MSE_{i1} of the same location are quite close to each other. In Appendix 4.2 (3), we provide MSE_{i2} and MSE_{t2} It show a similar phenomenon like that of the ANN model. There is no guarantee the MSE_{i1} is larger than MSE_{i2}, but MSE_{t1} is always larger than MSE_{t2}, which is quite interesting. We denote MSE_{t1} by MSE_1^{OLS} for the following model comparison.

MSE_{*1} / Location		1	2	3	4	5	*t*
the California Coast	B_1	0.208265	0.212102	0.206806	0.209543	0.209884	0.209320
	B_2	0.151332	0.155111	0.151991	0.153166	0.154565	0.153233
	B_3	0.150830	0.150458	0.150961	0.149010	0.149358	0.150123
the Great Lakes	L_1	0.340158	0.340399	0.345339	0.340600	0.346112	0.342522
	L_2	0.327851	0.323740	0.326463	0.327541	0.324036	0.325926
	L_3	0.351522	0.345923	0.346856	0.349007	0.346033	0.347868

the Gulf of Mexico	G_1	0.088600	0.086838	0.091190	0.087971	0.085869	0.088094
	G_2	0.262873	0.262783	0.265980	0.259386	0.258015	0.261807
	G_3	0.253615	0.246740	0.248613	0.253412	0.248629	0.250202

Table 4.2.1 The table shows MSE_{i1} and MSE_{t1} of the of the $FAR(3)$ model.

5 Introducing long-range dependence

Hurst observed in 1950 that Nile streamflows exhibited persistent excursions from their mean value. Since Hurst phenomenon was introduced, long-range dependence characteristics of the climatological time series have drawn attention of scientists. We want to explore whether this long-range dependence exists in our data and whether we could use it to improve our model performance. So, we introduce *FAR* model in this part. Proofs of all theorems introduced in this part can be found in ref. [4].

5.1 Introduction to long-range dependence

Long-range dependence may be defined in many ways. However, as pointed out by Hall (1997), the original motivation for the concept of long memory is closely related to the estimation of the mean of a stationary process. If the autocovariances of a stationary process are absolutely summable, then generally speaking, these processes are said to have short memory. On the contrary, a process has long memory if its autocovariances are not absolutely summable. And we introduce some definitions of long memory below.

Definition 5.1.1 A: Let $\gamma(h)=\langle y_t, y_{t+h}\rangle$ be the autocovariance function at lag h of the stationary process $\{y_t: t\in Z\}$. if

$$\sum_{h=-\infty}^{+\infty}|\gamma(h)|=\infty, \tag{5.1.1}$$

we call $\{y_t: t\in Z\}$ a long-memory process.

Definition 5.1.1 B: Let $\gamma(h)=\langle y_t, y_{t+h}\rangle$ be the autocovariance function at lag h of the stationary process $\{y_t: t\in Z\}$. if

$$\gamma(h)\sim h^{2d-1}l(h) \tag{5.1.2}$$

as $h\to\infty$, we call $\{y_t: t\in Z\}$ a long-memory process. d is the long-memory parameter and $l(\cdot)$ is a slowly varying function in Karamata's sense.

According to the Wold representation theorem, a stationary purely nondeterministic process may be expressed as:

$$y_t=\sum_{j=0}^{+\infty}\psi_j\varepsilon_{t-j}=\psi(B)\varepsilon_t \tag{5.1.3}$$

Where $\psi_0=1$, $\sum_{j=0}^{+\infty}\psi_j^2<+\infty$, $\{\varepsilon_t\}$ is a white noise sequence with variance σ^2. An alternative definition of long-memory behavior could be based directly on (5.1.3):

Definition 5.1.1 C: For process y_t, if ψ_j in (5.1.3) satisfy

$$\psi_j\sim j^{d-1}l(j) \tag{5.1.4}$$

for $j>0$, and $l(\cdot)$ is a slowly varying function, we call $\{y_t: t\in Z\}$ a long-memory process.

Unless further conditions are imposed, these definitions are not necessarily equivalent. Some relationships among these definitions are established in the following theorem:

Theorem 5.1.1: Let $\{y_t\}$ be a stationary process with Wold expansion (5.1.3). Assuming that $0 < d < \frac{1}{2}$, we have:

(a) If the process $\{y_t\}$ satisfies (5.1.4), it also satisfies (5.1.2).

(b) If the process $\{y_t\}$ satisfies (5.1.2), it also satisfies (5.1.1).

5.2 Introduction to FAR processes

A well-known class of long-memory models is the autoregressive fractionally integrated moving-average (*ARFIMA*) process introduced by Granger and Joyeux (1980) and Hosking (1981). An *ARFIMA*(p, d, q) process $\{y_t\}$ can be defined by

$$\phi(B)y_t = \theta(B)(1-B)^{-d}\varepsilon_t, \tag{5.2.1}$$

where $\phi(B)=1+\phi_1 B+ \cdots+\phi_p B^p$ and $\theta(B)= 1+\theta_1 B+ \cdots+\theta_q B^q$ are the autoregressive and moving-average operators, respectively, $\phi(B)$ and $\theta(B)$ have no common roots, $(1-B)^{-d}$ is a fractional differencing operator defined by the binomial expansion

$$(1-B)^{-d} = \sum_{j=0}^{+\infty}\eta_j B^j, \tag{5.2.2}$$

where

$$\eta_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)}, \tag{5.2.3}$$

for $d< \frac{1}{2}$, $d \neq 0, -1, -2, \cdots$ and $\{\varepsilon_t\}$ is a white noise sequence with a finite variance.

Specifically, let $q=0$, we get the *FAR* process.

Theorem 5.2.1 examines the existence of a stationary solution of the *FAR* process defined by equation (5.2.1), including its uniqueness, causality.

Theorem 5.2.1: Consider the *FAR* process defined by (5.2.1) with $q=0$. Assume $d \in (-1, \frac{1}{2})$. Then,

(a) If the roots of $\phi(\cdot)$ lie outside the unit circle $\{z: |z|= 1\}$, then there is a unique stationary solution of (5.2.1) given by

$$y_t = \sum_{j=-\infty}^{+\infty}\psi_j\varepsilon_{t-j}, \tag{5.2.4}$$

Where $\psi(z)=(1-z)^{-d}/\phi(z)$.

(b) If the zeros of $\phi(\cdot)$ lie outside the closed unit disk $\{z: |z|= 1\}$, then the solution y_t is causal.

According to Theorem 5.2.1, under the assumption that the roots of the polynomials $\phi(B)$ are outside the closed unit disk $\{z: |z|= 1\}$ and $d \in (-1, \frac{1}{2})$, the *FAR* process is stationary, causal. In this case we can write

$$y_t = (1-B)^{-d}\phi(B)^{-1}\theta(B)\varepsilon_t = \psi(B)\varepsilon_t. \tag{5.2.5}$$

The $MA(\infty)$ coefficients, ψ_j satisfy the following asymptotic relationships:

$$\psi_j = \frac{\theta(1)j^{d-1}}{\phi(1)\Gamma(d)} + O(j^{-1}), \tag{5.2.6}$$

As $j \to \infty$. So, an *FAR* process satisfies the Definition 5.1.1 C, and is a long memory process.

For *FAR* process, observe that the polynomial $\phi(B)$ in (5.2.1) can be written as

$$\phi(B) = \prod_{i=1}^{p} (1 - \rho_i B). \tag{5.2.7}$$

Assuming that all the roots of $\phi(B)$ have multiplicity one, the autocovariance function have the following form:

$$\gamma(h) = \sigma^2 \sum_{j=1}^{p} \xi_j C(d, p+i-h, \rho_j), \tag{5.2.8}$$

with

$$\xi_j = [\rho_j \prod_{i=1}^{p} (1-\rho_i \rho_j) \prod_{m \neq j} \prod_{i=1}^{p} (\rho_j - \rho_m)]^{-1}, \tag{5.2.9}$$

and

$$C(d, p+i-h, \rho_j) = \frac{\gamma_0(h)}{\sigma^2} [\rho^{2p} \beta(h) + \beta(-h) - 1], \tag{5.2.10}$$

where $\beta(h) = F(d+h, 1, 1-d+h, \rho)$ and $F(a, b, c, x)$ is the Gaussian hypergeometric function

$$F(a, b, c, x) = 1 + \frac{a \cdot b}{\gamma \cdot 1} x + \frac{a \cdot (a+1) \cdot b \cdot (b+1)}{\gamma(\gamma+1) \cdot 1 \cdot 2} x^2 + \cdots, \tag{5.2.11}$$

It can be shown that

$$\gamma(h) \sim c_\gamma |h|^{2d-1}, \tag{5.2.12}$$

As $|h| \to \infty$, where

$$c_\gamma = \frac{\sigma^2}{\pi |\phi(1)|^2} \Gamma(1-2d) \sin(\pi d), \tag{5.2.13}$$

which satisfies Definition 5.1.1 B.

5.3 Model estimate

We use the maximum likelihood method to estimate models.

Specifically, suppose we have l independent simulations of temperatures of a n-year-long period, i.e., for the process $T((y, d))$, $\frac{1}{365} \le y \le n$ we have realizations $T_i = \{T_i((y, d)) | \frac{1}{365} \le y \le n\}$, $i = 1, 2, \cdots, l$. We estimate the model by the following procedures:.

(a) To make the estimate more computationally efficient, set up a threshold s (the length of years we build the conditioning) for the past observations used to estimate the model.

(b) For each series T_i, derive 2n sub-series:

$$T_i^j = \begin{cases} \{T_i((y, d)), | \frac{1}{365} \le y \le j + \frac{243}{365}\}, & \text{if } j \le s \\ \{T_i((y, d)), | (j-s+1) \times 365 \le y \le j + \frac{243}{365}\}, & \text{if } j > s \end{cases}, \quad j = 0, 1, 2, \cdots, n-1 \tag{5.3.1}$$

and

$$\tilde{T}_i^j = \begin{cases} T_i = \{T_i((y, d)), | (1 \le y \le j + \frac{152}{365}\}, & \text{if } j > s \\ T_i = \{T_i((y, d)), | (j-s)+1 \le y \le j + \frac{152}{365}\}, & \text{if } j > s \end{cases}, \quad j = 0, 1, 2, \cdots, n-1 \tag{5.3.2}$$

Figure 5.3.1 shows how we construct the sub-series.

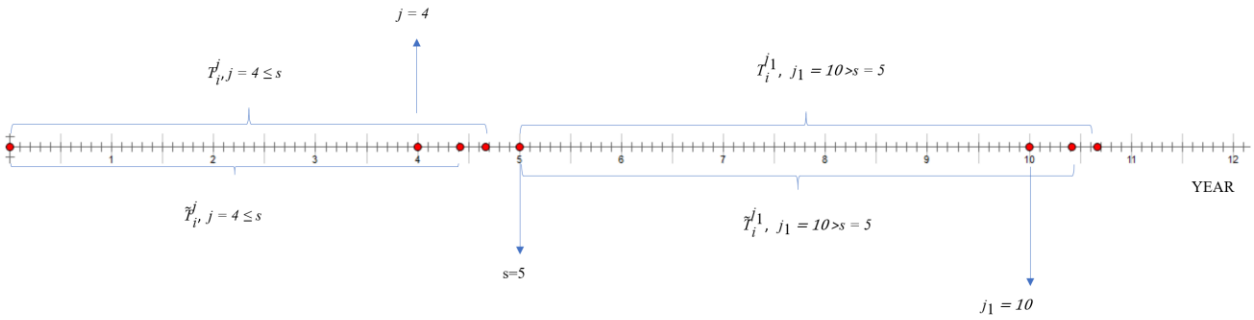


Figure 5.3.1

We let $s = 5, j = 4 < s$ and $j_1 = 10 > s$ to show how to construct the sub-series.

(c) Calculate the log-likelihood for the $AR(3)$ model of each series, get $llk_i^j(s)$ and $\widetilde{llk}_i^j(s)$, $i=1, 2, \cdots, l,\ j= 0, 1, 2, \cdots, n-1$. The conditional log-likelihood of the summer days of the i^{th} run j^{th} year is $llk_i^j(s)$-$\widetilde{llk}_i^j(s)$. So, conditional log-likelihood of the summer days of the whole data is $llk(s)= \sum_{i=1}^{l}\sum_{j=1}^{n-1} llk_i^j(s)$-$\widetilde{llk}_i^j(s)$. We set the initial values for the AR coefficients to be $\hat{\phi}_1^{ols}$, $\hat{\phi}_2^{ols}$, and $\hat{\phi}_3^{ols}$ (the estimates of the coefficients of the $AR(3)$ model by the least square method from Chapter 4. We also add a mean term μ_i for the process T_i, which is expected to be 0. We minimize the $llk(s)$ to get the maximum likelihood estimate for the $AR(3)$ models.

(d) Calculate the log-likelihood for the $FAR(3)$ model of each series, get $LLK_i^j(s)$ and $\widetilde{LLK}_i^j(s)$, $i=1, 2, \cdots, l,\ j= 0, 1, 2, \cdots, n-1$. The conditional log-likelihood of the summer days of the whole data $LLK(s)= \sum_{i=1}^{l}\sum_{j=1}^{n-1} LLK_i^j(s)$-$\widetilde{LLK}_i^j(s)$. We set the initial values for the AR parameters to be $\hat{\phi}_1^{AR}$, $\hat{\phi}_2^{AR}$, and $\hat{\phi}_3^{AR}$ (the estimates of the coefficients of the $AR(3)$ model from step (c)). We also add a mean μ_i for the process T_i, which is expected to be 0 and the initial values for the fractional parameter d is 0. We minimize the $LLK(s)$ to get the maximum likelihood estimate for the $FAR(3)$ models.

In this part, we use the whole data to build the model, i.e., $l = 50$ and $j = 50$.

We estimate the models for $s \in \{1, 3, 5, 7\}$. In Appendix 5.3 (1), we compare the log-likelihood of different values of s. Also, in Figure 5.3.2, we show the change of the $LLK(s)$ as s change. There is no guarantee that the log-likelihood is a monotonically increasing function of s. Meanwhile, because they are not nested models, we cannot use methods like likelihood ratio test to decide the value of s. L_2 is an exception, the $LLK(s)$ increases as s increases all the time. In other cases, when s increase from 1 to 5, $LLK(s)$ decreases, when s increase from 5 to 7, $LLK(s)$ decreases just a bit or even increases.

We also compare the MSE of one-step ahead prediction of different values of $s$in Appendix 5.3 (2), we find that when s is increased from 3 to 5, in some cases the MSE can still decrease a little bit, but when s is increased from 5 to 7, there is almost no improvement. Besides, estimating the model based on too long past time is not computationally effective. So, for the following parts, we only consider models with s equal to 5. And in the following parts, we eliminate s from the variable names.

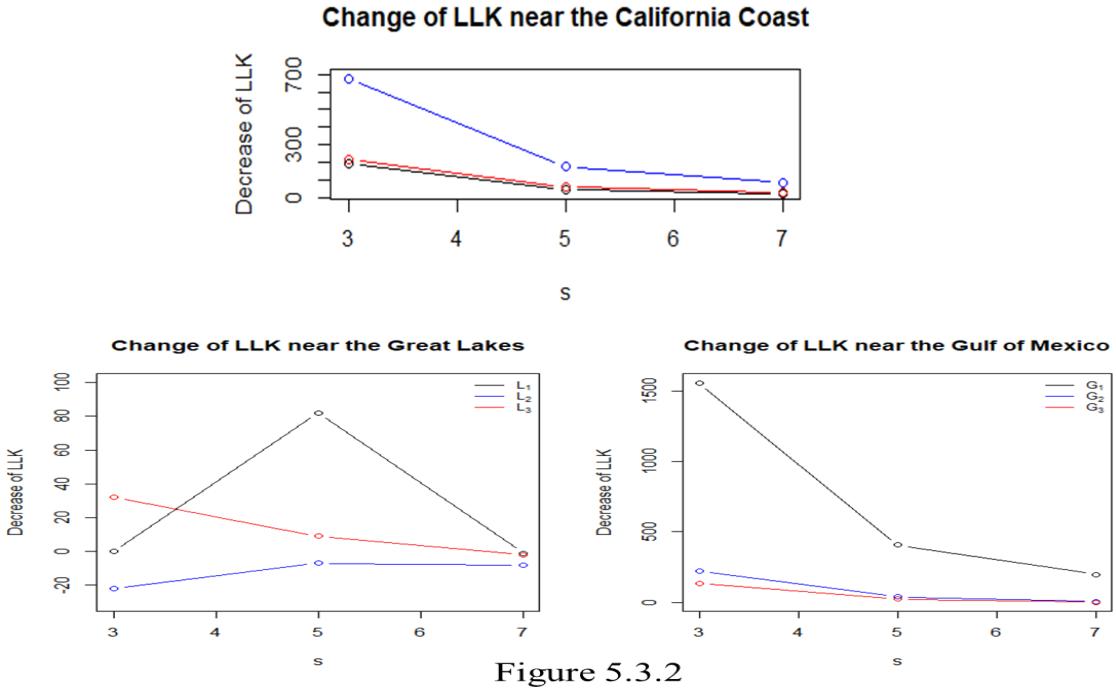


Figure 5.3.2

In the figure above, we show how *LLK(s)* as *s* changes, and for each point (s, y), it satisfies *y = LLK(s-2)-LLK(s)*.

In Table 5.3.1, we list the estimated coefficients of the *AR*(*3*) and *FAR*(*3*) models, from which we can see:

(1) $\hat{\mu}$ is close to 0, which is coincide with our expectation and indicate that we extract the trend successfully.

(2) The absolute value of $\hat{\phi}_i^{FAR}$, *i=1, 2, 3* are smaller than those of $\hat{\phi}_i^{FAR}$, *i=1, 2, 3*, but their signs are the same, $\hat{\phi}_1^{FAR}$, $\hat{\phi}_1^{AR}$, $\hat{\phi}_3^{FAR}$, and $\hat{\phi}_3^{AR}$ are positive, while $\hat{\phi}_2^{FAR}$ and $\hat{\phi}_1^{AR}$ are negative.

(3) $\hat{d}$ is always belongs to (0, 0.5), which indicate the solution is real.

(4) Comparing the values of *LLK* and the corresponding values of *llk*, if we do a likelihood ratio test, $\hat{d}$ is significant.

(5) $LLK - llk$ is an increasing function of $\hat{d}$, the larger the $\hat{d}$, the larger the $LLK - llk$.

In Appendix 5.3 (3), we show the roots of AR Polynomials of *FAR*(*3*) models, and all the roots lie outside the unit disk, so according to Theorem 4.1.1, the processes are causal. And all the roots of the AR polynomial of *FAR*(*3*) model lie outside the unit disk, according to Theorem 5.2.1, for each location, there is a unique stationary process given by $T(t) = \sum_{j=-\infty}^{+\infty} \psi_j \varepsilon_{t-j}$, where $\psi(z)=(1-z)^{-d}/\phi(z)$, and $T(t)$ is causal and invertible.

Location \ Model		*AR*(3)				*FAR*(3)					*LLK − llk*
		parameters				parameters					
		$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{\mu}$	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$	$\hat{d}$	$\hat{\mu}$	
the California Coast	B_1	1.346	-0.758	0.242	0.001	1.138	-0.605	0.155	0.199	0.004	867
	B_2	1.409	-0.785	0.230	0.000	1.278	-0.679	0.184	0.130	-0.001	270
	B_3	1.515	-0.948	0.287	0.001	1.400	-0.843	0.239	0.116	0.002	238
the Great Lakes	L_1	1.099	-0.513	0.159	0.000	0.961	-0.437	0.109	0.134	0.001	283
	L_2	1.127	-0.555	0.191	0.001	0.980	-0.473	0.139	0.143	0.002	324
	L_3	1.087	-0.505	0.164	0.001	1.000	-0.457	0.135	0.086	0.001	107
the Gulf of Mexico	G_1	1.374	-0.798	0.302	-0.003	1.098	-0.614	0.214	0.276	-0.006	1165
	G_2	1.101	-0.416	0.118	0.000	0.989	-0.355	0.093	0.110	0.001	147
	G_3	1.163	-0.552	0.211	0.001	0.928	-0.428	0.137	0.227	0.000	730

Table 5.3.1

The table shows the estimated coefficients and difference of log-likelihood of *AR(3)* and *FAR*(*3*) models estimated by MLE.

5.4 One-step ahead forecasting

Let $P_t = \overline{sp}\{y_t, y_{t-1}, \cdots, y_{t-r+1}\}$ denotes the Hilbert space generated by the observations $\{y_t, y_{t-1}, \cdots, y_{t-r+1}\}$, then the best linear predictor (BLP) of y_{t+1} based on its finite past is given by

$$\tilde{y}_{t+1} = E[y_{t+1}|P_t] = \phi_{t1} y_t + \cdots + \phi_{t(t-r+1)} y_{t-r+1}, \tag{5.4.1}$$

where $\phi_t = (\phi_{t1}, \cdots, \phi_{t(t-r+1)})'$ is the unique solution of the linear equation

$$\Gamma_t \phi_t = \gamma_t, \tag{5.4.2}$$

with $\Gamma_t = [\gamma(i-j)]_{i,j=1,\cdots,t-r+1}$ and $\gamma_t = [\gamma(1), \cdots, \gamma(t-r+1)]'$.

On each training set $train_i$, we use the method introduce in 5.3 to build *AR*(*3*) model AR_i and *FAR*(*3*) model FAR_i. Based on formula (5.4.1), we use model AR_i and FAR_i to make one-step prediction on each testing set $test_i$, and get the *MSE* on the training set MSE_i^{MLE} and MSE_i^{FAR}, calculate the average of MSE_i^{MLE} to get MSE_1^{MLE} and calculate the average of MSE_i^{FAR} to get MSE_1^{FAR}.

We define the percent difference of MSE_1^* to MSE_1^{ANN} as $(\frac{MSE_1^* - MSE_1^{ANN}}{MSE_1^{ANN}}) \times 100\%$. In Table 5.4.1, we show percent difference of a comparison of MSE_1^{OLS}, MSE_1^{MLE} and MSE_1^{FAR} to MSE_1^{ANN}.

Some points here need to be considered:

(1) MSE_1^{MLE} is larger than MSE_1^{FAR}, since *AR*(*3*) model is nested in the *FAR*(*3*) model.

(2) For locations near the Great Lakes, the *FAR* models work best among all the linear time series models. For B_1 and G_3, it even works better than the ANN models. However, there is no guarantee that the *FAR* models always work better than the *AR* model estimated by OLS. This reveals that effects of the long-range memory depend on certain local factors and vary from place to place. For the locations where *FAR* model performs worse, then *AR* model estimated by MLE, also works worse than that by OLS. Because the two methods are equal if the variance of the random part are the same, the problem could come from the estimate of the variance.

Location \ Model		*AR*(*OLS*)	*FAR*		*AR*(*MLE*)	
		Percent Difference	MSE_1^{FAR}	Percent Difference	MSE_1^{MLE}	Percent Difference
the California Coast	B_1	0.57	0.207703	-0.21	0.209334	0.58
	B_2	0.64	0.153789	1.00	0.154167	1.25
	B_3	0.65	0.149890	0.49	0.150203	0.70
the Great Lakes	L_1	2.15	0.341666	1.89	0.342522	2.15
	L_2	2.21	0.324992	1.92	0.325926	2.21
	L_3	2.72	0.347532	2.62	0.347868	2.72
the Gulf of Mexico	G_1	1.49	0.095491	10.01	0.096604	11.29
	G_2	0.68	0.262349	0.89	0.262680	1.02
	G_3	0.47	0.248955	-0.03	0.250571	0.62

Table 5.4.1

The table shows the Percent Difference of AR(OLS), AR(MLE) and FAR models to the ANN model.

5.5 Multistep ahead forecasting

In principle, the *FAR* model would show more advantages in long-range forecasting. So, in this part, we compare the performance of the *FAR* model and *AR* model in multistep ahead forecasting.

The best linear predictor of y_{t+h} based on the finite past P_t is

$$\tilde{y}_t(h)=\phi_{t1}(h)y_t+\cdots+\phi_{t(t-r+1)}(h)y_{t-r+1}, \tag{5.5.1}$$

where $\phi_t(h)=[\phi_{t1}(h), \cdots, \phi_{t(t-r+1)}(h)]'$ is the unique solution of the linear equation

$$\Gamma_t\phi_t(h) = \gamma_t(h), \tag{5.5.2}$$

with $\Gamma_t=[\gamma(i-j)]_{i,j=h,\cdots,t-r+h}$ and $\gamma_t=[\gamma(h), \cdots, \gamma(t-r+h)]'$.

We calculate *h-step* ($h\in\{1, 7, 14, 21, 28, 35\}$) ahead forecasting of *FAR* models and *AR* models for different values of *h*, on each training set $test_i$ first to get MSE_{hi}^{FAR} and then we calculate the average of MSE_{hi}^{FAR} to get MSE_h^{FAR} . Similarly, we get MSE_h^{AR}. We provide MSE_h^{FAR} and MSE_h^{AR} in Appendix 5.5 (1).

Table 5.5.1 shows the ratio $r_h = \frac{MSE_h^{FAR}}{MSE_h^{AR}}$. Figure 5.5.1 shows r_h within the 3 areas. We can see r_h decreases first as h increases from one day to about 2 to 3 weeks. Then it becomes an increasing function of h. The decreasing part of is coincide with our expectation, while the increasing part could be attributed to the loss of predictability of the AR.

Location \ Model		$FAR(3)$				$AR(3)$			
		s				s			
		1	*3*	5	7	1	3	5	7
the California Coast	B_1	0.20757	0.20735	0.20755	0.20755	0.20918	0.20917	0.20917	0.20917
	B_2	0.15383	0.15371	0.15369	0.15367	0.15420	0.15409	0.15406	0.15405
	B_3	0.14985	0.14984	0.14983	0.14984	0.15016	0.15015	0.15014	1.05014
the Great Lakes	L_1	0.34155	0.34154	0.34153	0.34153	0.34238	0.34238	0.34238	0.34238
	L_2	0.32494	0.32493	0.32492	0.32492	0.32585	0.32585	0.32585	0.32585
	L_3	0.34744	0.34743	0.34743	0.34743	0.34776	0.34776	0.34776	0.34776
the Gulf of Mexico	G_1	0.09584	0.09550	0.09542	0.09538	0.09695	0.09661	0.09652	0.09648
	G_2	0.26223	0.26221	0.26221	0.26221	0.26255	0.26253	0.26253	0.26253
	G_3	0.24883	0.24882	0.24882	0.24882	0.25043	0.25042	0.25042	0.25042

Table 5.5.1

The table shows the ratio $r_h = \frac{MSE_h^{FAR}}{MSE_h^{AR}}$ to compare the performance of AR model and FAR model on multi-step predictions.

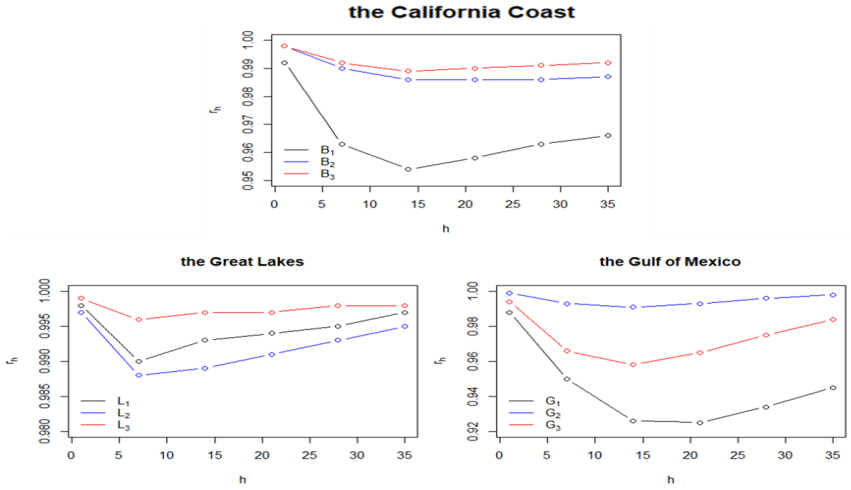


Figure 5.5.1

The Figure shows how $r_h = \frac{MSE_h^{FAR}}{MSE_h^{AR}}$ changes as h changes at different locations, from which we can compare the model performance.

5.6 Uncertainty of the forecasting

Except the prediction itself, we also interested in the uncertainty of the prediction. So, we will estimate the standard error of the prediction.

In on equation (5.6.1), we suppose the ϕ_t is fixed, then, all randomness comes from $\{y_t, y_{t-1}, \cdots, y_{t-r+1}\}$. We have

$$Var(\tilde{y}_t(h))=E(\tilde{y}_t(h)-y_{t+h})^2=E(\phi_{t1}(h)y_t+\cdots+\phi_{t(t-r+1)}(h)y_{t-r+1})^2 . \quad (5.6.1)$$

Notice that the RHS of equation (5.6.1) is just a linear combination of the autocovariance function and easy to calculate.

Suppose we have l independent simulations of temperatures of a n-year-long period,

Let

$$\widehat{Var}(\tilde{y}_t(h))=\begin{cases}\sum_{i=1}^{l}(\phi_{t1}(h)y_t^i+\cdots+\phi_{t(t-r+1)}(h)y_{(t-r+1)}^i)^2/l, & t>r\\ \sum_{i=1}^{l}(\phi_{t1}(h)y_t^i+\cdots+\phi_{tt}(h)y_1^i)^2/l, & t\le r\end{cases},$$

then $\widehat{Var}(\tilde{y}_t(h))$ is an estimate of $Var(\tilde{y}_t(h))$.

Let

$$\overline{Var}(\tilde{y}_t(h))=\sum_{i=0}^{n-1}\sum_{t=365i+152}^{t=365i+243}\widehat{Var}(\tilde{\boldsymbol{y}}_{\boldsymbol{t}\text{-}h}(h)) ,$$

then $\overline{Var}(\tilde{y}_t(h))$ can be considered as the theoretical values of the MSE_h^{FAR} we get in Part 5.5, so, if the model is fine, they should be close to each other.

We first calculate $\overline{Var}(\tilde{y}_t(h))$ on each training set $test_i$, to get $\overline{Var}(\tilde{y}_t(h))_i$ and calculate the average of $\overline{Var}(\tilde{y}_t(h))_i$ to get $\overline{Var}(\tilde{y}_t(h))_t$. $\overline{Var}(\tilde{y}_t(h))_t$ can be found in Appendix 5.6, we also provide the $\overline{\mathrm{Var}}(\tilde{y}_t(h))$ for AR models.

We can compare the values in Appendix 5.5 (1) and Appendix 5.6. In most cases, they are close, which reveals that the models are fine. But for G_1, as h increases, $\overline{Var}(\tilde{y}_t(h))$ and $\overline{Var}(\tilde{y}_t(h))$ become increasingly more different. Meanwhile percent difference of MSE_1^{FAR} of G_1 is very large, greater than 10%. It is possible that FAR model is not suitable for the place. B_2 has similar problems, but not as serious as that of G_1. AR model estimated by MLE also has similar problems.

Now I want to explain why the FAR model does not work in G_1 and B_2 and why AR model estimated by OLS seems to work better than both FAR model and AR model estimated by MLE. We compare the estimated coefficients of AR(MLE) and AR(OLS). In most locations, they are very close, but in G_1, $|\hat{\phi}_1^{OLS}|$ is much larger than $|\hat{\phi}_1^{MLE}|$ and $|\hat{\phi}_2^{OLS}|$ is much larger than $|\hat{\phi}_2^{MLE}|$, like on training set $train_1$, $\hat{\phi}_1^{OLS}$=1.62, $\hat{\phi}_2^{OLS}$=-1.01, while $\hat{\phi}_1^{MLE}$=1.37 and $\hat{\phi}_2^{MLE}$=-0.80. In comparison, for G_2, on training set $train_1$, $\hat{\phi}_1^{OLS}$=1.11, $\hat{\phi}_2^{OLS}$=-0.37, while $\hat{\phi}_1^{MLE}$=1.10 and $\hat{\phi}_2^{MLE}$=-0.42. Because the difference between $\hat{\phi}_3^{MLE}$ and $\hat{\phi}_3^{OLS}$ are much smaller, we eliminate then in the fowling discussion. If $T_{t-1}^*, T_{t-2}^* \in [a, 2]$, a >0 or $T_{t-1}^*, T_{t-2}^* \in [-2, a]$, a <0 , if we increase $|\hat{\phi}_1|$ and $|\hat{\phi}_2|$ simultaneously, the effect can be traded off. But if T_t^* are positive or negative

alternatively, things become different. We suppose T^*_{t-1}, T^*_{t-2}, T^*_{t-3} are positive negative, positive, separately. Figure shows 5.6.1 $\hat{h}_2(x_2)$ of G_1 and we focus on $\hat{h}_2(x_2)|_{x_2 \text{ usual}}$. Because $\hat{h}_2(x_2)|_{x_2 \text{ usual}}>0$, $\hat{\phi}_2$ are negative, if we use MLE and use T^*_{t-1}, T^*_{t-2} to estimate T^*_t and T^*_{t-2}, T^*_{t-3} to estimate T^*_{t-1} and denote by res_1 and res_2 the two residuals separately, res_1 are more likely to be positive while res_2 are more likely to be negative. Then if we increase $|\hat{\phi}_1|$ and $|\hat{\phi}_2|$ simultaneously, it is likely that res_1 decreases while res_2 increases but res_1 still positive while now res_2 positive. Because in the model we eliminate the intercept, the sum of residuals is not forced to be 0, so the case is possible. In this case, we may achieve a smaller MSE. This can only achieve when we use OLS because we suppose that the responses are conditionally independent given the past observations. Similar explanations can be generalized to explain why *AR*(OLS) works no worse than *AR*(MLE) and even better than *FAR* models in other locations.

It is difficult to check whether our explanation above is correct. But if it is right, there would be many residuals of *AR*(OLS) forced to be positive, so in G_1, the number of positive residuals of *AR*(OLS) should be larger than that of *AR*(MLE). The Table 5.6.1 below compare the residuals of G_1, G_2 and G_3, showing the number of positive residuals on each testing subset. We can see more residuals of *AR*(OLS) are positive than those of *AR*(MLE) in training subsets except $train_1$ in G_1, which does not happen in G_2 and G_3. This could suggests that our explanation is reasonable.

So, when the mean function is non-linear, the *AR*(OLS) model may have more "freedom" to achieve a smaller MSE, the cost of which is a systematical bias. It is difficult to decide whether it is better in this case.

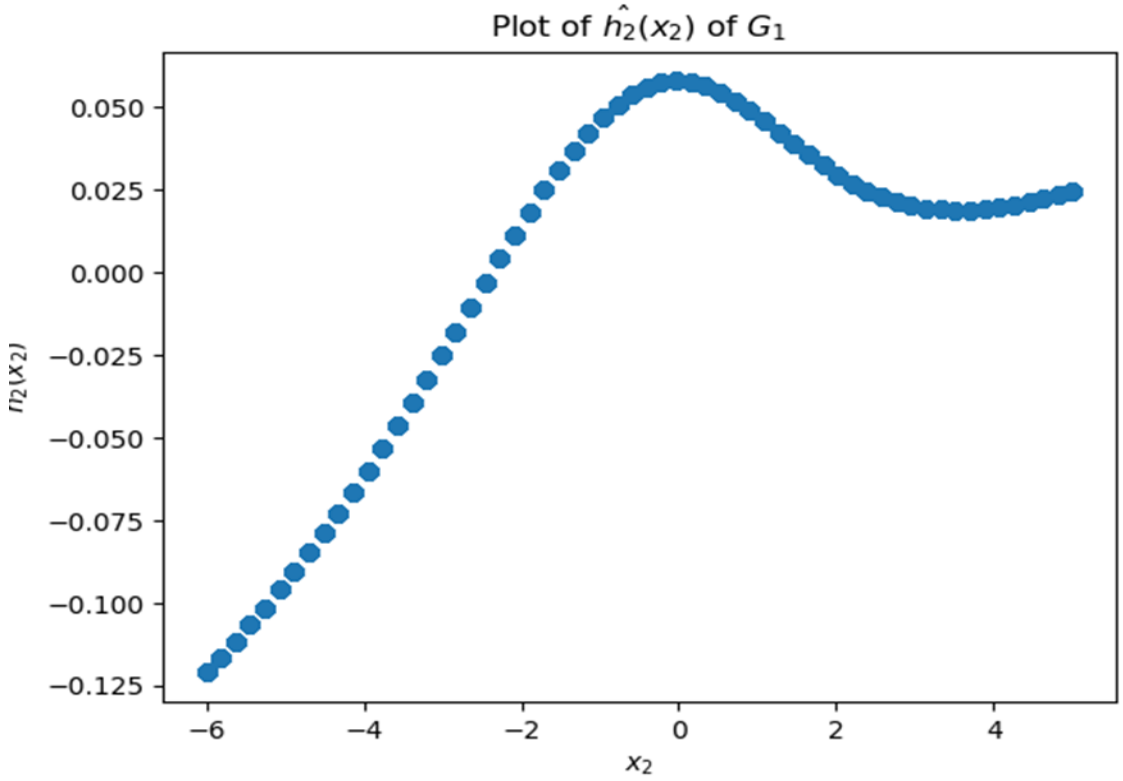


Figure 5.6.1

The figure shows $\hat{h}_2(x_2)$ of G_1

Location	Subset (i)	OLS	MLE	Difference
G_1	1	24174	24177	-3
	2	23898	23655	243
	3	24166	24006	160
	4	23750	23546	204
	5	23765	23493	272
G_2	1	24213	24273	-60
	2	24288	24273	15
	3	23893	23891	2
	4	23857	23883	-26
	5	24002	24012	-10
G_3	1	24199	24227	-28
	2	24152	24230	-78
	3	23680	23675	5
	4	23838	23878	-40
	5	23788	23836	-48

Table 5.6.1

The table shows the number of positive residuals of *AR*(OLS) and *AR*(MLE) on each testing sets near the Gulf of Mexico. On each testing subset, the total number of predictions are 46000. We use the data on the testing sets just because they are easy to track, the properties are similar on the training set.

5.7 Predictions for unnormalized observations

Compared with the normalized data, we are more interested in the performance of the model on the original data.

Theorem 5.7.1: If $\tilde{y}_t$ is the BLP of y_t, then $a+b\tilde{y}_t$ is the BLP of $a+by_t$ (a and b are constant).

We consider $\hat{v}(t)$ and $\hat{m}(t)$ are fixed. According to Theorem 5.7.1 and equation (2.2.1), the BLP of $\bar{\bar{T}}(t)$ is

$$\hat{\bar{\bar{T}}}^{BLP}(t)=\sqrt{\hat{v}(t)}\hat{T}_t^{BLP}+\hat{m}(t). \tag{5.7.1}$$

We map the $\hat{T}_t^{BLP}$ attained from Part 5.8 back according to equation (5.7.1) to get $\hat{\bar{\bar{T}}}^{BLP}(t)$ on each testing set $test_i$, and get the *MSE* of $\hat{\bar{\bar{T}}}^{BLP}(t)$ on $test_i$, MSE_i^{FAR}, calculate the average of MSE_i^{FAR} to get MSE_1^{FAR} .

In Table 5.7.2, we provide the $\sqrt{MSE_1^{FAR}}$, which shows, on average, the absolute difference of the prediction is less than 0.7^0C, for the Great Lakes, it is a bit large, about 1^0C to 1.4^0C.

Location	the California Coast			the Great Lakes			the Gulf of Mexico		
	B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
$\sqrt{MSE_1^{FAR}}$	0.45	0.54	0.50	1.33	1.19	1.24	0.66	0.50	0.31

Table 5.7.1

The table shows $\sqrt{MSE_1^{FAR}}$, from which we can see on average how many degrees Celsius our predictions are different from the true values.

We want to compare the model performance of ANN, *FAR* model *AR* model estimated by OLS and MLE on the original data. We compare the percent differences of MSE_1^{FAR}, MSE_1^{MLE} and MSE_1^{OLS} to MSE_1^{ANN}which is defined in Part 5.4 in Table 5.7.2. Here we get MSE_1^{AR} to MSE_1^{ANN} in the same way as we get MSE_1^{FAR}.

Location \ Model		*FAR*		*AR*(*OLS*)		*AR*(*MLE*)	
		MSE_1^{FAR}	Percent Difference	MSE_1^{OLS}	Percent Difference	MSE_1^{MLE}	Percent Difference
the California Coast	B_1	0.201545	-0.22	0.203125	0.57	0.203138	0.57
	B_2	0.289347	1.18	0.287951	0.69	0.290053	1.43
	B_3	0.245789	0.66	0.245741	0.64	0.245845	0.69
the Great Lakes	L_1	1.777938	2.01	1.782014	2.24	1.781987	2.24
	L_2	1.424283	2.03	1.428055	2.30	1.428053	2.30
	L_3	1.527450	2.69	1.528870	2.79	1.528865	2.79
the Gulf of Mexico	G_1	0.442092	10.05	0.407630	1.47	0.447216	11.32
	G_2	0.254330	0.88	0.253876	0.69	0.254662	1.00
	G_3	0.095009	-0.01	0.095483	0.49	0.095618	0.63

Table 5.7.2

The table compares the percent differences of MSE_1^{FAR}, MSE_1^{MLE} and MSE_1^{OLS} to MSE_1^{ANN} of the original data.

Basically, percent differences are enlarged the after taking the variation of the original data into account. Performances of different models do not change much. Where the *FAR* model performs well like the Great Lakes, it still works good, where performs bad still bad. But in places the performances of *FAR* and *AR* model are close, the comparative performances can change a bit, like in B_3, *FAR* model works better on the normalized data while a bit worse on the raw data.

We then want to look into the length of the predictability of the *AR* model and *FAR* model. Based on formula (5.7.1), map the h-step ahead predictions we get from Part 5.5 back to the original scale, and use the same method as that in Part 5.5 to get MSE_h^{FAR} of the original data, which is in Appendix 5.7 (2).

On each training set train_i, compute the average of $\bar{\bar{T}}(t)$ through the 40 runs, then we use the average to estimate $\bar{\bar{T}}(t)$ and get *MSE* on train_i, then compute the average of *MSE* on train_i, which is denoted by MSE^{Ave}. It is in Appendix 5.7 (3).

In Appendix 5.7 (3), we show the percentile reduction of MSE_h^{FAR} and MSE_h^{AR} from MSE^{Ave}, R_h^{FAR} and R_h^{AR}, respectively. The property of R_h^{AR} is distinct. R_h^{AR} reduced to about *2%* in most cases when *h* is about 2 to 3 weeks and decreases very slow after that. For all the locations, it "converges" to about *2%*. So, we think the predictability of *AR* model can be effective for about 2 to 3 weeks. R_h^{AR} is more complicated. For some places which the *FAR* model works well, like B_1, even h is larger than 4 weeks, R_h^{FAR} is still quite large, reduces stable and fast. For places like this, the predictability may last longer than 4 weeks. If we do longer forecast, we may still get some useful information. In places like L_1, after about 3 weeks, although R_h^{FAR} still decreases, but it decreases very slow and tend to "converge", for these places, the predictability may last about 3 weeks. When *h* is large than that, we may get little information but not much. G_1 is again a very special case for the above analysis. Although based on the former analysis, we think *FAR* model may not be suitable for this place, R_h^{FAR} there is still very large, reduces stable and fast even *h* is larger than 4 weeks. Considering that we take the past random part into account in the *FAR* model, and the difference between MLE estimate and OLS estimate of *AR* model, the problem could happen to the estimate of the variance when we use a linear model to approximate the non-linear model. Figure 5.7.1 shows R_h^{FAR} of the examples we give above.

We can also map $\widehat{Var}(\tilde{y}_t(h))$ attained from Part 5.6 back to the original scale to evaluate the uncertainty of the prediction and test whether the models are fine.

In Appendix 5.7 (4), we provide $\overline{Var}(\tilde{y}_t(h))_t$ of the original scale, we can compare that with the MSE_h^{FAR} in Appendix 5.7 (1). Similar with the case of the normalized data, the model in G_1 is problematic, and in B_2, the model is also problematic but not so serious.

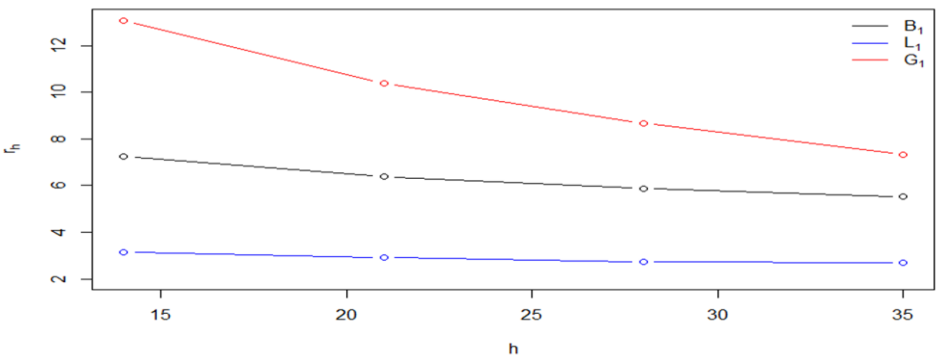


Figure 5.7.1

Plots of R_h^{FAR} in B_1, L_1 and G_1, to show the trend when h is large, we begin from $h = 14$

Finally, we want to compare the performance of ANN and *FAR* models on multi-step predictions. We try step$\in\{3, 5, 7, 14, 21, 28, 35\}$. We've got the outcome of *FAR* models in Appendix 5.7 (3). For ANN model, we use the models from Part 3.3 to make multi-step predictions. Suppose we predict T_t and step = h:

(1) Take T_{t-h}, T_{t-h-1} and T_{t-h-2} as predictors to predict T_{t-h+1}, get $\hat{T}_{t-h+1}$.

Take $\hat{T}_{t-h+1}$, T_{t-h}, and T_{t-h-1} as predictors to predict T_{t-h+2}, get $\hat{T}_{t-h+2}$.

…

Take $\hat{T}_{t-1}$, $\hat{T}_{t-2}$ and $\hat{T}_{t-3}$ as predictor to predict T_t, get $\hat{T}$.

(2) Use formula (5.7.1) to get $\hat{\hat{\hat{T}}}_t$.

(3) Calculate MSE of the original data.

(4) Repeat procedures (1) to (3) on each testing set $test_i$, get MSE_{hi}^{ANN}. Calculate the average of MSE_{hi}^{ANN} to get MSE_{ht}^{ANN} and denote by MSE_h^{ANN}.

We provide MSE_{hi}^{ANN} and MSE_{ht}^{ANN} in Appendix 5.7. (5). In Table 5.7.3, we provide the percent difference of MSE_h^{FAR} to MSE_h^{ANN}. We can see, when h is larger than 3, in all locations, $MSE_h^{FAR} < MSE_h^{ANN}$. When h = 3, just in the locations where the *FAR* model performs really worse for the one-step prediction, like G_1, $MSE_h^{FAR} > MSE_h^{ANN}$. This shows the advantage of the *FAR* model over ANN on multi-step prediction. Besides, the percent difference first decreases then increases as h increases. This could also be due to the loss of predictability of the ANN model. ANN model could be more suitable for fit the data but not for long-time ahead predictions.

Location \ h		3	5	7	14	21	28	35
the California Coast	B_1	-2.07	-3.18	-3.71	-4.80	-4.73	-4.38	-4.05
	B_2	-0.05	-0.68	2.41	1.84	-1.50	-1.49	-1.38
	B_3	-0.28	-0.66	-1.01	-1.57	-1.64	-1.57	-1.47
the Great Lakes	L_1	-0.96	-2.72	-3.42	-3.65	-3.50	-3.31	-3.30
	L_2	-1.04	-4.16	-6.00	-8.01	-8.10	-7.94	-7.77
	L_3	-0.76	-4.64	-7.09	-9.87	-10.10	-10.06	-10.02
the Gulf of Mexico	G_1	4.64	-2.58	-4.41	-6.04	-6.62	-6.25	-5.39
	G_2	0.49	-0.83	-1.47	-2.13	-2.07	-1.88	-1.71
	G_3	-1.46	-2.57	-3.20	-4.07	-3.44	-2.47	-1.64

Table 5.7.3

This table shows the percent difference of MSE_h^{FAR} to MSE_h^{ANN}, which is defined as $(\frac{MSE_h^{FAR} - MSE_h^{ANN}}{MSE_h^{ANN}})\times 100\%$

6 Conclusion

Under our assumption in the previous parts and evaluating models from the point of view of *MSE*, we can get some important conclusions:

(1) Conditional mean function of T_t , $E(T_t|\cdot)$ can be well approximate by function $\hat{f}_3$ of lag order 3. $\hat{f}_3$ is nonlinear, but close to additive. The nonlinearity varies from location to location.

(2) $AR(3)$ model can be used to model T_t . In most cases, this approximation will not loss a lot. Different estimation methods give outcomes with some differences. The ordinary least square method performs "no worse" than the MLE, which we explain in Chapter 5.6.

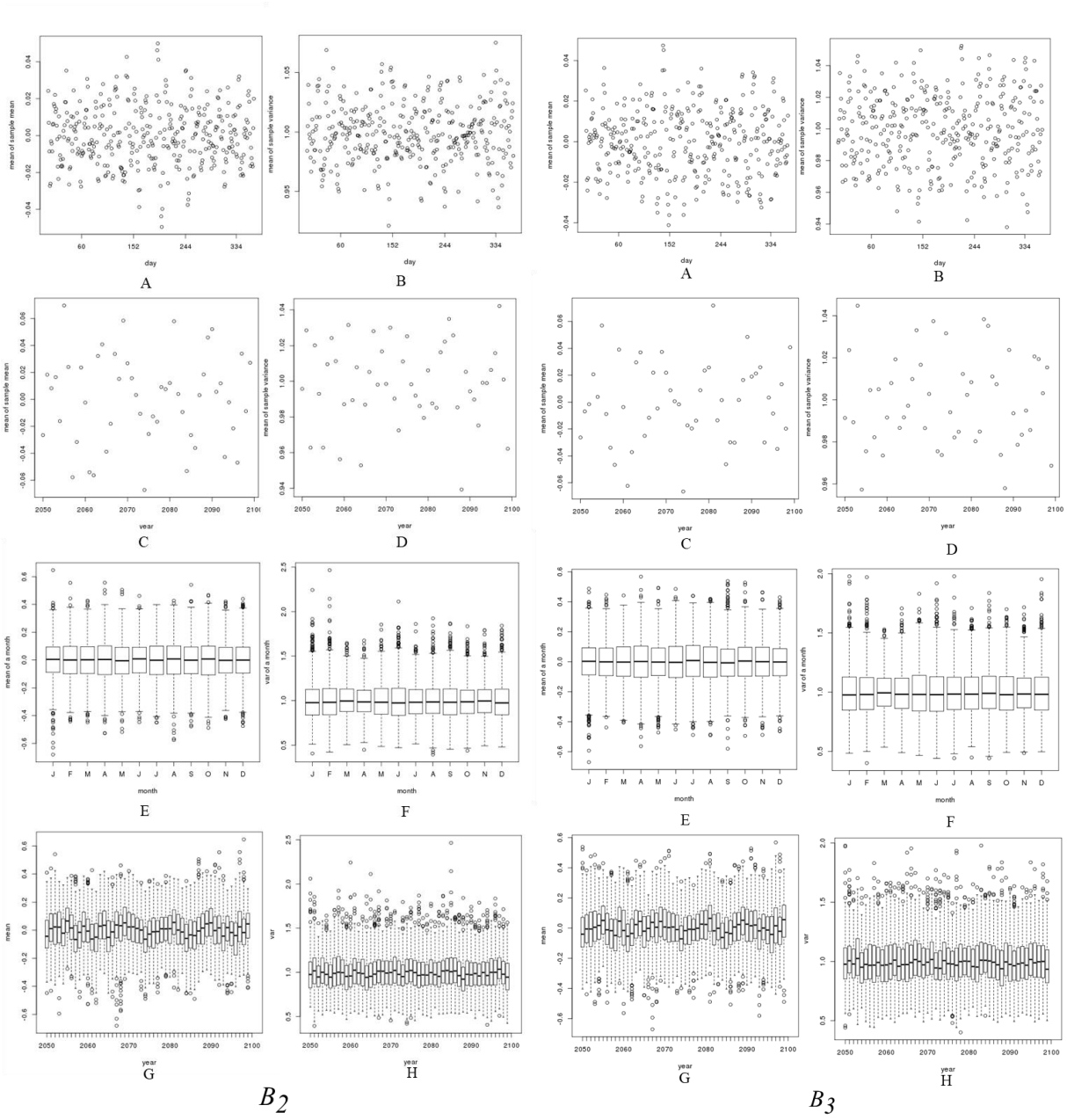
(3) Because the long-memory parameter d is always significant, the long-range effect exists in the temperature data. However, the effect varies from place to place. For model fitting or one-step prediction, in some places, the $FAR(3)$ model works even better than the non-linear ANN model. In most places, the loss is not large. But when the non-linear function $\hat{f}_3$ have certain features, the $FAR(3)$ could work poor, and the problems may happens to the estimate of the variance of the random part because we use linear function to approximate the nonlinear mean function. But for long-time ahead prediction, $FAR(3)$ model works much better than the ANN model.

(4) Comparing the $AR(3)$ model and the $FAR(3)$ model both estimated by MLE, the predictability of $FAR(3)$ model lasts longer than the $AR(3)$ model. For $AR(3)$ model, the predictability can last about 2 to 3 weeks. And the differences among different locations are not significant. In contrast, the length of predictability of the $FAR(3)$ model varies from place to place. Although its predictability is always better than the $AR(3)$ model, in some places, it can last longer than 4 to 5 weeks, while in some places, it may only last about 3 weeks.

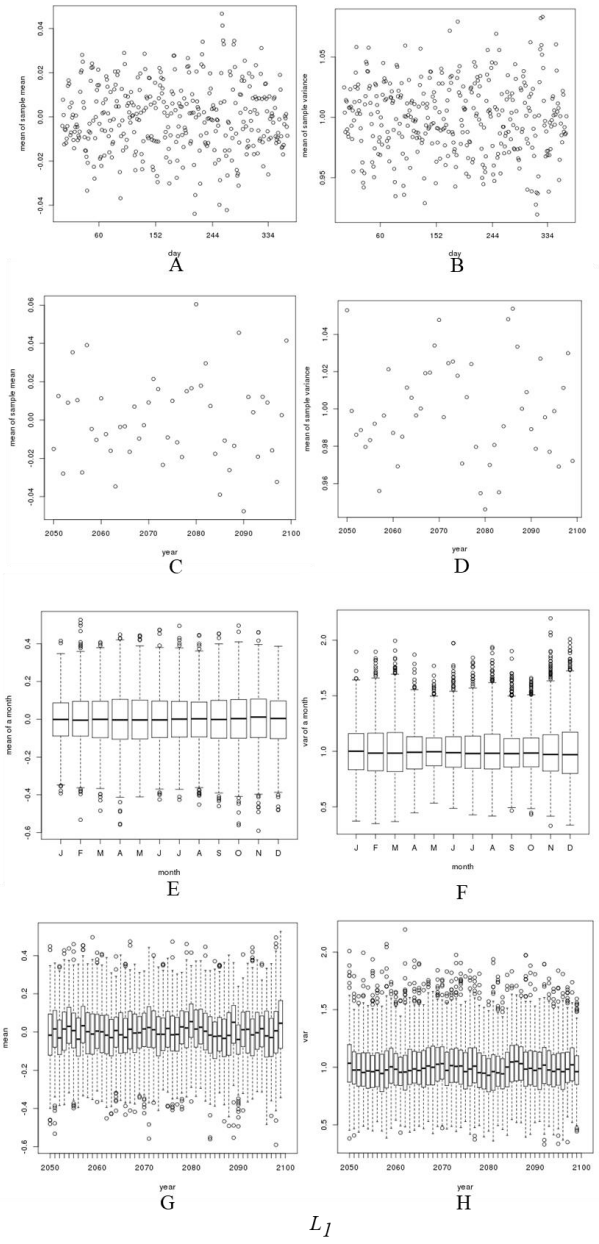
A Appendix

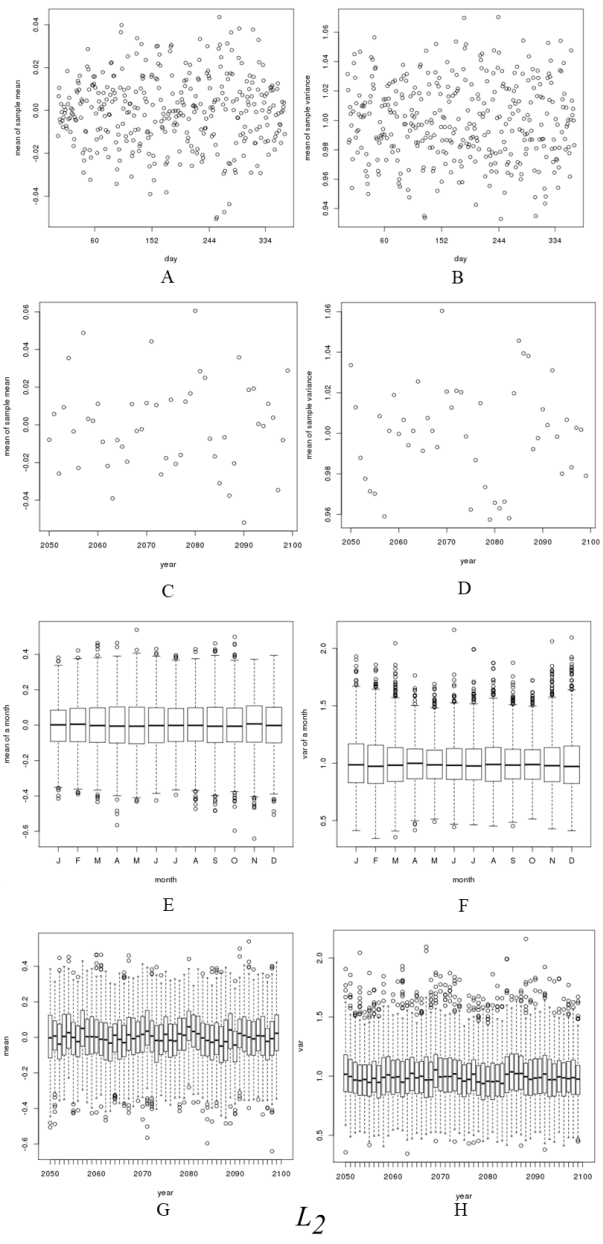
2.3

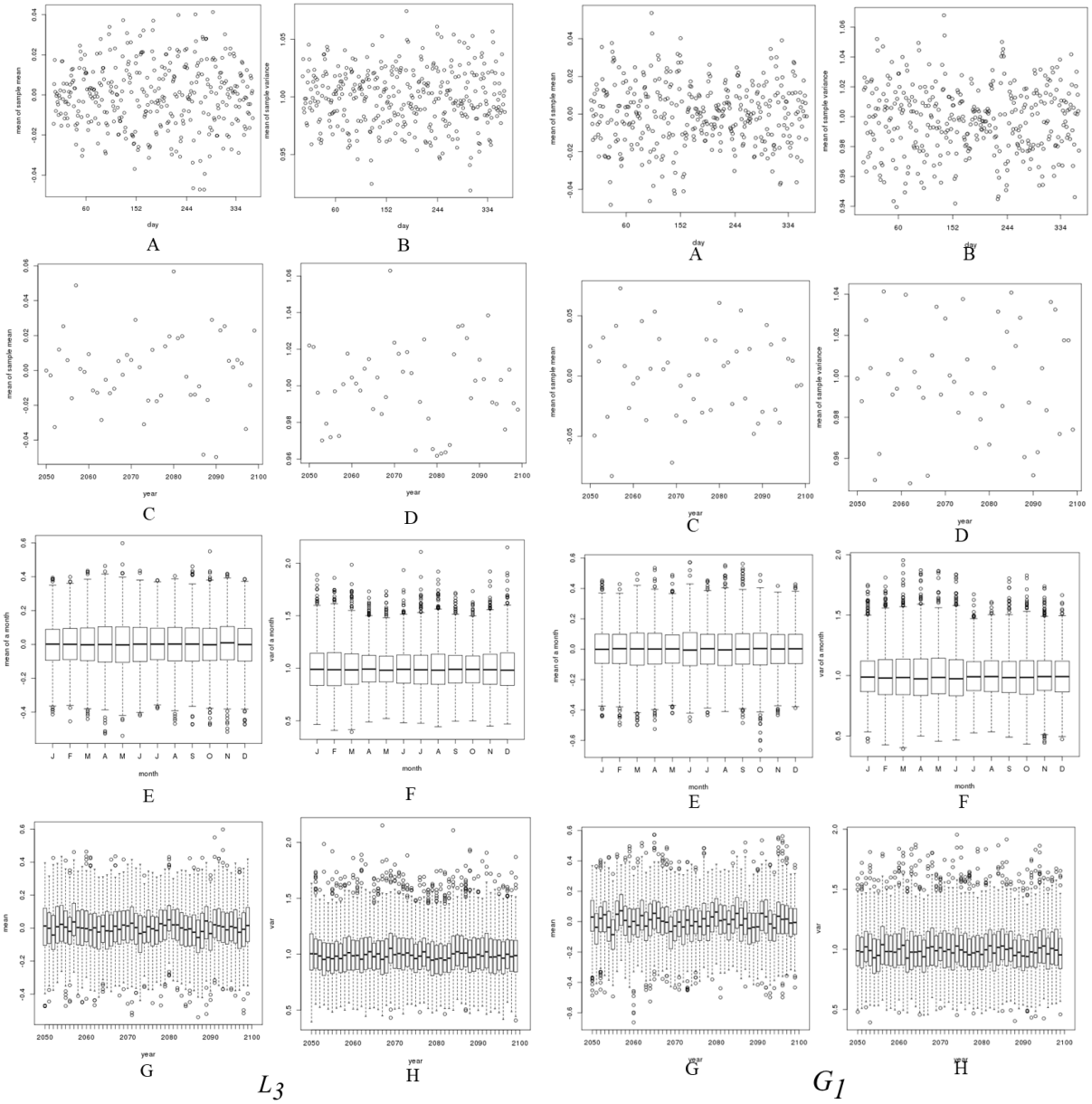
The figures below are the plots of T_t corresponding to Part 2.3 for all the locations except B_1, which are shown in Figure 2.3.1.



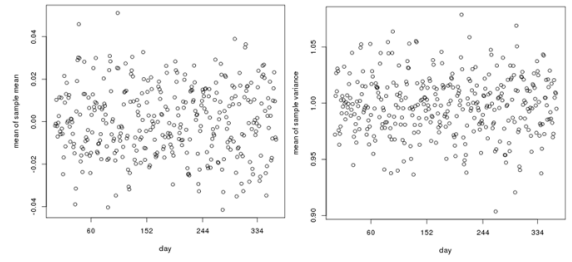
B_2 B_3

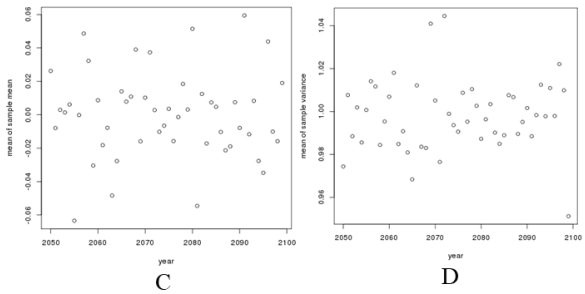

mean of sample mean
mean of sample variance
day
A
B
mean of sample mean
mean of sample variance
year
C
D
mean of a month
var of a month
month
E
F
mean
var
year
G
H
L_1

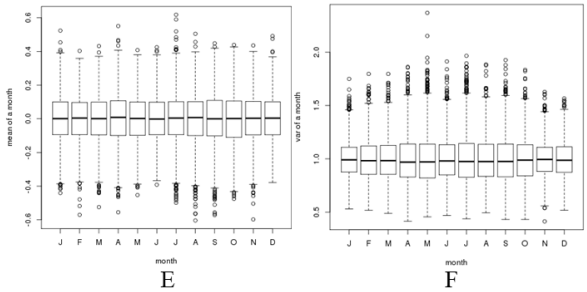

mean of sample mean
mean of sample variance
day
A
B
mean of sample mean
mean of sample variance
year
C
D
mean of a month
var of a month
month
E
F
mean
var
year
G
H
L_2

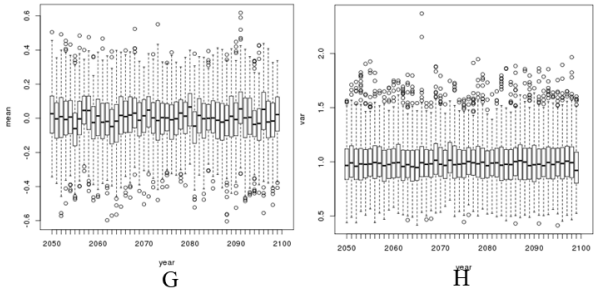

mean of sample mean
day
A
mean of sample variance
day
B
mean of sample mean
year
C
mean of sample variance
year
D
mean of a month
month
E
var of a month
month
F
mean
year
G
var
year
H
mean of sample mean
day
A
mean of sample variance
day
B
mean of sample mean
year
C
mean of sample variance
year
D
mean of a month
month
E
var of a month
month
F
mean
year
G
var
year
H

L_3 G_1

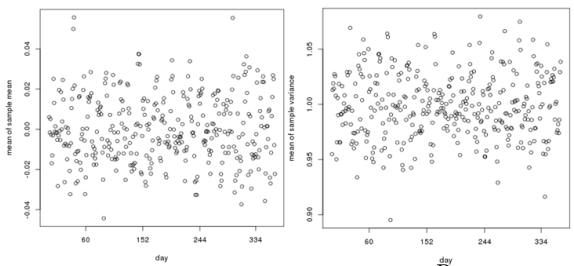

mean of sample mean
day
A
mean of sample variance
day
B

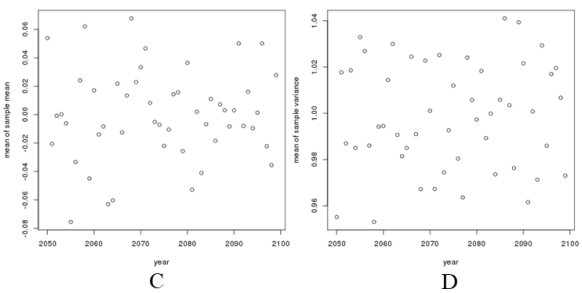

mean of sample mean
year
C
mean of sample variance
year
D

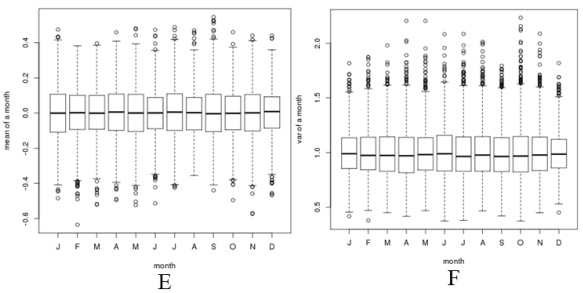

mean of a month
month
E
var of a month
month
F

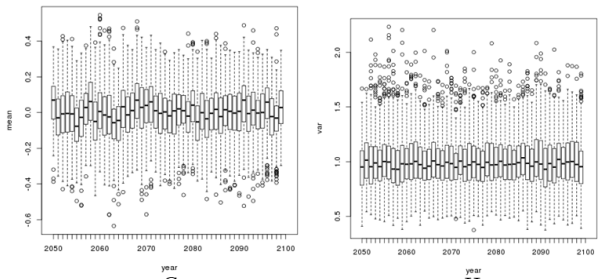

mean
year
G
var
year
H

G_2


mean of sample mean
day
A
mean of sample variance
day
B


mean of sample mean
year
C
mean of sample variance
year
D

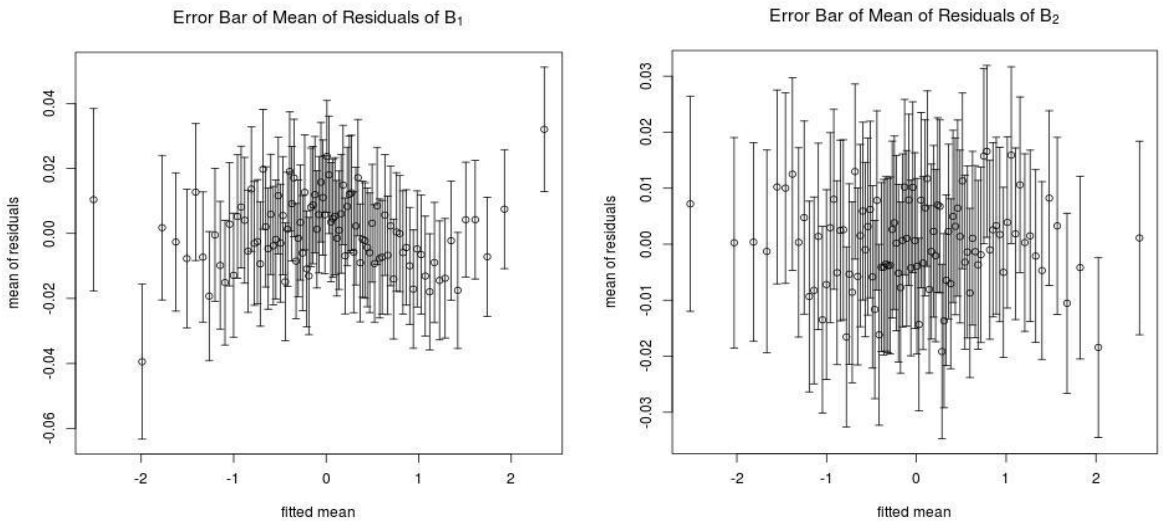

mean of a month
month
E
var of a month
month
F

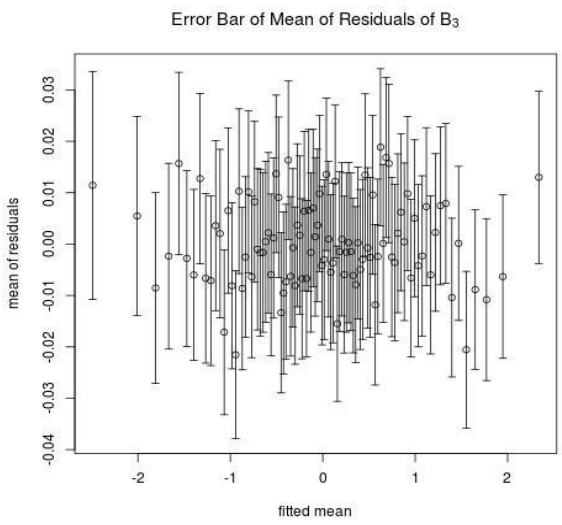

mean
year
G
var
year
H

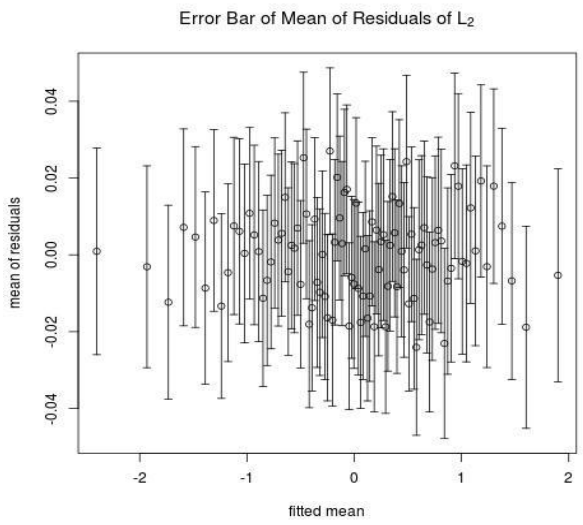
G_3

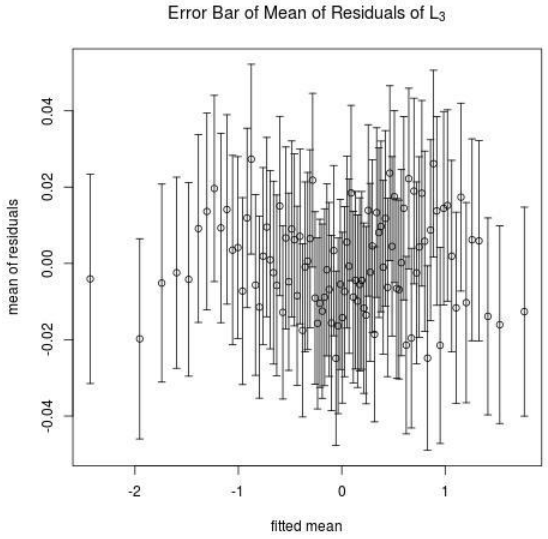
3.1

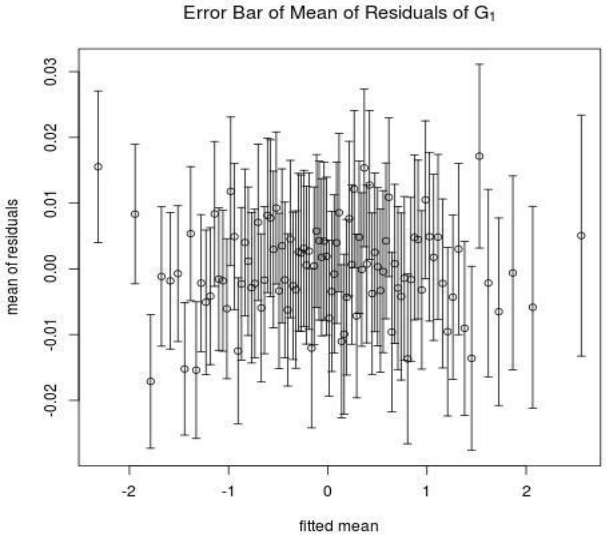
In this part, we briefly explain how we estimate use ANN to estimate $E(T_t|\cdot)$. We use python function sklearn.neural_network.MLPRegressor to estimate the model. We consider the data as time series data for each location separately, so, it has no complicated structures, we just use one hidden layer with 50 neurons to estimate the model. And based on our experience, increasing the number of layers or neurons cannot improve the fitting. We use the sigmoid function for the activation. We choose the solver 'lbfgs', a quasi-Newton Method. Here we need to notice that as we mentioned in 3.1, we need to check whether the error term $\hat{\varepsilon}_t$ is independent of $\hat{f}_i$, which suggests whether the model is ample if the assumption holds. In general, the residuals $\hat{\varepsilon}_t$ from stochastic gradient methods like "sgd" and "adam" are highly correlated with $\hat{f}$ but with different patterns in different runs. One method to solve this problem is to build different stochastic models and use the averaged model. We tried that methods and build 40 independent stochastic models to average models, the $\hat{\varepsilon}_t$ perform better, showing a much weaker correlation with $\hat{f}$ but the models work worse than 'lbfgs' method and building many models is not efficient. We set the batch size to be 92, the length of a summer. All other parameters are the default ones. The figures below are error bars of the mean of residuals versus the estimated mean of all locations except L_1.

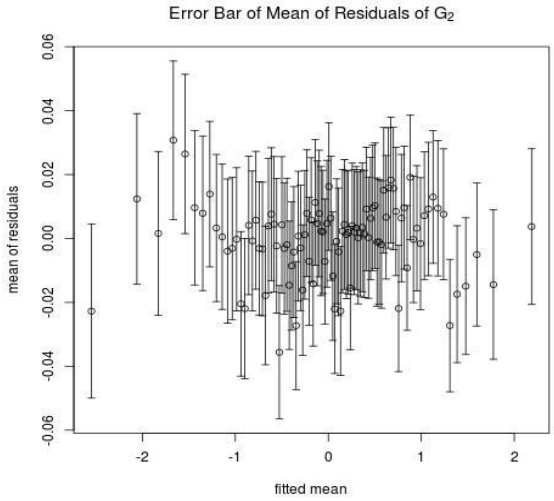


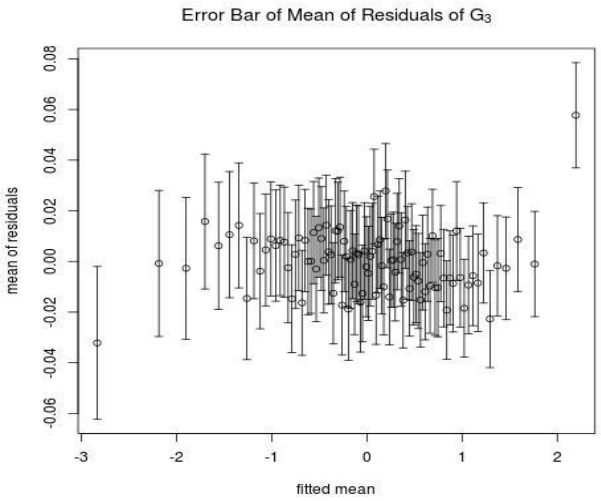

Error Bar of Mean of Residuals of B_3
mean of residuals
fitted mean


Error Bar of Mean of Residuals of L_2
mean of residuals
fitted mean


Error Bar of Mean of Residuals of L_3
mean of residuals
fitted mean


Error Bar of Mean of Residuals of G_1
mean of residuals
fitted mean


Error Bar of Mean of Residuals of G_2
mean of residuals
fitted mean


Error Bar of Mean of Residuals of G_3
mean of residuals
fitted mean

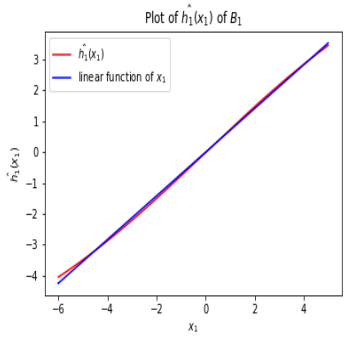
3.2

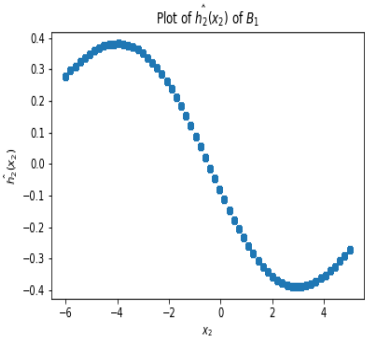
(1) Table below shows how the *MSE* changes in the second loop for all the locations, r_i' represents r_i in the second loop.

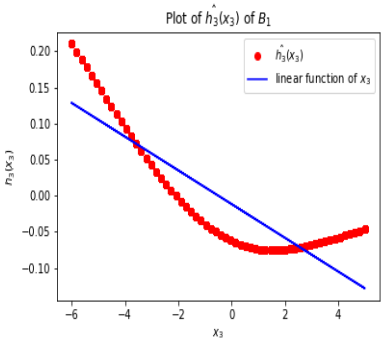
Location \ i		$\tilde{T}'$	r_1'	r_2'	r_3'
the California Coast	B_1	0.2515414	0.2515418	0.2515430	0.2515432
	B_2	0.1893602	0.1893629	0.1893634	0.1893633
	B_3	0.0722586	0.0722586	0.0722595	0.0722598
the Great Lakes	L_1	0.0410371	0.0410378	0.0410381	0.0410393
	L_2	0.1058308	0.1058309	0.1058309	0.1058312
	L_3	0.13303368	0.13303375	0.13303497	0.13303523
the Gulf of Mexico	G_1	0.1619355	0.1619357	0.1619357	0.1619368
	G_2	0.0106419	0.0106419	0.0106419	0.0106419
	G_3	0.0599974	0.0599974	0.0599975	0.0599975

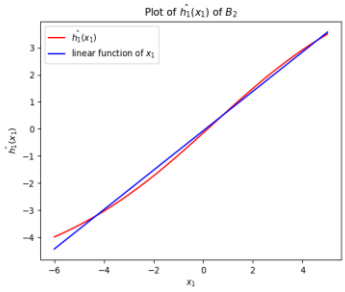
(2)

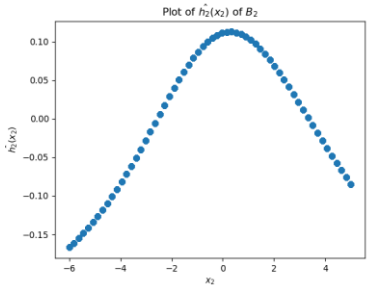
The figures below are of the three main effects of all locations.

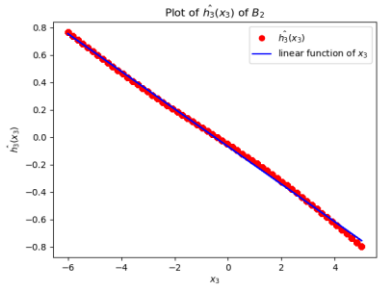

Plot of $\hat{h_1}(x_1)$ of B_1
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

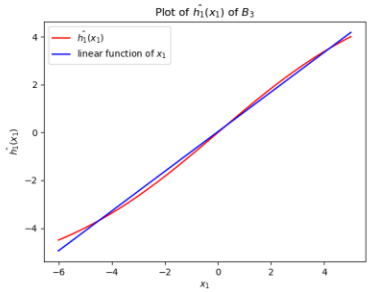

Plot of $\hat{h_2}(x_2)$ of B_1
$\hat{h_2}(x_2)$
x_2

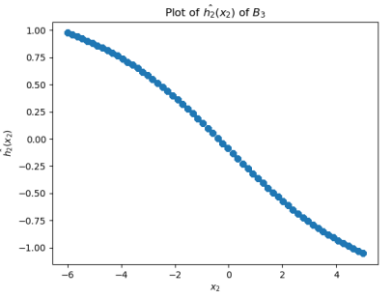

Plot of $\hat{h_3}(x_3)$ of B_1
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

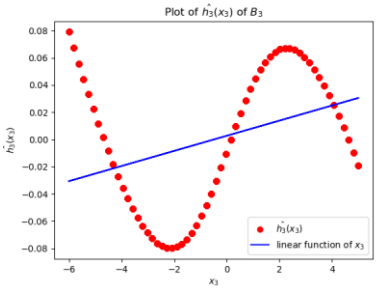

Plot of $\hat{h_1}(x_1)$ of B_2
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

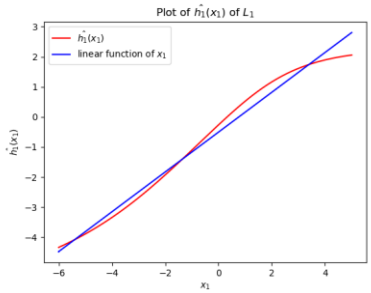

Plot of $\hat{h_2}(x_2)$ of B_2
$\hat{h_2}(x_2)$
x_2

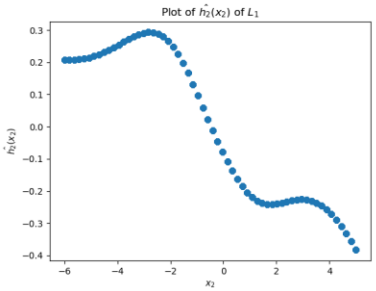

Plot of $\hat{h_3}(x_3)$ of B_2
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

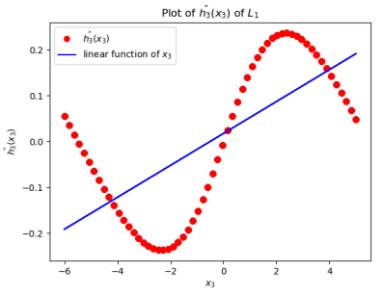

Plot of $\hat{h_1}(x_1)$ of B_3
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

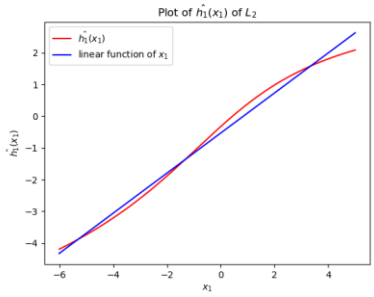

Plot of $\hat{h_2}(x_2)$ of B_3
$\hat{h_2}(x_2)$
x_2

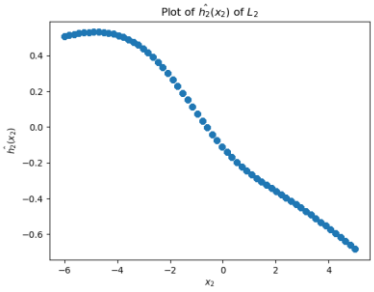

Plot of $\hat{h_3}(x_3)$ of B_3
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

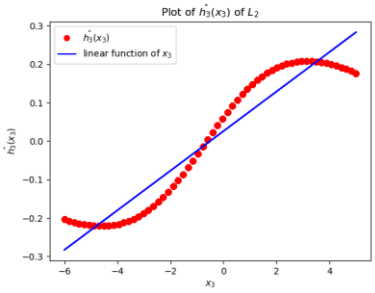

Plot of $\hat{h_1}(x_1)$ of L_1
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

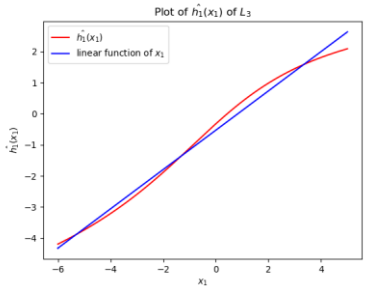

Plot of $\hat{h_2}(x_2)$ of L_1
$\hat{h_2}(x_2)$
x_2

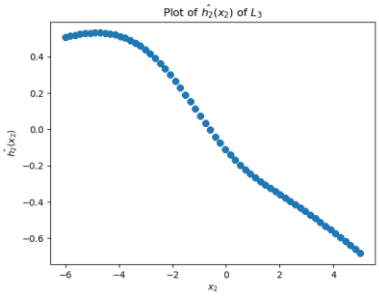

Plot of $\hat{h_3}(x_3)$ of L_1
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

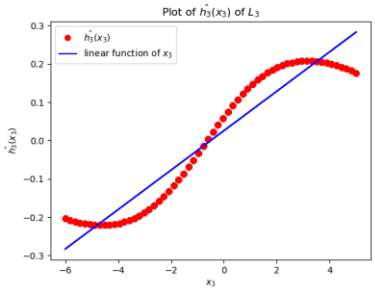

Plot of $\hat{h_1}(x_1)$ of L_2
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

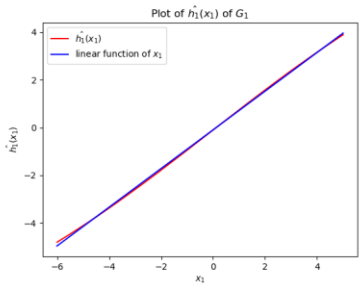

Plot of $\hat{h_2}(x_2)$ of L_2
$\hat{h_2}(x_2)$
x_2

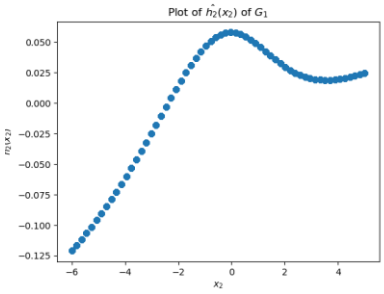

Plot of $\hat{h_3}(x_3)$ of L_2
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

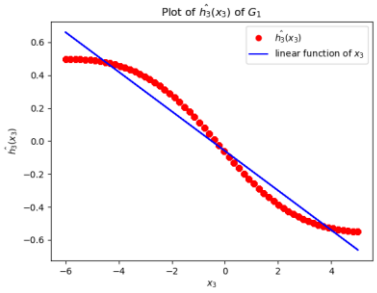

Plot of $\hat{h_1}(x_1)$ of L_3
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

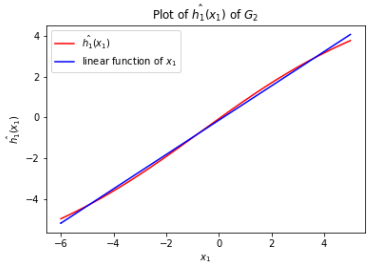

Plot of $\hat{h_2}(x_2)$ of L_3
$\hat{h_2}(x_2)$
x_2

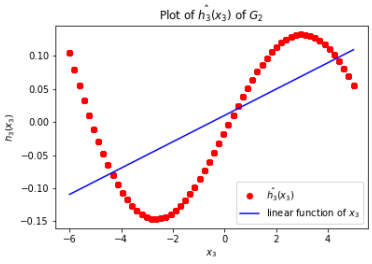

Plot of $\hat{h_3}(x_3)$ of L_3
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

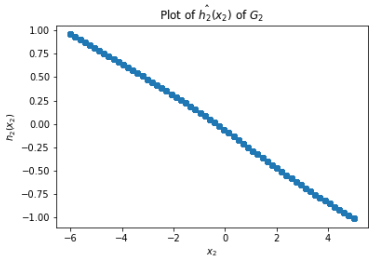

Plot of $\hat{h_1}(x_1)$ of G_1
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1

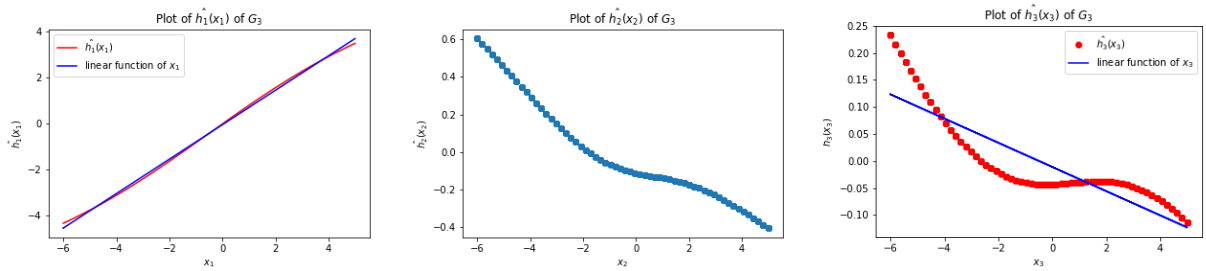

Plot of $\hat{h_2}(x_2)$ of G_1
x_2


Plot of $\hat{h_3}(x_3)$ of G_1
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

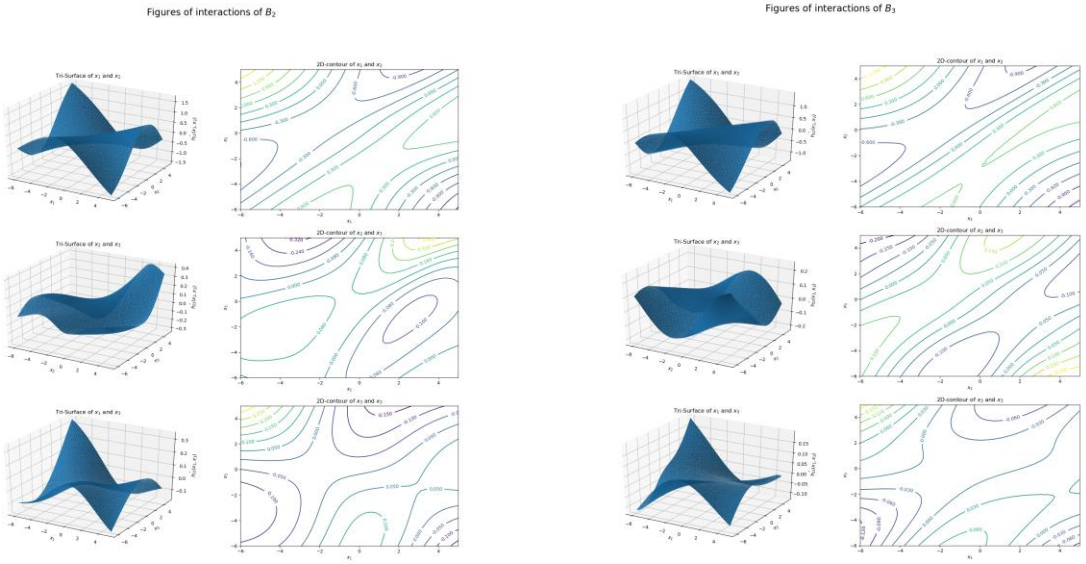

Plot of $\hat{h_1}(x_1)$ of G_2
$\hat{h_1}(x_1)$
linear function of x_1
$\hat{h_1}(x_1)$
x_1


Plot of $\hat{h_3}(x_3)$ of G_2
$\hat{h_3}(x_3)$
linear function of x_3
$\hat{h_3}(x_3)$
x_3

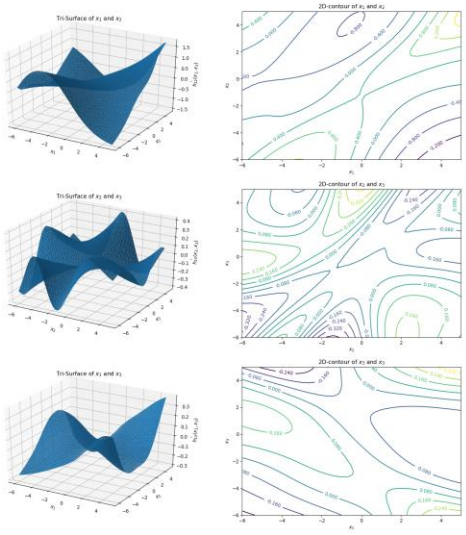

Plot of $\hat{h_2}(x_2)$ of G_2
$\hat{h_2}(x_2)$
x_2



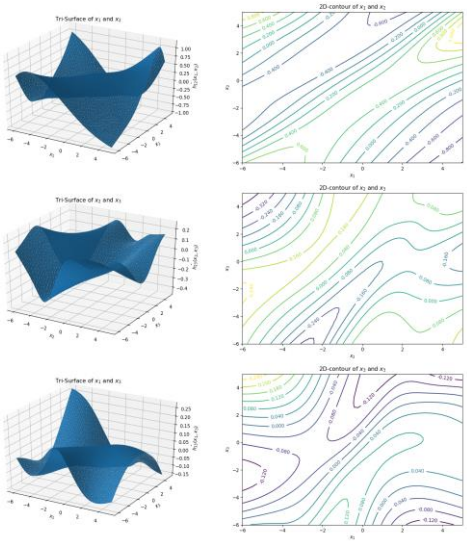
The figures below are the Tri-surfaces and 2D contours of all two-way interactions of all locations except B_1.



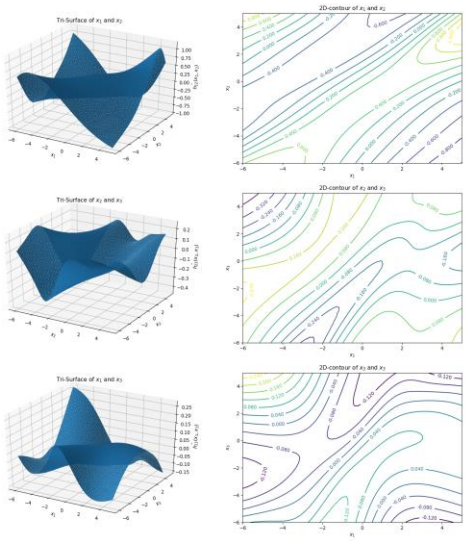
Figures of interactions of L_1



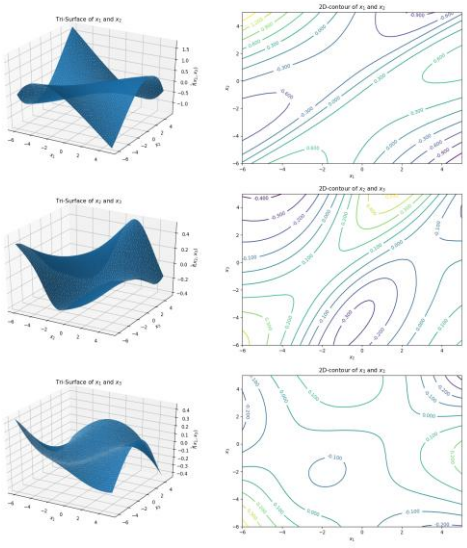
Figures of interactions of L_2

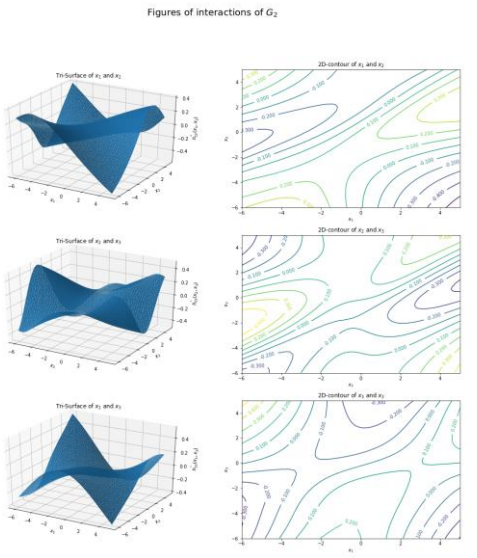


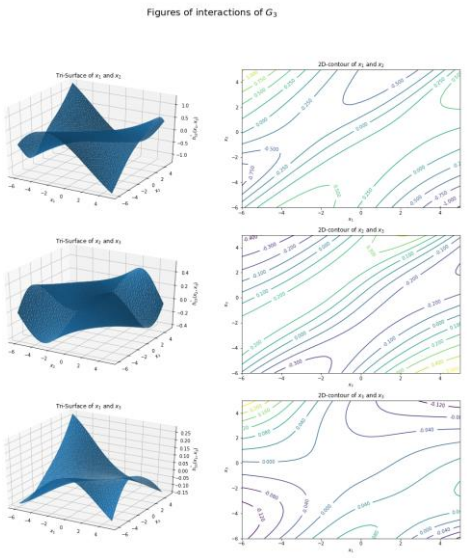
Figures of interactions of L_3



Figures of interactions of G_1







3.3

The table below is *MSE* of the ANN model on the training sets corresponding to Part 3.3.

Location \ MSE_*^{ANN}		1	2	3	4	5	t
the California Coast	B_1	0.208911	0.208235	0.207591	0.207668	0.207689	0.208019
	B_2	0.152865	0.152285	0.151703	0.152779	0.151341	0.152195
	B_3	0.148972	0.149666	0.148972	0.149013	0.148742	0.149073
the Great Lakes	L_1	0.335202	0.333587	0.334410	0.335893	0.335772	0.334973
	L_2	0.317743	0.317849	0.317924	0.319076	0.320013	0.318521
	L_3	0.337354	0.337942	0.337817	0.337817	0.339250	0.338036
the Gulf of Mexico	G_1	0.086594	0.086804	0.086374	0.086898	0.086890	0.086712
	G_2	0.259934	0.260077	0.259466	0.259831	0.259664	0.259794
	G_3	0.249384	0.248933	0.247874	0.249174	0.248867	0.248846

4.2

(1) The table shows the estimated coefficients of $AR(3)$ models.

Location \ $\hat{\phi}$		$\hat{\phi}_1$					$\hat{\phi}_2$					$\hat{\phi}_3$				
	$train_i$	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
the California Coast	B_1	1.352	1.353	1.352	1.353	1.352	-0.765	-0.767	-0.764	-0.767	-0.765	0.242	0.244	0.242	0.243	0.242
	B_2	1.481	1.483	1.482	1.48	1.483	-0.878	-0.883	-0.879	-0.877	-0.879	0.258	0.261	0.259	0.258	0.259
	B_3	1.537	1.534	1.535	1.535	1.535	-0.978	-0.976	-0.975	-0.975	-0.975	0.297	0.297	0.295	0.295	0.295
the Great Lakes	L_1	1.099	1.102	1.099	0.097	1.098	-0.513	-0.515	-0.511	-0.509	-0.513	0.161	0.161	0.159	0.158	0.161
	L_2	1.128	1.129	1.128	1.126	1.126	-0.555	-0.556	-0.556	-0.552	-0.556	0.19	0.19	0.19	0.19	0.192
	L_3	1.089	1.088	1.088	1.086	1.086	-0.508	-0.506	-0.504	-0.503	-0.506	0.165	0.164	0.164	0.165	0.166
the Gulf of Mexico	G_1	1.615	1.618	1.617	1.617	1.615	-1.007	-1.016	-1.01	-1.014	-1.01	0.314	0.32	0.315	0.318	0.317
	G_2	1.107	1.106	1.105	1.106	1.105	-0.367	-0.367	-0.365	-0.369	-0.364	0.075	0.077	0.077	0.079	0.075
	G_3	1.174	1.174	1.174	1.172	1.173	-0.531	-0.531	-0.531	-0.53	-0.531	0.188	0.188	0.19	0.19	0.19

(2) The table shows the roots of the *AR* polynomials.

roots \ Location			the California Coast			the Great Lakes			the Gulf of Mexico		
	$train_i$		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
x_1	1	R	1.31	1.28	1.30	1.44	1.38	1.43	1.14	1.33	1.24
		I	0	0	0	0	0	0	0	0	0
	2	R	1.31	1.28	1.31	1.44	1.38	1.43	1.14	1.32	1.24
		I	0	0	0	0	0	0	0	0	0
	3	R	1.30	1.28	1.31	1.44	1.38	1.42	1.14	1.32	1.24
		I	0	0	0	0	0	0	0	0	0
	4	R	1.31	1.28	1.31	1.45	1.38	1.42	1.15	1.32	1.24
		I	0	0	0	0	0	0	0	0	0
	5	R	1.31	1.28	1.31	1.44	1.38	1.42	1.14	1.33	1.24
		I	0	0	0	0	0	0	0	0	0
x_2	1	R	0.93	1.06	0.99	0.87	0.77	0.83	1.03	1.78	0.79
		I	1.52	1.38	1.26	1.88	1.79	1.89	1.31	2.62	1.91
	2	R	0.93	1.05	0.99	0.88	0.77	0.83	1.02	1.72	0.79

		I	1.51	1.37	1.26	1.88	1.79	1.89	1.30	2.62	1.91
	3	*R*	0.93	1.06	1.00	0.89	0.77	0.83	1.03	1.71	0.78
		I	1.52	1.38	1.26	1.89	1.79	1.90	1.31	2.63	1.91
	4	*R*	0.93	1.06	1.00	0.89	0.76	0.81	1.02	1.67	0.78
		I	1.52	1.38	1.26	1.89	1.80	1.90	1.30	2.60	1.91
	5	*R*	0.93	1.06	1.00	0.87	0.76	0.81	1.02	1.76	0.78
		I	1.52	1.38	1.26	1.88	1.79	1.89	1.31	2.64	1.91
x_3	*1*	*R*	0.93	1.06	0.99	0.87	0.77	0.83	1.03	1.78	0.79
		I	-1.52	-1.38	-1.26	-1.88	-1.79	-1.89	-1.31	-2.62	-1.91
	2	*R*	0.93	1.05	0.99	0.88	0.77	0.83	1.02	1.72	0.79
		I	-1.51	-1.37	-1.26	-1.88	-1.79	-1.89	-1.30	-2.62	-1.91
	3	*R*	0.93	1.06	1.00	0.89	0.77	0.83	1.03	1.71	0.78
		I	-1.52	-1.38	-1.26	-1.89	-1.79	-1.90	-1.31	-2.63	-1.91
	4	*R*	0.93	1.06	1.00	1.44	0.76	0.81	1.02	1.67	0.78
		I	-1.52	-1.38	-1.26	-1.89	1.80	-1.90	-1.30	-2.60	-1.91
	5	*R*	0.93	1.06	1.00	0.87	0.76	0.81	1.02	1.76	0.78
		I	-1.52	-1.38	-1.26	-1.88	-1.79	-1.89	-1.31	-2.64	-1.91

(3) The table shows MSE_{i2} and MSE_{t2} for $AR(3)$ models.

MSE_{*2} / Location		1	2	3	4	5	t
the California Coast	B_1	0.210048	0.209601	0.208683	0.208991	0.208896	0.209244
	B_2	0.153742	0.153313	0.152679	0.153825	0.152342	0.153180
	B_3	0.149952	0.150693	0.149973	0.150035	0.149784	0.150087
the Great Lakes	L_1	0.342831	0.341024	0.341597	0.343298	0.343343	0.342419
	L_2	0.325378	0.325329	0.325026	0.326399	0.327119	0.325850
	L_3	0.346925	0.347605	0.347193	0.348558	0.348649	0.347786
the Gulf of Mexico	G_1	0.087900	0.088158	0.087650	0.088238	0.088323	0.088054
	G_2	0.261894	0.262088	0.261280	0.261679	0.261638	0.261716
	G_3	0.250689	0.250148	0.249094	0.250370	0.250181	0.250096

5.3

(1) The table shows the log-likelihood of $AR(3)$ and $FAR(3)$ models based on different past time s.

Location \ Model		$AR(3)$				$FAR(3)$			
		s				s			
		1	3	5	7	1	3	5	7
the California Coast	B_1	-147045	-146860	-146814	-146792	-146180	-145991	-145946	-145924
	B_2	-120260	-119581	-119404	-119320	-119987	-119309	-119133	-119050
	B_3	-110193	-109975	-109918	-109890	-109953	-109738	-109680	-109652
the Great Lakes	L_1	-203203	-203135	-203130	-203130	-202929	-202929	-202847	-202848
	L_2	-197427	-197456	-197464	-197473	-197110	-197132	-197139	-197147
	L_3	-204953	-204924	-204918	-204920	-204852	-204820	-204811	-204813
the Gulf of Mexico	G_1	-108618	-107058	-106651	-106455	-107447	-105893	-105487	-105290
	G_2	-178456	-178232	-178193	-178193	-178306	-178084	-178046	-178042
	G_3	-171825	-171682	-171660	-171660	-171086	-170952	-170930	-170930

(2) The table shows the MSE of one-step ahead prediction of $FAR(3)$ and $AR(3)$ models with different threshold s.

Location \ Model		$FAR(3)$				$AR(3)$			
		s				s			
		1	*3*	5	7	1	3	5	7
the California Coast	B_1	0.20757	0.20735	0.20755	0.20755	0.20918	0.20917	0.20917	0.20917
	B_2	0.15383	0.15371	0.15369	0.15367	0.15420	0.15409	0.15406	0.15405
	B_3	0.14985	0.14984	0.14983	0.14984	0.15016	0.15015	0.15014	1.05014
the Great Lakes	L_1	0.34155	0.34154	0.34153	0.34153	0.34238	0.34238	0.34238	0.34238
	L_2	0.32494	0.32493	0.32492	0.32492	0.32585	0.32585	0.32585	0.32585
	L_3	0.34744	0.34743	0.34743	0.34743	0.34776	0.34776	0.34776	0.34776
the Gulf of Mexico	G_1	0.09584	0.09550	0.09542	0.09538	0.09695	0.09661	0.09652	0.09648
	G_2	0.26223	0.26221	0.26221	0.26221	0.26255	0.26253	0.26253	0.26253
	G_3	0.24883	0.24882	0.24882	0.24882	0.25043	0.25042	0.25042	0.25042

(3) The table shows the roots of the AR Polynomials.

Location \ Model		*AR*(3)						*FAR*(3)					
		Solutions of the AR polynomial						Solutions of the AR polynomial					
		x_1		x_2		x_3		x_1		x_2		x_3	
		R	I	R	I	R	I	R	I	R	I	R	I
the California Coast	B_1	1.30	0.00	0.92	1.53	0.92	-1.53	1.82	0.00	1.04	1.57	1.04	-1.57
	B_2	1.28	0.00	1.07	1.50	1.07	-1.50	1.48	0.00	1.11	1.57	1.11	-1.57
	B_3	1.30	0.00	1.00	1.29	1.00	-1.29	1.48	0.00	1.02	1.33	1.02	-1.33
the Great Lakes	L_1	1.45	0.00	0.89	1.88	0.89	-1.88	1.91	0.00	1.05	1.92	1.05	-1.92
	L_2	1.38	0.00	0.76	1.79	0.76	-1.79	1.73	0.00	0.84	1.86	0.84	-1.86
	L_3	1.43	0.00	0.83	1.89	0.83	-1.89	1.63	0.00	0.88	1.94	0.88	-1.94
the Gulf of Mexico	G_1	1.17	0.00	0.74	1.51	0.74	-1.51	1.52	0.00	0.68	1.62	0.68	-1.62
	G_2	1.32	0.00	1.10	2.28	1.10	-2.28	1.50	0.00	1.16	2.41	1.16	-2.41
	G_3	1.25	0.00	0.69	1.83	0.69	-1.83	1.68	0.00	0.72	1.96	0.72	-1.96

5.5

(1)

h \ Location		the California Coast			the Great Lakes			the Gulf of Mexico		
		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
1	*FAR*	0.207703	0.153789	0.149890	0.341666	0.324992	0.347532	0.095491	0.262349	0.248955
	AR	0.209334	0.154167	0.150203	0.342522	0.325926	0.347868	0.096604	0.262681	0.250571
7	*FAR*	0.931072	0.965264	0.980997	0.979929	0.972512	0.985752	0.820009	0.967290	0.909444
	AR	0.967041	0.975357	0.988667	0.989698	0.984062	0.989609	0.863218	0.974425	0.941046
14	*FAR*	0.949592	0.984275	0.987834	0.993754	0.988244	0.996785	0.886983	0.989587	0.946694
	AR	0.995267	0.998104	0.998510	1.001233	0.999522	0.999870	0.958146	0.998404	0.988246
21	*FAR*	0.958494	0.987432	0.990455	0.996154	0.991256	0.997698	0.914303	0.994349	0.964439
	AR	1.000395	1.001318	1.000624	1.001743	1.000548	1.000219	0.988326	1.000941	0.999210
28	*FAR*	1.001150	0.987922	0.991758	0.997242	0.993370	0.998281	0.931594	0.996951	0.976820
	AR	0.963722	1.001723	1.000925	1.001768	1.000637	1.000241	0.997837	1.001241	1.001363
35	*FAR*	0.967202	0.988898	0.992864	0.998256	0.995292	0.998736	0.945550	0.998779	0.985972

	AR	1.001255	1.001799	1.000971	1.001770	1.000647	1.000243	1.000751	1.001280	1.001777

(2) The table below shows the ratio shows the ratio $r_h = \frac{MSE_h^{FAR}}{MSE_h^{AR}}$.

5.6 The table below shows $\overline{Var}(\tilde{y}_t(h))_t$ of different step *h*.

Location / h		the California Coast			the Great Lakes			the Gulf of Mexico		
		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
1	*FAR*	0.208317	0.155188	0.151570	0.346426	0.326446	0.351063	0.096354	0.264466	0.251681
	AR	0.210118	0.155490	0.151956	0.347818	0.327514	0.351532	0.097499	0.264982	0.253521
7	*FAR*	0.927094	0.861146	0.957310	0.994208	0.977891	0.997090	0.491255	0.899138	0.867449
	AR	0.973806	0.871497	0.964235	1.012077	0.995720	1.004658	0.511874	0.910119	0.909866
14	*FAR*	0.955456	0.890572	0.982375	1.005821	0.993788	1.006086	0.536110	0.921856	0.907549
	AR	0.996444	0.898191	0.985937	1.017065	1.004540	1.010381	0.559128	0.929303	0.944360
21	*FAR*	0.964217	0.895498	0.986329	1.009013	0.997799	1.007515	0.550448	0.925241	0.920133
	AR	0.996991	0.899016	0.986475	1.017091	1.004631	1.010418	0.564098	0.929704	0.945929
28	*FAR*	0.968943	0.897673	0.988051	1.010658	0.999848	1.008178	0.558362	0.926709	0.927001
	AR	0.997005	0.899042	0.986488	1.017091	1.004632	1.010418	0.564641	0.929712	0.946002
35	*FAR*	0.971995	0.898974	0.989075	1.011685	1.001129	1.008575	0.563633	0.927580	0.931496
	AR	0.997005	0.899043	0.986489	1.017091	1.004632	1.010418	0.564700	0.929712	0.946005

5.7

(1) The table below shows MSE_1^{FAR} of the unnormalized data $\overline{\overline{T}}(t)$

i / Location		1	2	3	4	5	*t*
the California Coast	B_1	0.200113	0.203356	0.200199	0.201927	0.202130	0.201545
	B_2	0.286274	0.291869	0.288093	0.287757	0.292744	0.289347
	B_3	0.246715	0.244850	0.246451	0.246451	0.244476	0.245789
the Great Lakes	L_1	1.764595	1.775464	1.786658	1.768591	1.794384	1.777938
	L_2	1.428992	1.420508	1.426903	1.430853	1.414159	1.424283

	L_3	1.538196	1.526485	1.523665	1.529381	1.519522	1.527450
the Gulf of Mexico	G_1	0.444730	0.436269	0.455061	0.440870	0.433531	0.442092
	G_2	0.254796	0.254299	0.258400	0.252458	0.251698	0.254330
	G_3	0.095880	0.094046	0.094705	0.095680	0.094733	0.095009

(2) The table below shows MSE_h^{FAR} and MSE_h^{AR} of the original data $\bar{\bar{T}}(t)$.

Location / h		the California Coast			the Great Lakes			the Gulf of Mexico		
		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
1	*FAR*	0.201545	0.289348	0.245317	1.777938	1.424283	1.52745	0.442092	0.25433	0.095009
	AR	0.203138	0.290053	0.245845	1.781987	1.428053	1.528865	0.447216	0.254662	0.095618
3	*FAR*	0.78467	1.433762	1.292634	4.747927	3.939646	4.027132	2.644956	0.74854	0.289687
	AR	0.802726	1.441859	1.29881	4.778142	3.968855	4.03769	2.715799	0.751195	0.294962
5	*FAR*	0.886874	1.788901	1.567001	5.017375	4.190659	4.270731	3.495177	0.891347	0.331312
	AR	0.916507	1.804411	1.578028	5.057824	4.230779	4.285021	3.639107	0.896555	0.340434
7	*FAR*	0.904467	1.861427	1.604553	5.078239	4.251416	4.329583	3.801522	0.931779	0.346085
	AR	0.939641	1.881089	1.617804	5.122972	4.296752	4.345701	4.000348	0.938743	0.357855
14	*FAR*	0.922601	1.905949	1.617565	5.144348	4.313124	4.374291	4.117116	0.953195	0.36024
	AR	0.967339	1.933189	1.635946	5.178703	4.358638	4.387994	4.443852	0.961774	0.375661
21	*FAR*	0.931184	1.912273	1.622101	5.155943	4.324555	4.37799	4.245487	0.957747	0.366893
	AR	0.972333	1.939595	1.639558	5.181226	4.362785	4.389434	4.584026	0.964214	0.379776
28	*FAR*	0.936266	1.913557	1.624213	5.166253	4.334092	4.381203	4.325915	0.960278	0.371457
	AR	0.97307	1.940424	1.640057	5.181358	4.363178	4.389542	4.627919	0.964504	0.38058
35	*FAR*	0.939758	1.915842	1.626041	5.166906	4.342187	4.383221	4.389604	0.962052	0.374777
	AR	0.973174	1.940579	1.640133	5.181367	4.363217	4.38955	4.641232	0.964543	0.380733
MSE^{Ave}		0.994769	1.986099	1.677221	5.311063	4.474479	4.500744	4.736623	0.986800	0.390052

(3) The table show the percentile reduction of MSE_h^{FAR} and MSE_h^{AR} from MSE^{Ave}.

Location / h		the California Coast			the Great Lakes			the Gulf of Mexico		
		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
1	*FAR*	79.74	85.43	85.37	66.52	68.17	66.06	90.67	74.23	75.64
	AR	79.58	85.40	85.34	66.45	68.08	66.03	90.56	74.19	75.49
3	*FAR*	21.12	27.81	22.93	10.60	11.95	10.52	44.16	24.14	25.73
	AR	19.31	27.40	22.56	10.03	11.30	10.29	42.66	23.88	24.38
5	*FAR*	10.85	9.93	6.57	5.53	6.34	5.11	26.21	9.67	15.06
	AR	7.87	9.15	5.91	4.77	5.45	4.79	23.17	9.15	12.72
7	*FAR*	9.08	6.28	4.33	4.38	4.99	3.80	19.74	5.58	11.27
	AR	5.54	5.29	3.54	3.54	3.97	3.44	15.54	4.87	8.25
14	*FAR*	7.25	4.04	3.56	3.14	3.61	2.81	13.08	3.41	7.64
	AR	2.76	2.66	2.46	2.49	2.59	2.51	6.18	2.54	3.69
21	*FAR*	6.39	3.72	3.29	2.92	3.35	2.73	10.37	2.94	5.94
	AR	2.26	2.34	2.25	2.44	2.50	2.47	3.22	2.29	2.63
28	*FAR*	5.88	3.65	3.16	2.73	3.14	2.66	8.67	2.69	4.77
	AR	2.18	2.30	2.22	2.44	2.49	2.47	2.29	2.26	2.43
35	*FAR*	5.53	3.54	3.05	2.71	2.96	2.61	7.33	2.51	3.92
	AR	2.17	2.29	2.21	2.44	2.49	2.47	2.01	2.26	2.39

(4) The table shows $\overline{Var}(\tilde{y}_t(h))_t$ of the original scale.

Location / h		the California Coast			the Great Lakes			the Gulf of Mexico		
		B_1	B_2	B_3	L_1	L_2	L_3	G_1	G_2	G_3
1	*FAR*	0.202475	0.300696	0.248381	1.791117	1.423755	1.541267	0.447167	0.254983	0.095675
	AR	0.204224	0.301278	0.249010	1.798297	1.428420	1.543330	0.452503	0.255486	0.096372
3	*FAR*	0.780345	1.324528	1.268510	4.790762	3.932489	4.059555	1.801627	0.723816	0.282074
	AR	0.801129	1.331955	1.273777	4.784138	3.973251	4.076073	1.842689	0.727391	0.288378
5	*FAR*	0.880073	1.600678	1.529315	5.076066	4.184518	4.310966	2.114841	0.832741	0.315091
	AR	0.915684	1.615430	1.539043	5.157432	4.250211	4.338997	2.183519	0.840390	0.326662
7	*FAR*	0.901098	1.668578	1.568745	5.140362	4.264975	4.377529	2.279872	0.866896	0.329759
	AR	0.946495	1.688615	1.580100	5.232703	4.342749	4.410760	2.375667	0.877502	0.345874
14	*FAR*	0.928667	1.725597	1.609838	5.200410	4.334312	4.417024	2.488040	0.888000	0.345003
	AR	0.968499	1.740340	1.615664	5.258488	4.381217	4.435885	2.594982	0.895998	0.358987
21	*FAR*	0.937183	1.735144	1.616318	5.216915	4.351806	4.423295	2.554584	0.892064	0.349787
	AR	0.969031	1.741938	1.616546	5.258624	4.381612	4.436045	2.618045	0.896385	0.359584
28	*FAR*	0.941776	1.739359	1.619141	5.225421	4.360744	4.426206	2.591311	0.893480	0.352398
	AR	0.969044	1.741988	1.616568	5.258625	4.381616	4.430046	2.620567	0.896393	0.359611
35	*FAR*	0.944743	1.741879	1.620819	5.230732	4.366329	4.427949	2.615772	0.894319	0.354106
	AR	0.969045	1.741990	1.616568	5.258625	4.381616	4.436046	2.620843	0.896393	0.359612

(5) The table show the percentile reduction of MSE_h^{ANN} of the original scale.

Location \ h	i		3	5	7	14	21	28	35
the California coast	B_1	1	0.798615	0.918246	0.944220	0.985358	0.996566	0.998422	0.998726
		2	0.812959	0.931143	0.956573	0.985801	0.994406	0.996498	0.996881
		3	0.794545	0.905956	0.926265	0.950658	0.956781	0.957898	0.958083
		4	0.799624	0.914653	0.937167	0.959918	0.967172	0.969290	0.969769
		5	0.800560	0.910137	0.932322	0.963617	0.971938	0.973514	0.973824
		t	0.801261	0.916027	0.939309	0.96907	0.977373	0.979124	0.979457
	B_2	1	1.426048	1.798955	1.881346	1.881346	1.953488	1.954819	1.955128
		2	1.445290	1.827393	1.827393	1.914378	1.986209	1.987653	1.987855
		3	1.425706	1.780736	1.780736	1.81147	1.897415	1.898284	1.898401
		4	1.439298	1.810562	1.810562	1.887805	1.939143	1.940262	1.940528
		5	1.435752	1.788065	1.788065	1.862669	1.930230	1.931323	1.931600
		t	1.434419	1.801142	1.81762	1.871534	1.941297	1.942468	1.942702
	B_3	1	1.295011	1.581396	1.629781	1.659204	1.666009	1.666987	1.667234
		2	1.304758	1.597630	1.641846	1.665083	1.670879	1.671871	1.671992
		3	1.303291	1.583810	1.625526	1.638599	1.643394	1.644318	1.644478
		4	1.292584	1.567840	1.611255	1.631291	1.635229	1.636091	1.636249
		5	1.285412	1.556043	1.596599	1.622939	1.629912	1.631139	1.631354
		t	1.296211	1.577344	1.621001	1.643423	1.649085	1.650081	1.650261
the Great Lakes	L_1	1	4.812502	5.201162	5.293073	5.372435	5.377668	5.377948	5.377969
		2	4.700831	5.050330	5.146203	5.212228	5.214729	5.214839	5.214845
		3	4.812297	5.157786	5.265867	5.352570	5.355372	5.355481	5.355486
		4	4.794220	5.15199	5.258153	5.340851	5.344236	5.344419	5.344428
		5	4.850541	5.228137	5.327958	5.417393	5.422284	5.422577	5.422596
		t	4.794078	5.157881	5.258251	5.339095	5.342858	5.343053	5.343065
	L_2	1	4.018321	4.434843	4.607270	4.803114	4.822162	4.823877	4.824044

		2	3.928048	4.311367	4.455691	4.592455	4.603661	4.604724	4.604833
		3	3.964821	4.345896	4.504096	4.673926	4.690270	4.691954	4.692127
		4	4.011973	4.387739	4.528083	4.697807	4.718469	4.721303	4.721667
		5	3.982305	4.382182	4.518567	4.676387	4.694411	4.696611	4.696863
		t	3.981094	4.372405	4.522741	4.688738	4.705795	4.707694	4.707907
	L_3	1	4.082571	4.524822	4.729207	4.962632	4.983205	4.984773	4.984897
		2	4.012006	4.446975	4.648429	4.865829	4.885575	4.887374	4.887536
		3	4.028373	4.424247	4.593104	4.766631	4.783570	4.785271	4.785444
		4	4.096634	4.503434	4.664088	4.823932	4.836680	4.837805	4.837899
		5	4.070577	4.493371	4.664640	4.847425	4.860268	4.861173	4.861233
		t	4.058032	4.478570	4.659894	4.853290	4.869860	4.871279	4.871402
the Gulf of Mexico	G_1	1	2.565556	3.609419	3.996673	4.382391	4.541528	4.609862	4.636473
		2	2.479181	3.546940	3.936535	4.326910	4.496627	4.559353	4.585083
		3	2.590357	3.648914	4.038063	4.455251	4.625787	4.685784	4.707977
		4	2.503287	3.574801	3.979062	4.427319	4.580493	4.654192	4.678612
		5	2.499481	3.558736	3.934834	4.316042	4.487383	4.562402	4.591122
		t	2.527572	3.587762	3.977033	4.381583	4.546364	4.614319	4.639853
	G_2	1	0.743384	0.895098	0.945603	0.981701	0.986168	0.986864	0.986964
		2	0.744712	0.896546	0.942712	0.970992	0.974178	0.974758	0.974892
		3	0.756848	0.911967	0.956991	0.983282	0.988125	0.988807	0.988922
		4	0.739550	0.897418	0.941875	0.967511	0.971251	0.971987	0.972122
		5	0.739951	0.893114	0.941043	0.965990	0.970202	0.970773	0.970856
		t	0.744889	0.898829	0.945645	0.973895	0.977985	0.978638	0.978751
	G_3	1	0.294791	0.341896	0.361937	0.382367	0.386702	0.387546	0.387708
		2	0.288769	0.333492	0.351847	0.370863	0.375577	0.376427	0.376592
		3	0.292654	0.337313	0.353098	0.369937	0.374785	0.375733	0.375957
		4	0.298438	0.345078	0.361943	0.378349	0.382237	0.383226	0.383429
		5	0.295312	0.342420	0.358725	0.376093	0.380561	0.381394	0.381533
		t	0.293993	0.340040	0.357510	0.375522	0.379972	0.380865	0.381044

References

[1] Riahi, K., Rao, S., Krey, V. et al. *RCP 8.5—A scenario of comparatively high greenhouse gas emissions.* Climatic Change (2011) 109: 33. https://doi.org/10.1007/s10584-011-0149-y

[2] R. L. Sriver, C. E. Forest, and K. Keller. *Effects of initial conditions uncertainty on regional climate variability: An analysis using a low-resolution CESM ensemble.* Geophysical Research Letters, 42(13):5468–5476, 2015.

[3] Robert H. Shumway David S. Stoffer (2016). *Time Series Analysis and Its Applications With R Examples Fourth Edition.* New York: Springer.

[4] Palma, Wilfredo (2007). *Long-memory time series: theory and methods.* Hoboken: John Wiley & Sons, Inc.

[5] Jianqing Fan and Qiwei Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer-Verlag.

[6] Hastie, Tibshirani and Friedman (2009). *The Elements of Statistical Learning 2nd edition*. New York: Springer-Verlag.

[7] Wolfgang K. Härdle, Brenda L. Cabrera, Ostap Okhrin & Weining Wang. *Localizing Temperature Risk*. Journal of the American Statistical Association Vol. 111, Iss. 516, 2016.

[8] K. A. McKinnon, A. Rhines, M. P. Tingley & P. Huybers. *Long-lead predictions of eastern United States hot days from Pacific sea surface temperatures*. Nature Geoscience 9, 389–394 (2016).