HW7 Details

Things to submit:

1) report scale of your choice and plot of predicted average annual mean of the minimum temperature on 100 by 100 grid spanning all stations. plot should look like heatmap with legend(example in book)

2) regularize regression from 1) with **lasso** and do the same thing. report scales(at least six scales), plots, regularization constants, and briefly address "How many predictors does your model use? How does prediction error change with the number of predictors?"

3) regularize regression now with **elastic net**. try at least 3 different alpha values and briefly explain how this value affect the model (eg: does alpha affect predicted value? In general, any notable observations will suffice)

4) all your **source code**

5) a **README** file that says how to execute your code and which piece of code is associated with which question

6) a **group** plain text file. Of the form ***[full name], [NetID]*** per line (CSV style).

Note:

1) each group member should submit at least **group** file. The first person (line) in your **group** file should submit a **HW7.zip** file that contains all items from 1 - 6. You may work in a group up to 3 people.

Assignment 7 Report
Harrison Kiang hkiang2, Umberto Ravaioli urjav2, Annlin Sheih sheih2

Q1.

1) Use a kernel method (section 9.1.1) to smooth this data, and predict the average annual temperature at each point on a 100x100 grid spanning the weather stations. You should use each data point as a base point, and you should search over a range of at least six scales, using cross-validation to choose the scale. Use a Gaussian kernel. Plot your prediction as an image. Compare to the Kriging result that you can find in figure 4 here .

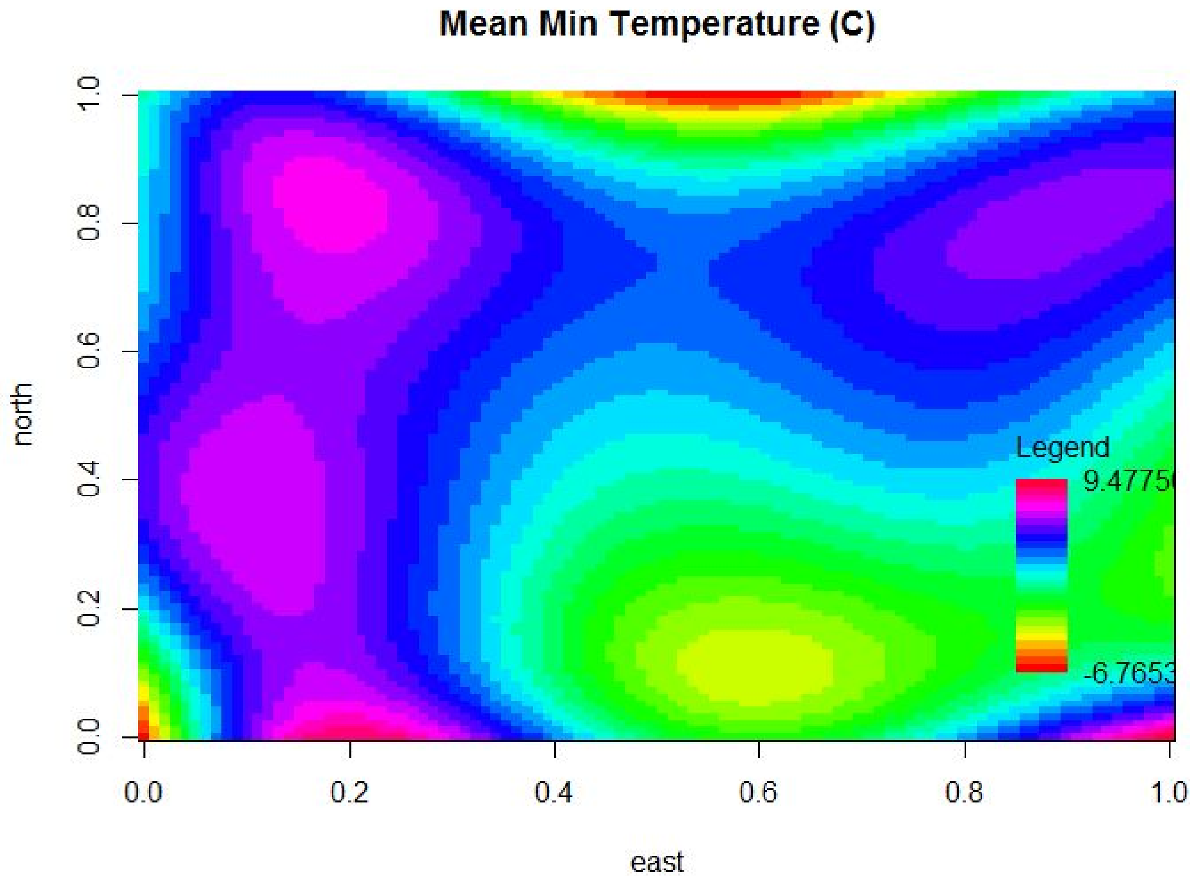
2) Regularize this kernel method (section 9.1.1; worked example 9.1; associated code) using the lasso, and predict the average annual temperature at each point on a 100x100 grid spanning the weather stations. You should choose the regularization constant using cross validation (cv.glmnet will do the work for you; read the manual!). You should use each data point as a base point, and you use a range of at least six scales. Plot your prediction as an image. Compare to the Kriging result that you can find in figure 4 here . How many predictors does your model use? How does prediction error change with the number of predictors?

   Scales: 10000, 20000, 50000, 75000, 100000, 150000, 200000, 300000, 500000000

3) Now investigate the effect of different choices of elastic net constant (alpha)

Actual Answers:

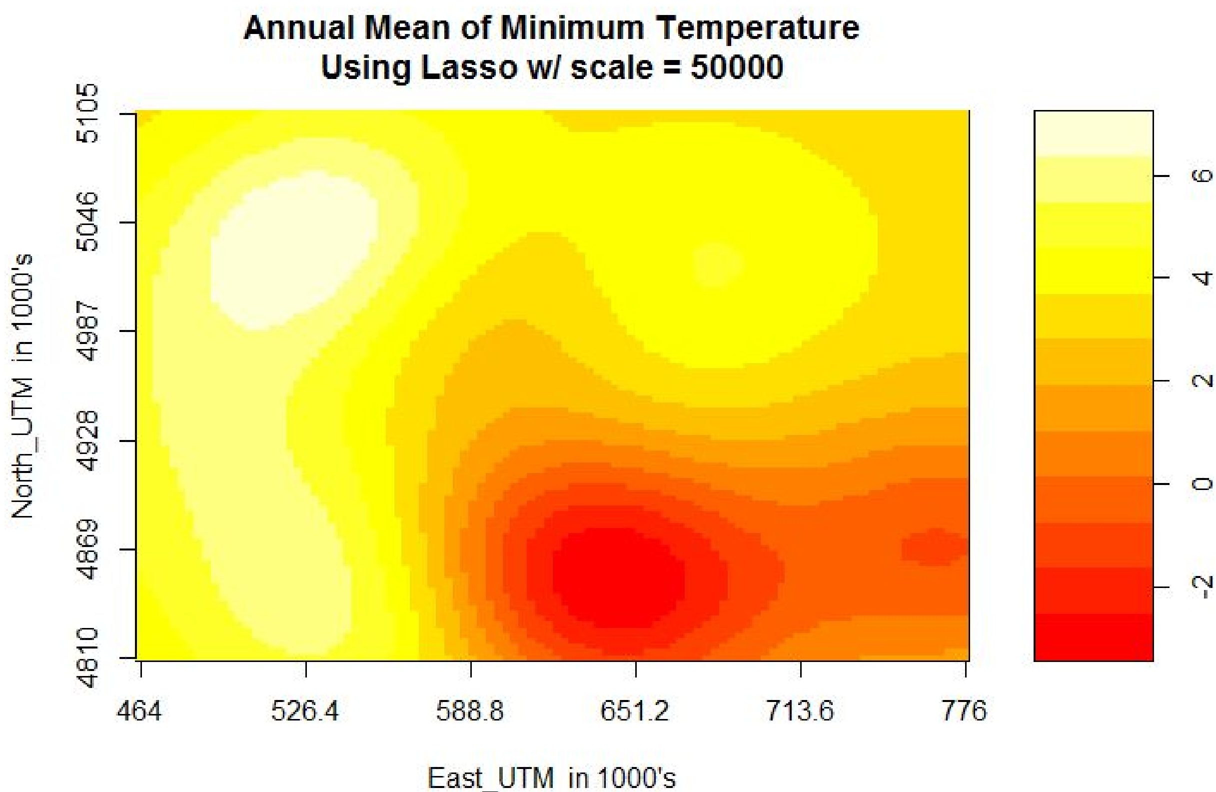
1)

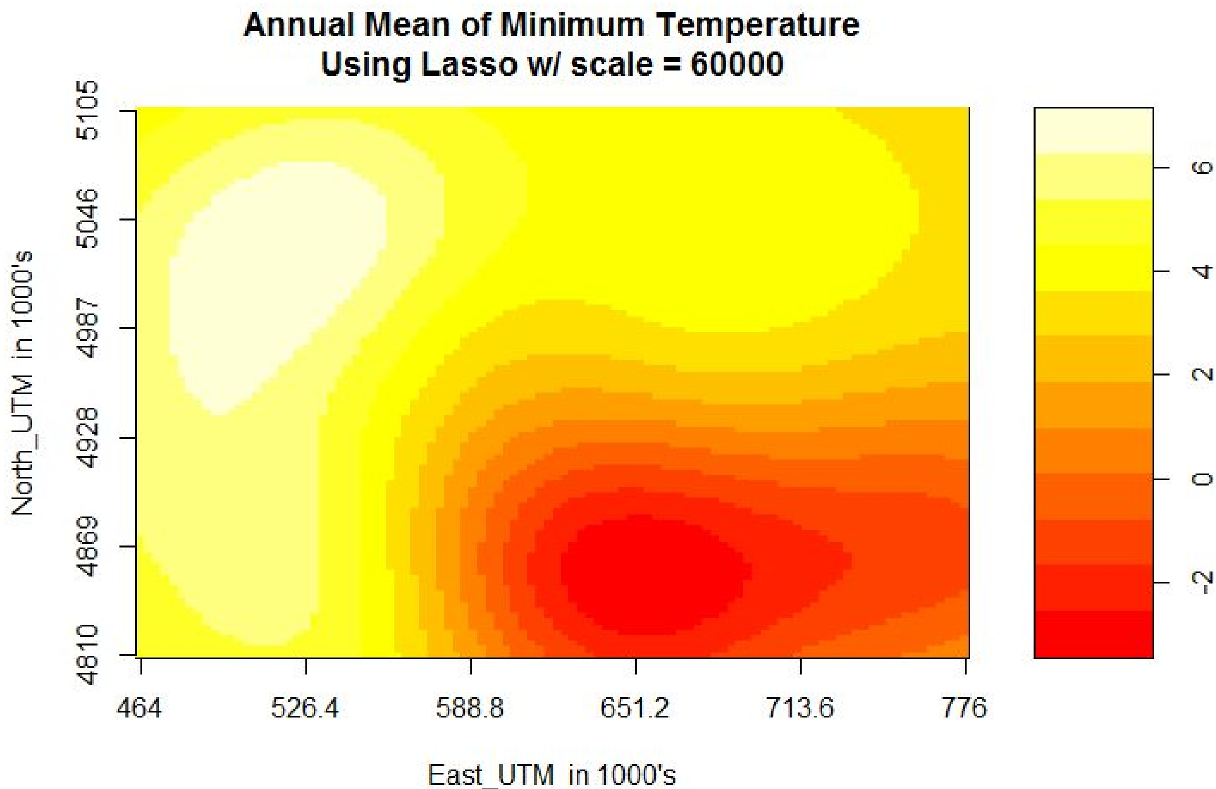


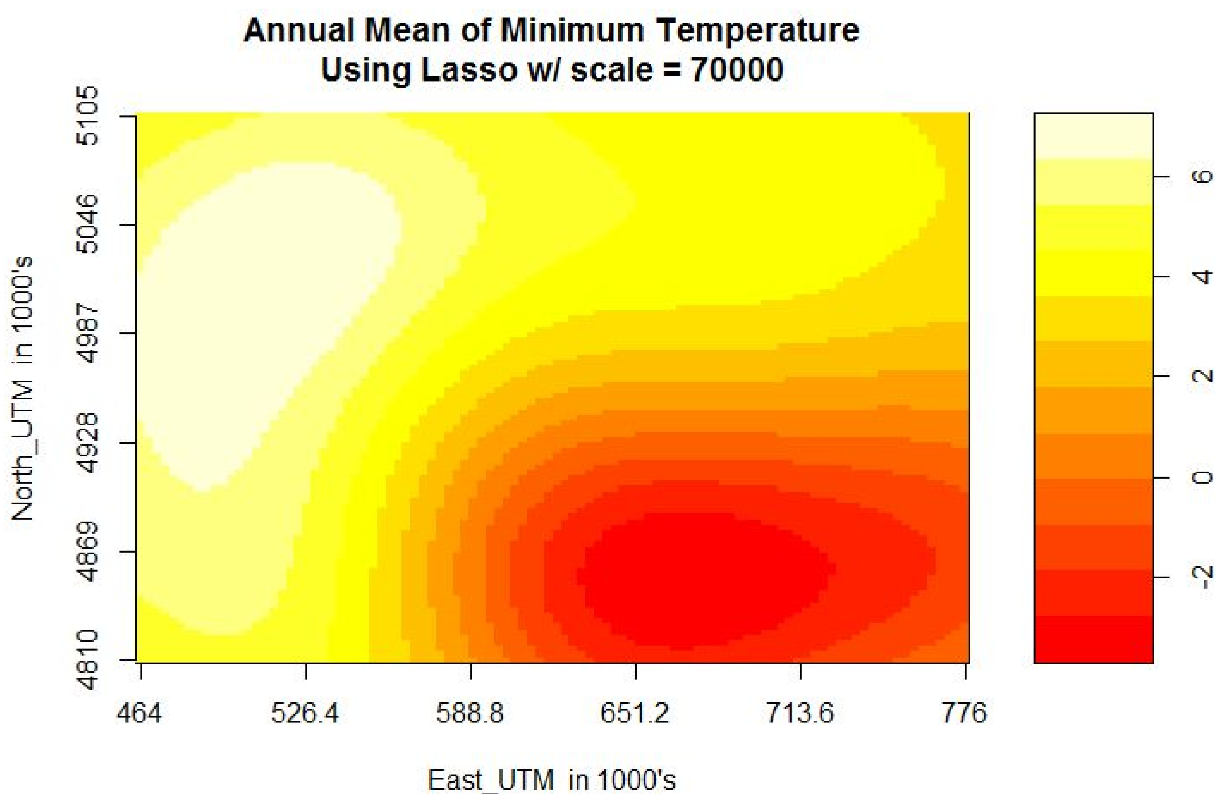
The best scale (not squared) of 68000000000 was chosen for this regression to produce the image above. Comparing to the given figure 4 the structure of our regression very closely matches in the region focused by the figure. However, there are differences with our model not reaching as significant highs or lows as the figure. We believe that is because the given figure focused only on regressing over a very small slice of the map whereas our regression needed to compensate for the entire weather station region.

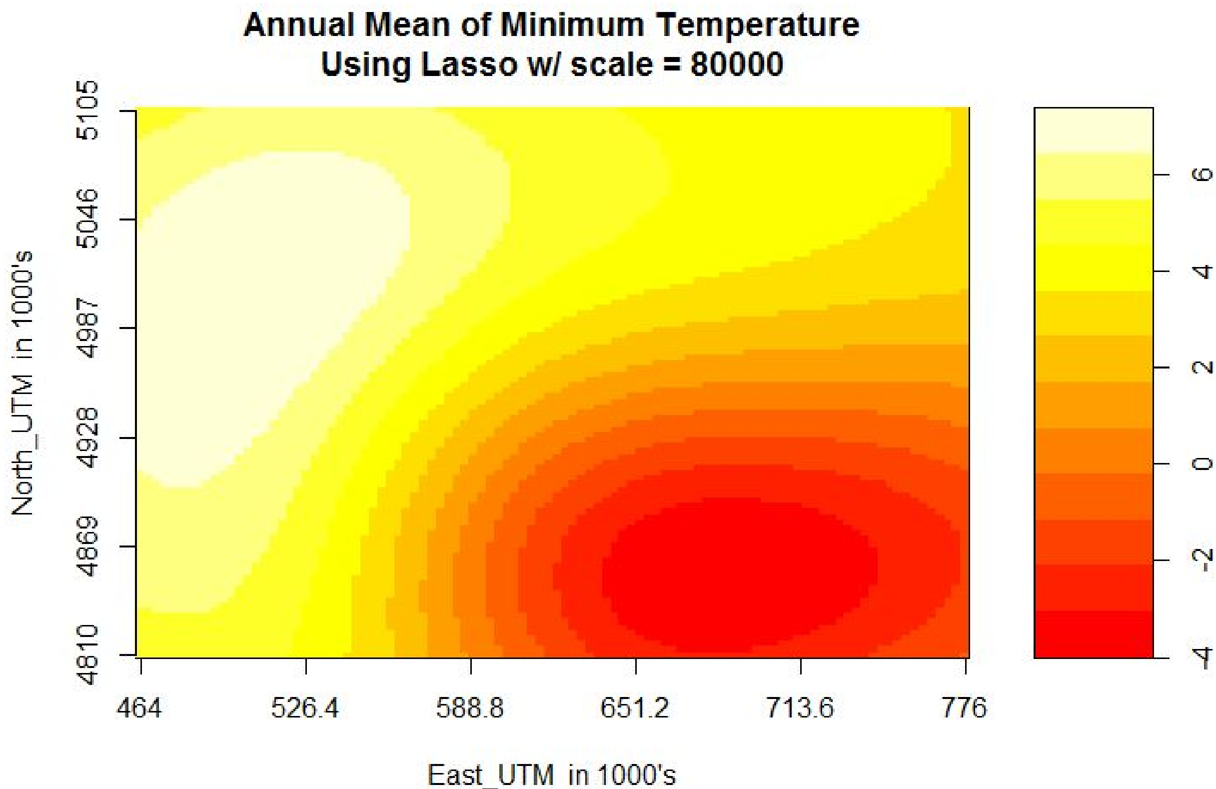
Selecting a scale was extremely difficult. Cross validation allowed us to throw out very poor scale choices, but unfortunately cross validation on simply the weather station data failed to come up with a good scale factor. Once applied to the entire grid of 100x100 points, cross validated scale points could result in poor final images. This is largely because many of the smoothed over points were very far away from the relatively clustered weather stations so a scale that worked well for distant points would not work well for the other nearby weather station. The clear conclusion that can be drawn from this is that one single best scale can't be chosen, instead one should use multiple scales with regularization so that each weather station can effectively impact nearby points and distant ones.

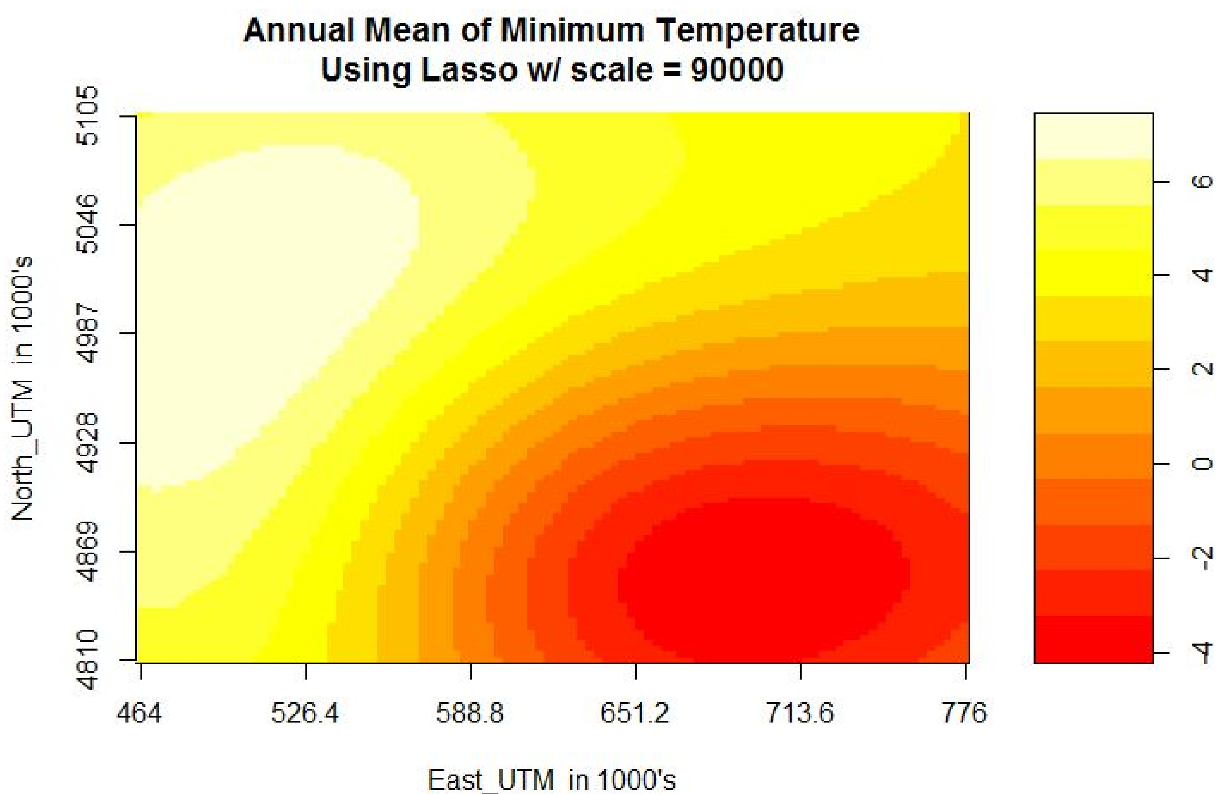
2)



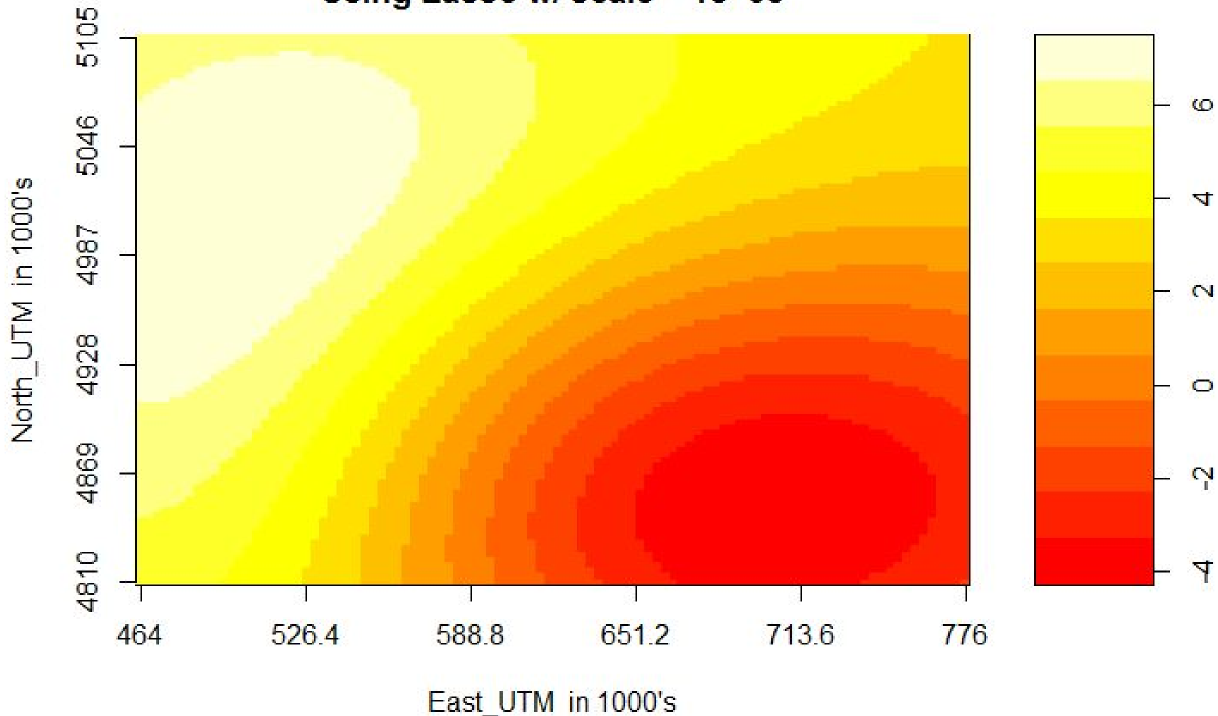

Annual Mean of Minimum Temperature
Using Lasso w/ scale = 60000
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2


Annual Mean of Minimum Temperature
Using Lasso w/ scale = 70000
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2

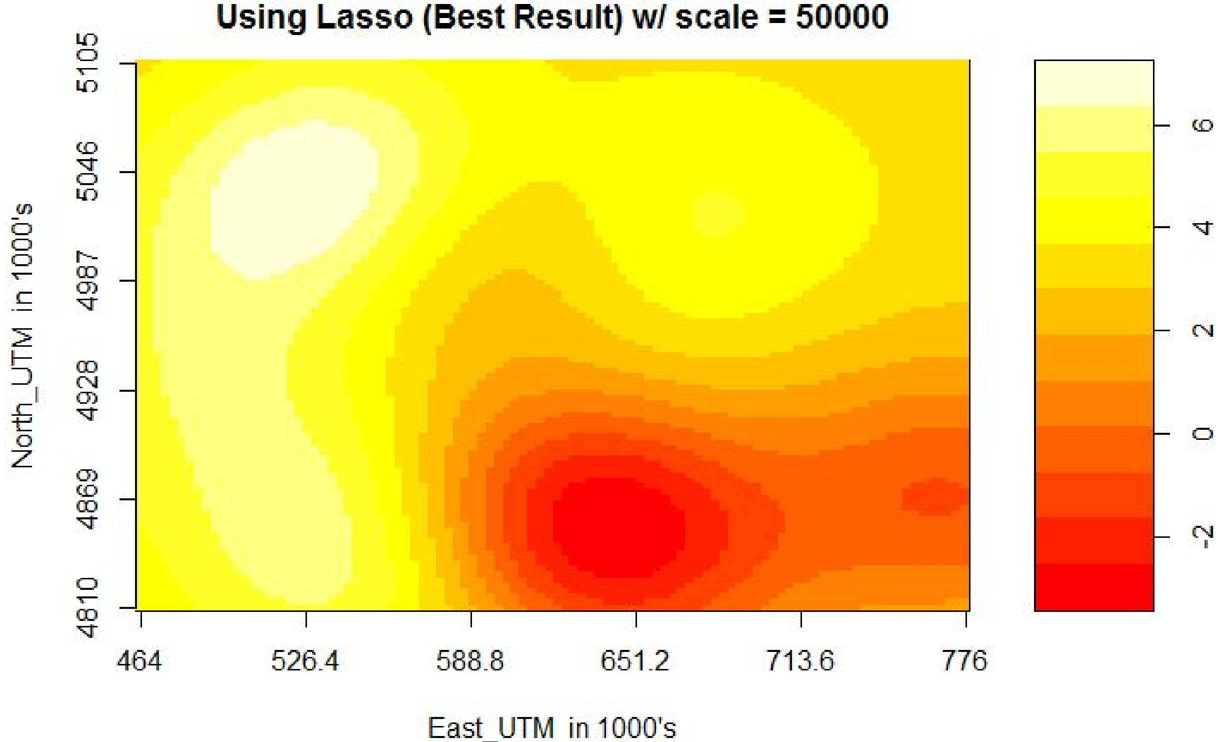

Annual Mean of Minimum Temperature
Using Lasso w/ scale = 80000
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2
-4


Annual Mean of Minimum Temperature
Using Lasso w/ scale = 90000
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2
-4

Annual Mean of Minimum Temperature
Using Lasso w/ scale = 1e+05



Annual Mean of Minimum Temperature
Using Lasso (Best Result) w/ scale = 50000

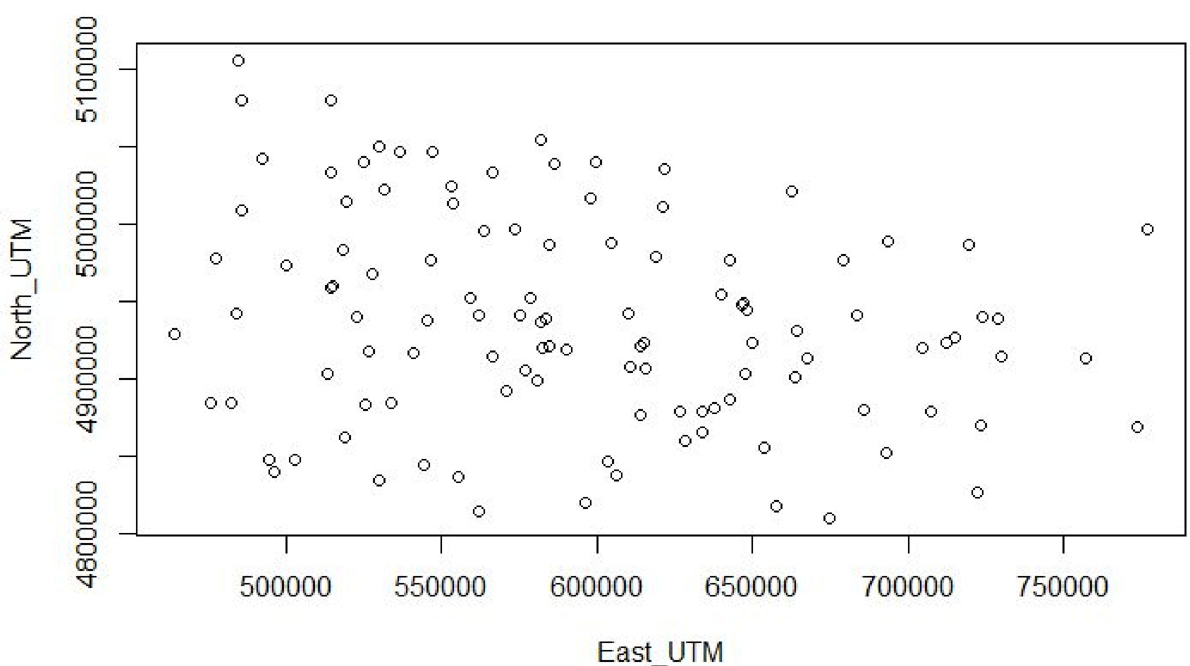


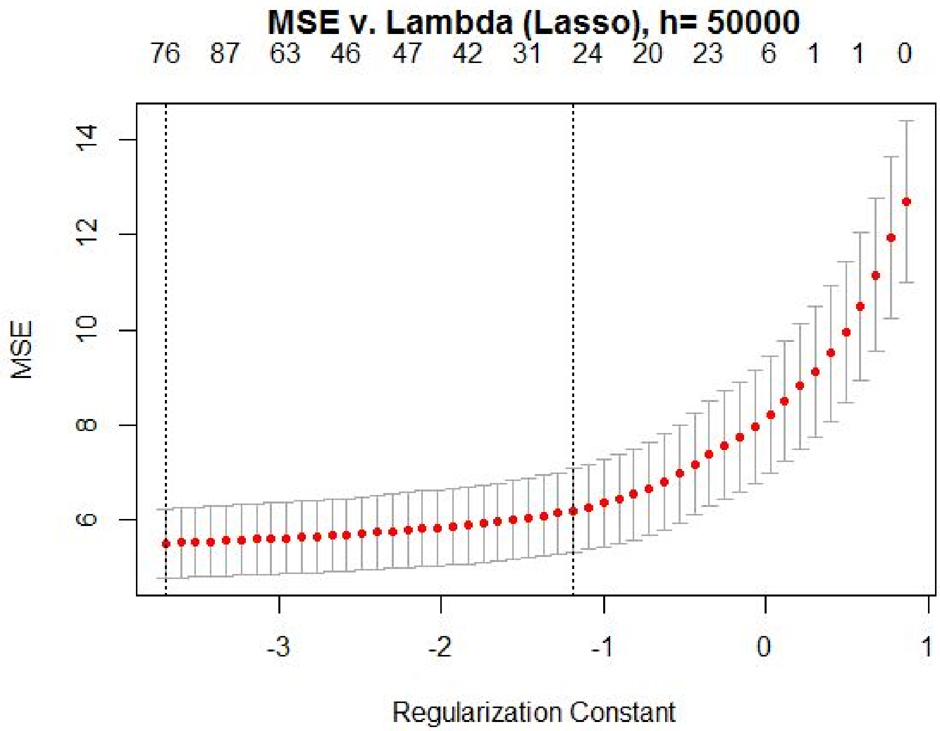
Above heatmap for scale of 50000 and regularization constant of 0.02479147

Scales tested: 50000 to 100000 in increments of 2500 (denoted as srange in source code)
Best scale: 50000 (denoted as bestScale in source code)
Best Corresponding regularization constant: 0.02479147 (bestLambda_lasso in source code)

Trend observed: the higher the scale, the lower the annual min temp for all points in the 100x100 array becomes, as the output of the Gaussian kernel varies inversely with the scale.

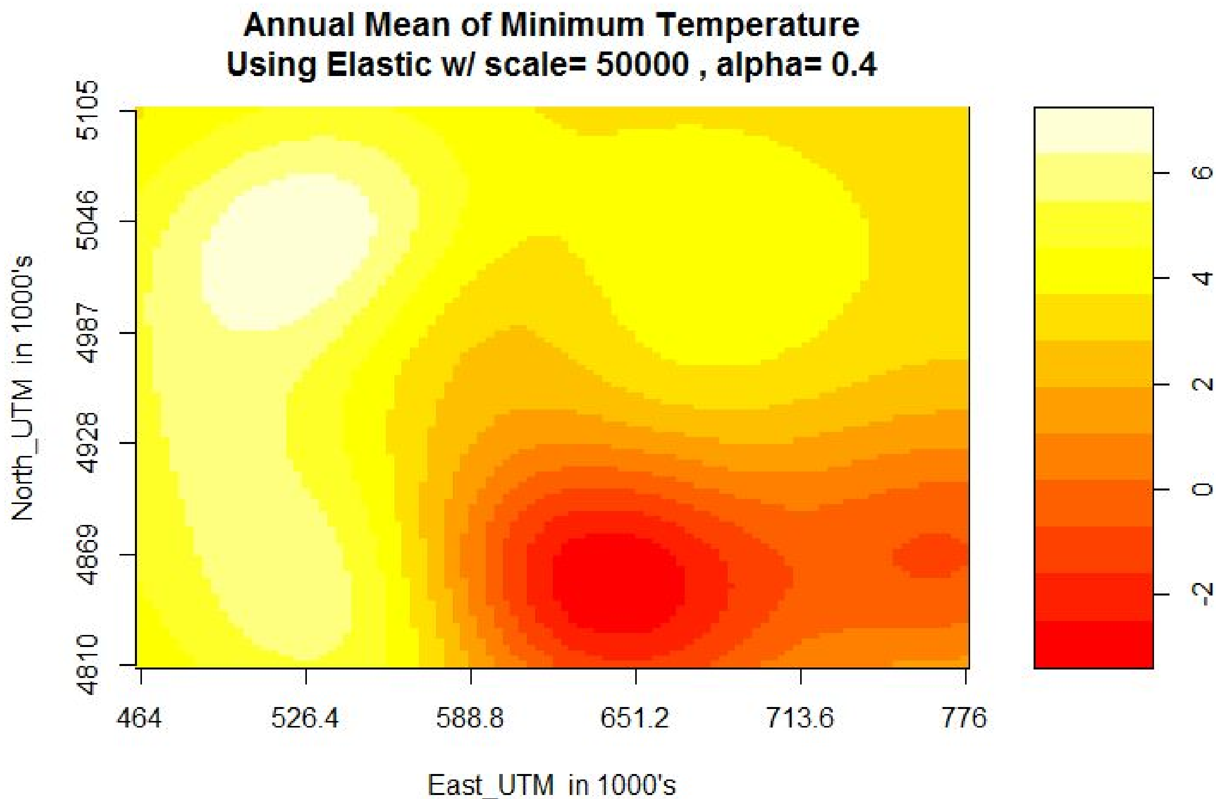
Compared to Figure 4 provided by the University of Edinburgh (http://www.geos.ed.ac.uk/homes/s0198247/Geostats.html), there are similarities with exception of the upper right corner, as there are not many data points or weather stations in that region. Below are the plots of the locations; the higher the concentration of locations, the better correlation between Figure 4 and the heatmap generated by our lasso prediction.

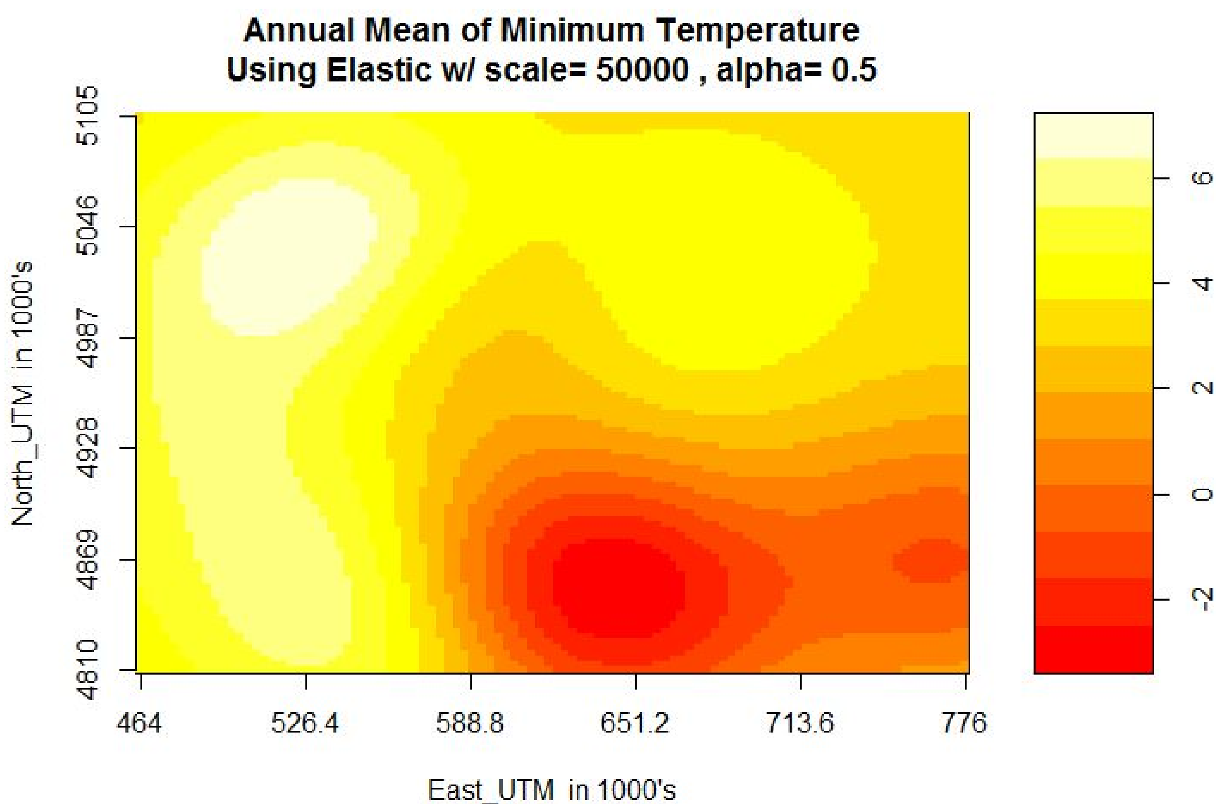


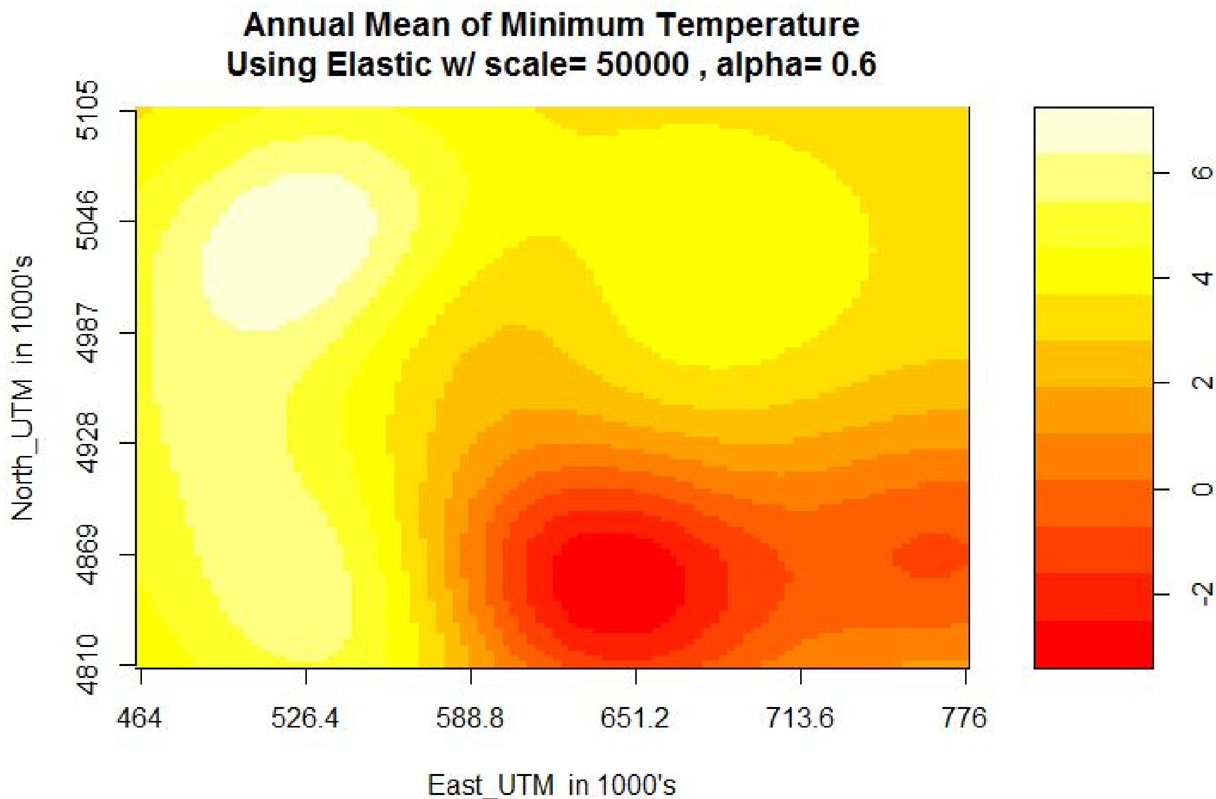


Our model used 76 predictors. Mean square error decreases from 0 to 24 predictors, where the regularization constant results in the residual hitting a knee, where using more predictors yields diminishing returns.

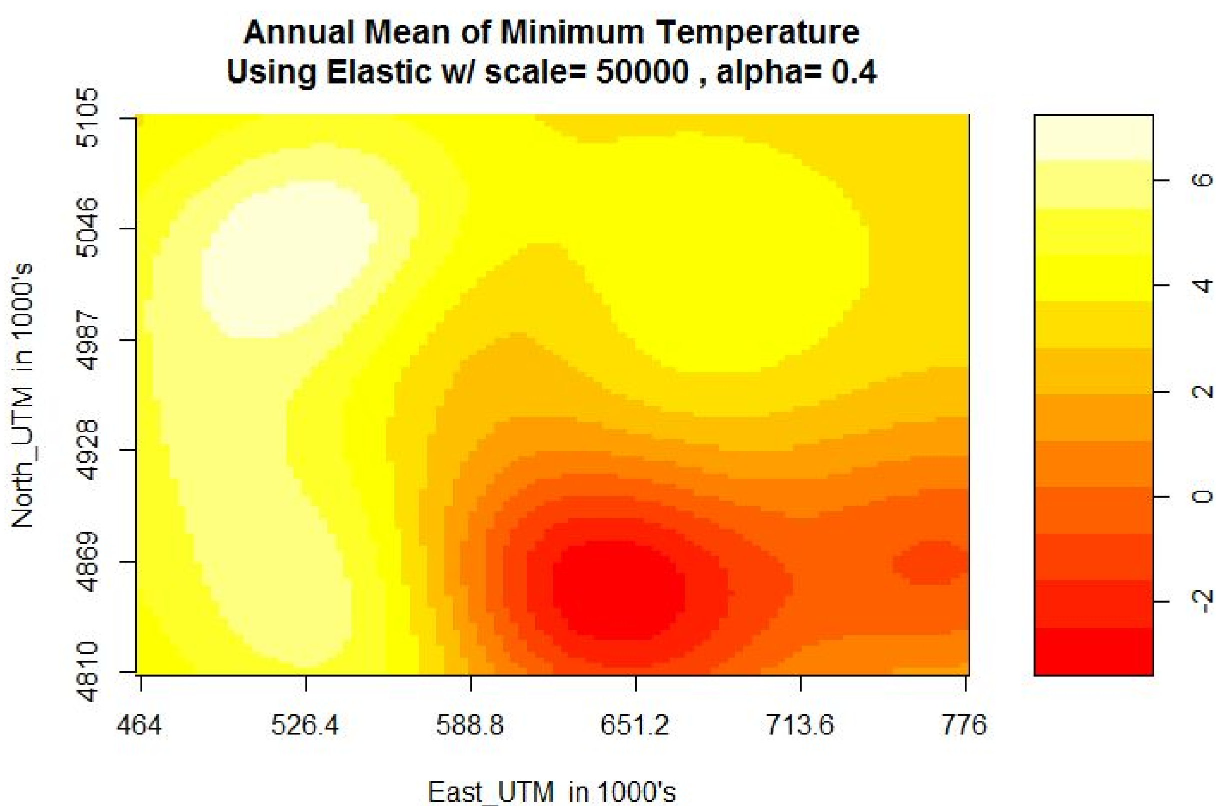
3)


Annual Mean of Minimum Temperature
Using Elastic w/ scale= 50000 , alpha= 0.4
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2


Annual Mean of Minimum Temperature
Using Elastic w/ scale= 50000 , alpha= 0.5
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2



Annual Mean of Minimum Temperature
Using Elastic w/ scale= 50000 , alpha= 0.6
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2



Annual Mean of Minimum Temperature
Using Elastic w/ scale= 50000 , alpha= 0.4
North_UTM in 1000's
5105
5046
4987
4928
4869
4810
464
526.4
588.8
651.2
713.6
776
East_UTM in 1000's
6
4
2
0
-2

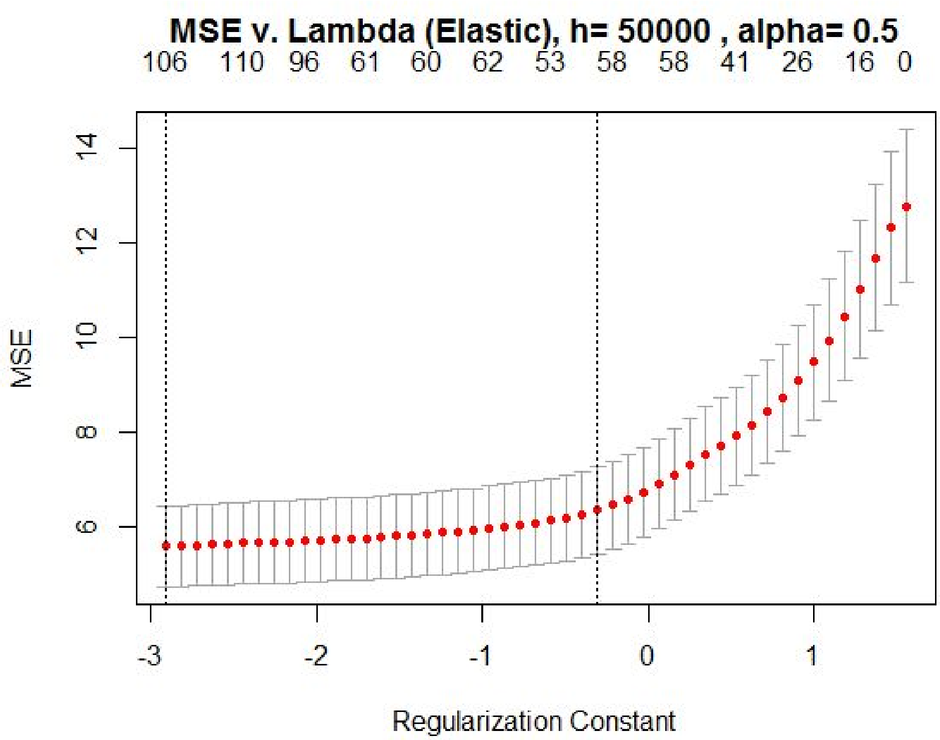
Above heatmap for scale of 50000 and regularization constant of 0.05441722, alpha = 0.4

Scales tested: 50000 to 100000 in increments of 2500 (denoted as srange in source code)
Best scale: 50000 (denoted as bestScale_elastic in source code)
Best Corresponding regularization constant: 0.05441722 (bestLambda_elastic in source code)

Alphas tested: 0.3, 0.4, 0.5 (denoted as net_alphas in source code)
Best alpha: 0.4

Trend observed: the alphas have little to no effect on the heatmap.

Compared to Figure 4 provided by the University of Edinburgh (http://www.geos.ed.ac.uk/homes/s0198247/Geostats.html), there are similarities with exception of the upper right corner, as there are not many data points or weather stations in that region. Below are the plots of the locations; the higher the concentration of locations, the better correlation between Figure 4 and the heatmap generated by our lasso prediction (see above corresponding section for lasso).



Our model used 106 predictors. Mean square error decreases from 0 to 58 predictors, where the regularization constant results in the residual hitting a knee, where using more predictors yields diminishing returns.