Linear and Nonlinear Postprocessing of Ensemble Forecasts

Ranran Wang[1] and Caren Marzban[1,2]

[1]Department of Statistics

[2]Applied Physics Laboratory

Univ. of Washington, Seattle, Washington

Generalities

NWP forecasts have problems.

Correction of forecasts = post-processing = MOS.

Long history:

- Glahn, Lowry 1972
- Wilson, Vallee 2002
- Marzban 2003
- Pasini et al.
- Marzban, Sandgathe, Kalnay 2005
- Marzban, Leyton, Colman 2007
- Gneiting, Raftery 2005
- Berrocal, Raftery, Gneiting 2007
- Wilson, Beauregard, Raftery, Verret 2007
- Sloughter, Raftery, Gneiting 2007
- Gneiting, Kleiber, Schlather 2009
- Bao, Gneiting, Raftery, Grimit, Guttorp 2009

Method

Here, we postprocess **ensemble** forecasts.

At a given station, at a given time, one has

- M forecasts (from an M-member ensemble),

- of meteorological variables (temp, wind, etc), and

- surface observations.

Goal: Develop a map from forecasts to obs.

Statistical Models:

Regression: temperature and wind (U and V)

Classification: yes/no prcp.

Raw: ensemble mean and ensemble proportion.

Linear:multiple regression (LM) and linear discriminant analysis (LDA)

Nonlinear:Neural Network (NN), minimizing MSE or cross-entropy.

Method ...

Reduce dimensionality of input data.

2-level hierarchical approach:

Step 1:

$$\begin{aligned} T_{fit} &= regress(T_{obs} \sim T_1 + T_2 + \cdots + T_{10}) \\ U_{fit} &= regress(U_{obs} \sim U_1 + U_2 + \cdots + U_{10}) \\ V_{fit} &= regress(V_{obs} \sim V_1 + V_2 + \cdots + V_{10}) \end{aligned}$$

Step 2:

LM1	LM2
$T_{obs} \sim T_1 + T_2 + \cdots + T_{10},$	$T_{obs} \sim T_1 + T_2 + \cdots + T_{10} + U_{fit} + V_{fit},$
$U_{obs} \sim U_{fit} + V_{fit},$	$U_{obs} \sim U_{fit} + V_{fit} + T_{fit},$
$V_{obs} \sim U_{fit} + V_{fit}.$	$V_{obs} \sim U_{fit} + V_{fit} + T_{fit}.$
LDA1	LDA2
$P_{obs} \sim P_1 + P_2 + \cdots + P_{10}$	$P_{obs} \sim P_1 + P_2 + \cdots + P_{10} + T_{obs} + U_{obs} + V$

Similarly, for NNs.

Input data = residuals (to filter out seasonality).

residuals = observed - conditional median.

Performance

Measures:

Mean Squared Error (MSE) for regression,

Area under ROC curve (AUC) for classification.

Assessment:

Cross-validation (on years) $\rightarrow$ sampling distribution of training and validation performance measures.

Compare ensemble mean, LM, and NN.

Compare ensemble proportion, LDA, and NN.

In cross-validation, models are paired.

Must look at pair-wise difference of models.

E.g. Where is zero on the distribution of

MSE_{vld}(LM)- MSE_{vld}(NN) ?

Vary H = number of hidden nodes.

Data (from Fanyou Kong and Steve Leyton)

Regional fine-scale ensemble based on WRF-ARW Version 3.0.1

Ten members, each with unique initial perturbations and varying physics options and land-use tables.

Initial perturbations = bred vectors produced by 6-hourly breeding cycle, using the North America Regional Reanalysis (NARR) as background fields.

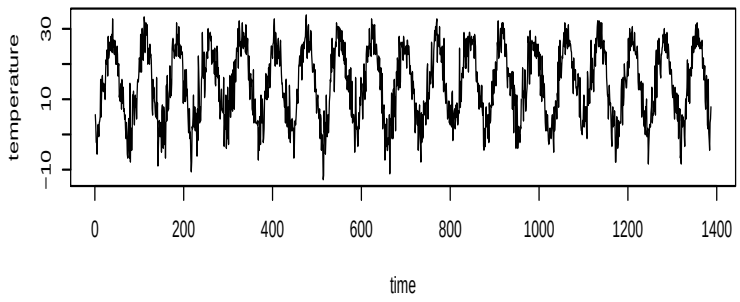
Three-domain, two-way nesting, innermost domain = 15 km grid spacing.

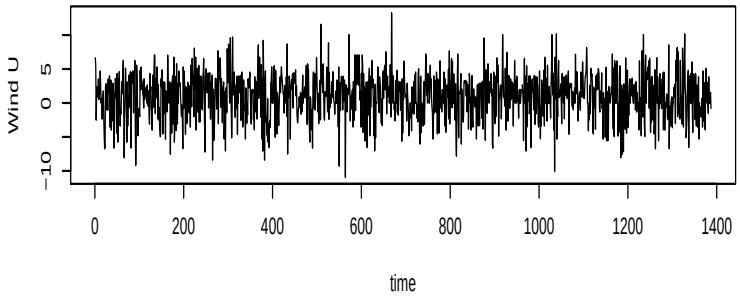
19 years (1987-2005), 48-hr re-forecasts of temperature, wind, and prcp, at 90 stations across the continental US.

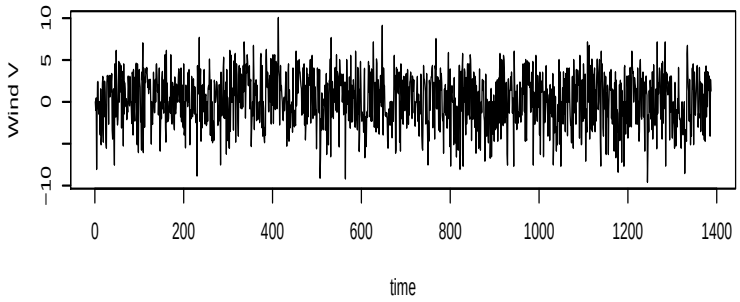
5-day intervals, leading to 1387 cases.

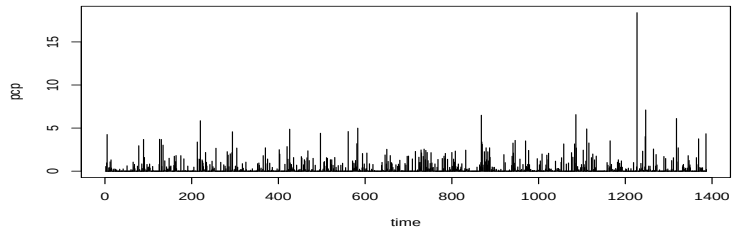
Plus surface obs (analysis, for prcp) at each station.

Surface obs at KSEA

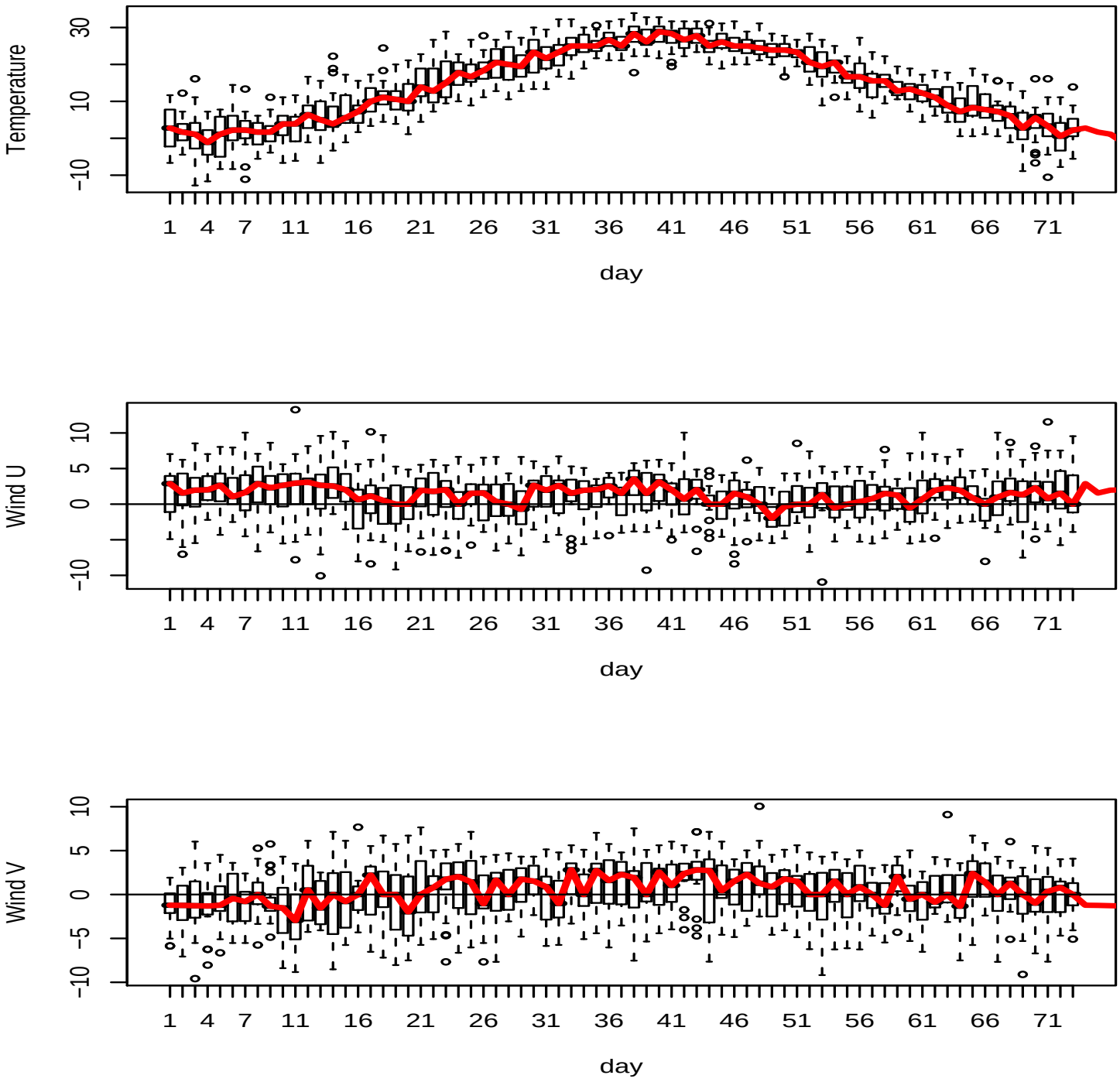




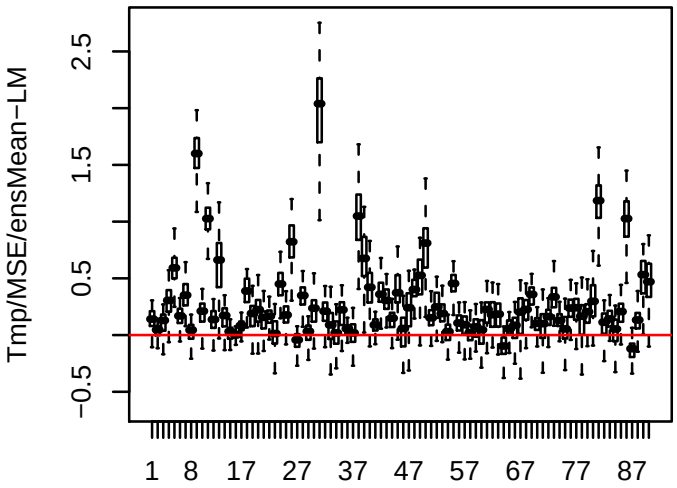


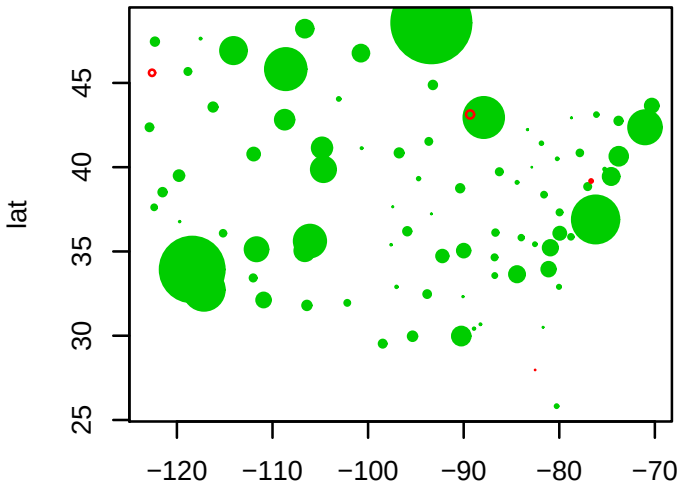


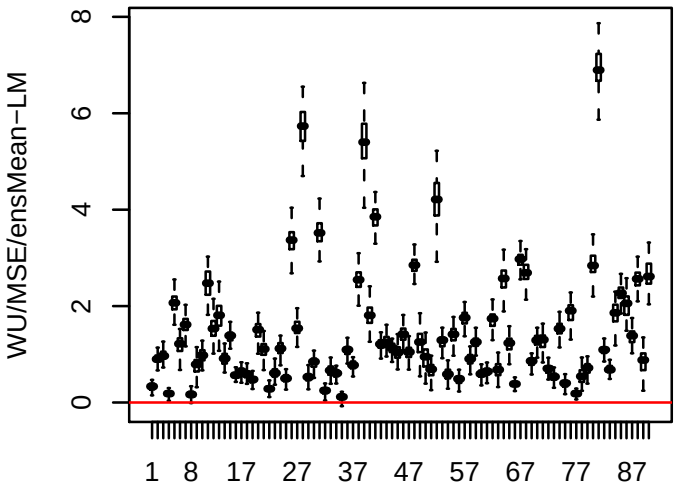
Conditional Medians

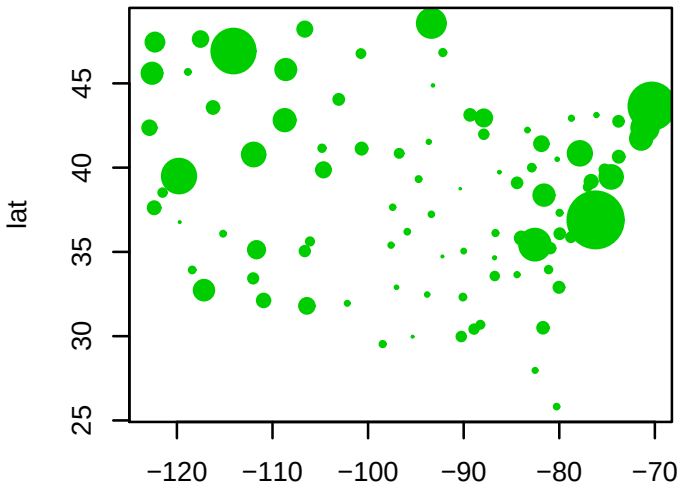


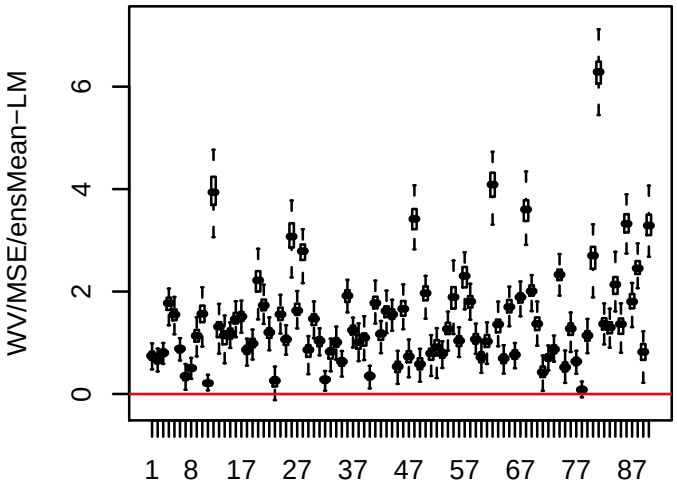
Ensemble Mean - LM

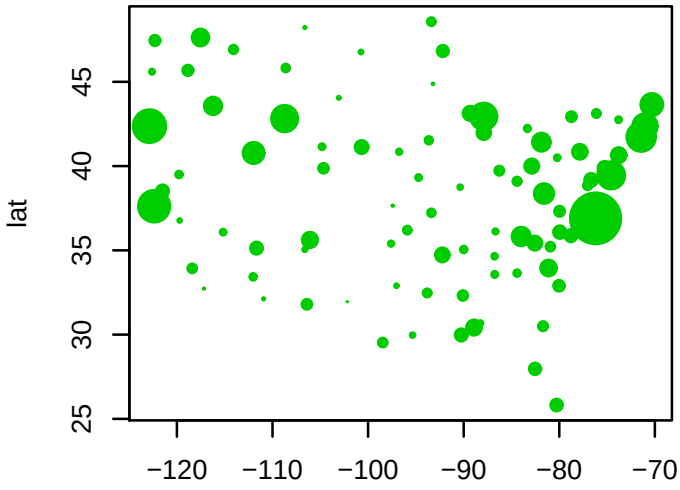




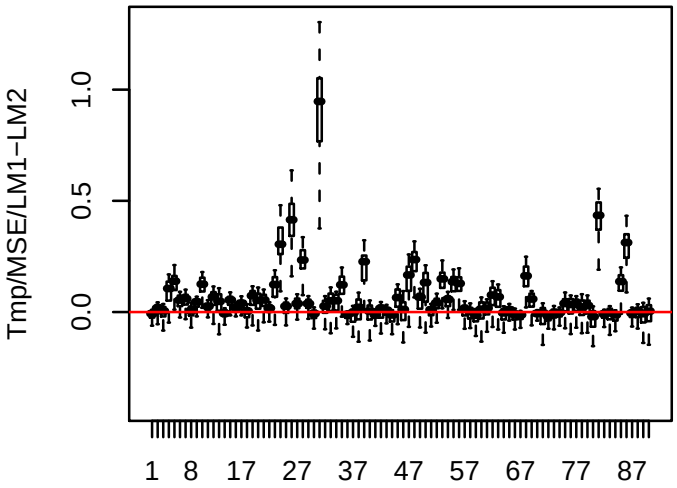


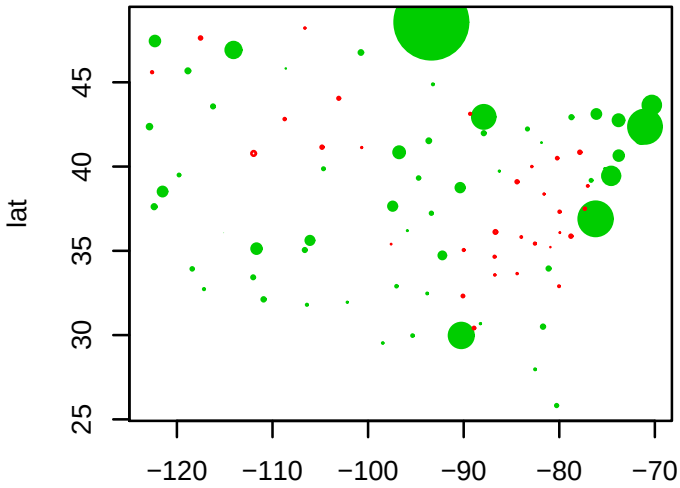


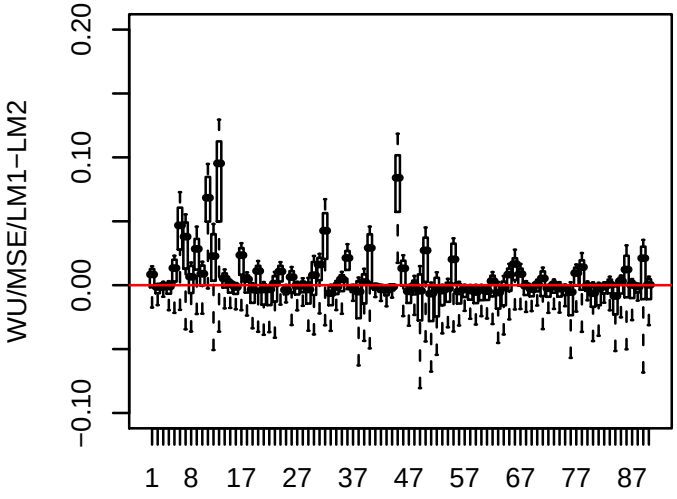


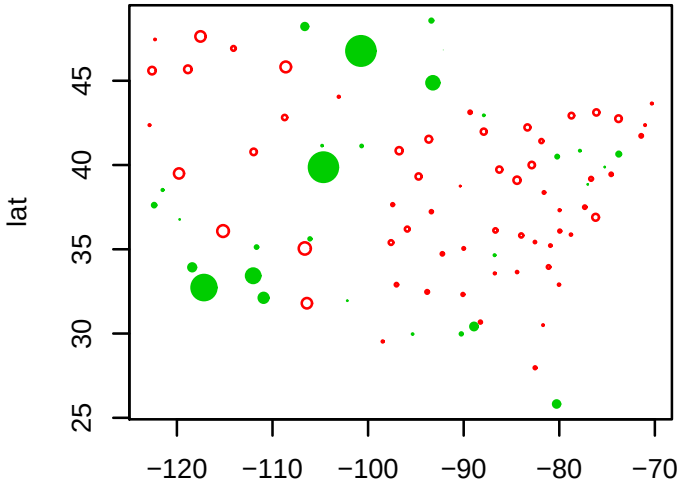


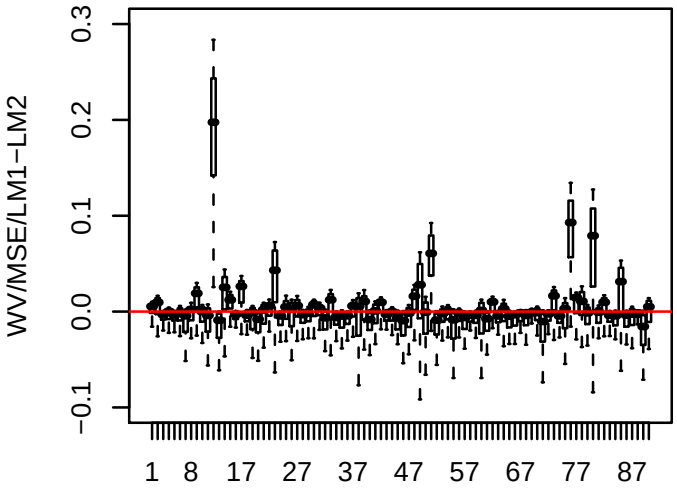
LM1 - LM2

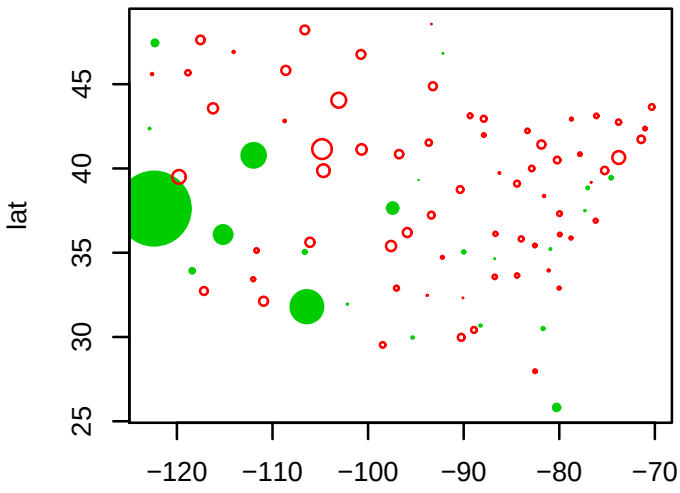

Tmp/MSE/LM1−LM2
1.0
0.5
0.0
1 8 17 27 37 47 57 67 77 87


lat
45
40
35
30
25
−120 −110 −100 −90 −80 −70

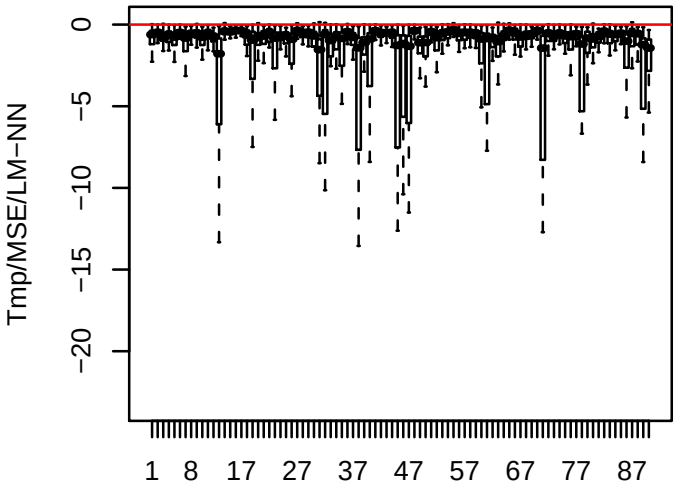

WU/MSE/LM1−LM2
0.20
0.10
0.00
−0.10
1 8 17 27 37 47 57 67 77 87

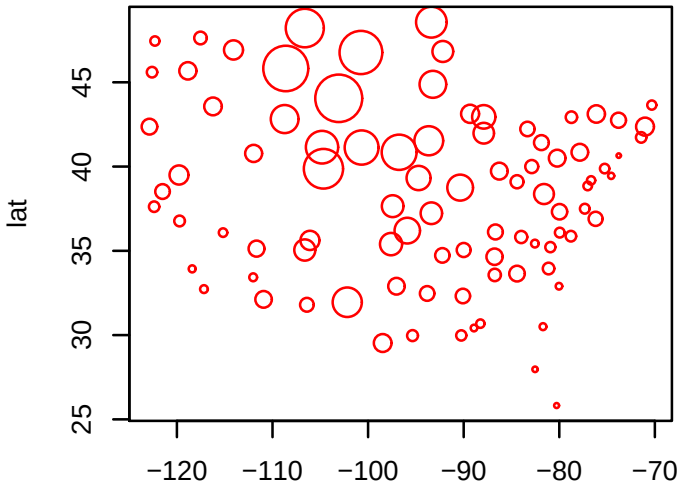

lat
45
40
35
30
25
−120 −110 −100 −90 −80 −70

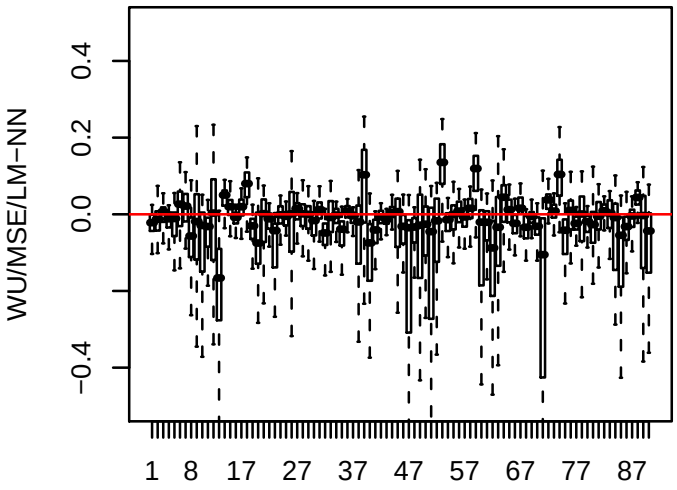

WV/MSE/LM1−LM2
0.3
0.2
0.1
0.0
−0.1
1 8 17 27 37 47 57 67 77 87

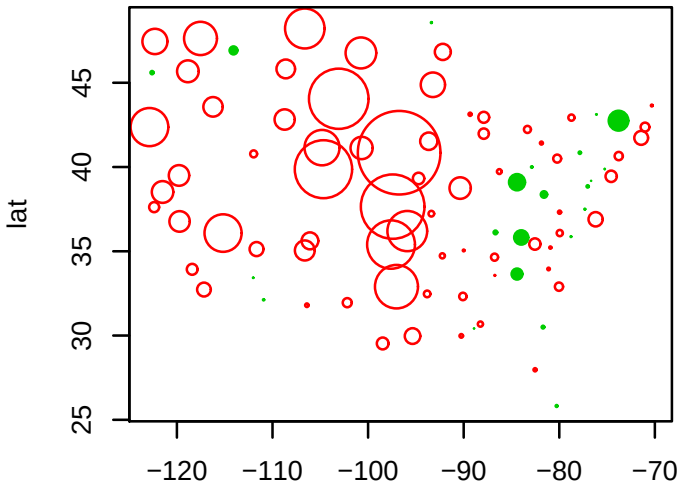

lat
45
40
35
30
25
−120 −110 −100 −90 −80 −70

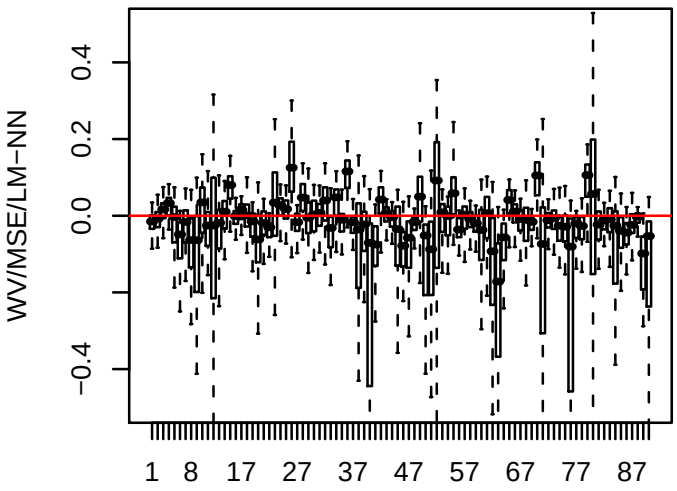
LM - NN

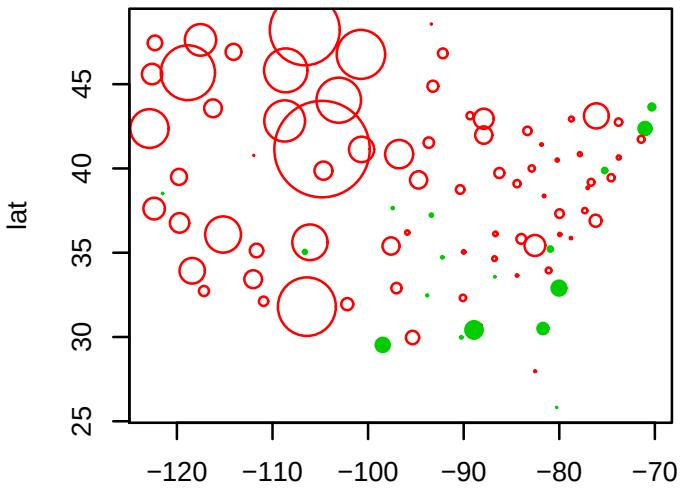

Tmp/MSE/LM−NN
0
−5
−10
−15
−20
1
8
17
27
37
47
57
67
77
87


lat
45
40
35
30
25
−120
−110
−100
−90
−80
−70

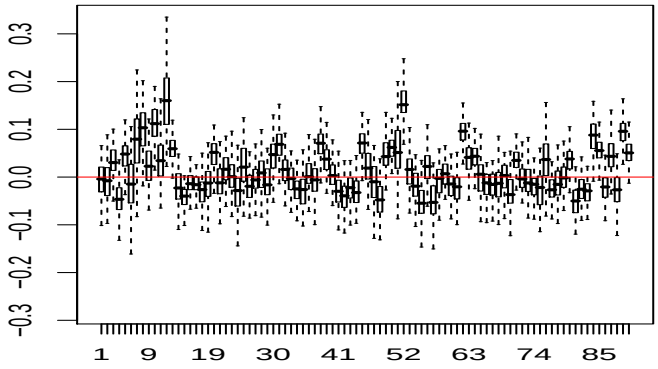

WU/MSE/LM−NN
0.4
0.2
0.0
−0.4
1
8
17
27
37
47
57
67
77
87

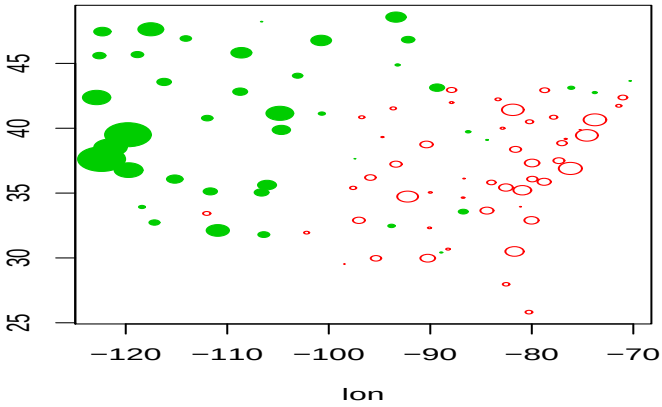

lat
45
40
35
30
25
−120
−110
−100
−90
−80
−70


WV/MSE/LM−NN
0.4
0.2
0.0
−0.4
1
8
17
27
37
47
57
67
77
87

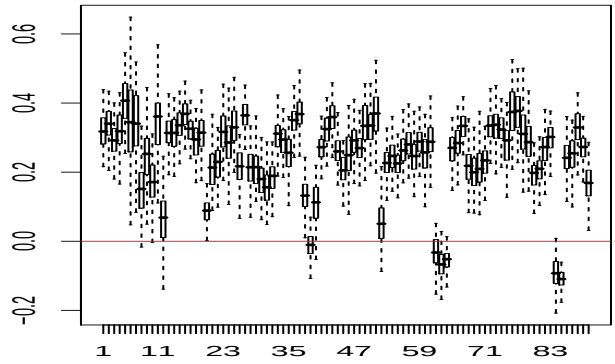

lat
45
40
35
30
25
−120
−110
−100
−90
−80
−70

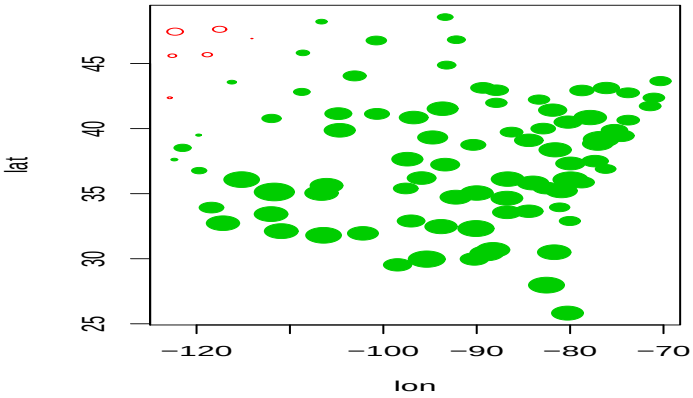
LDA1 - LDA2



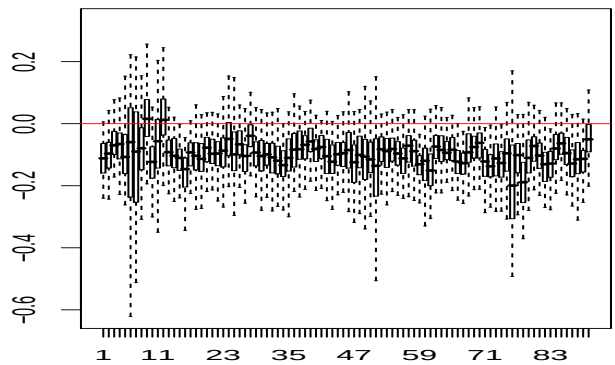


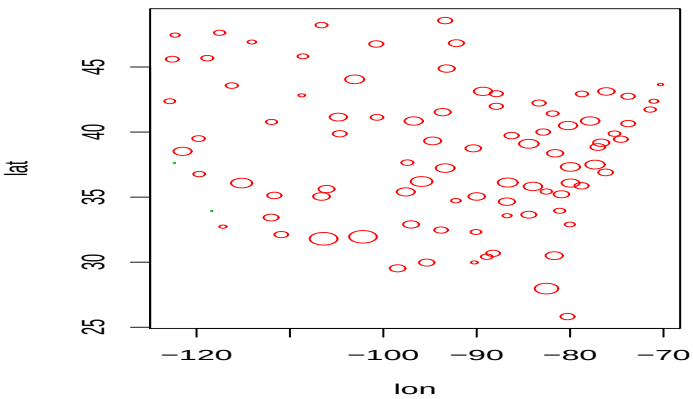
LDA - Ensemble Proportion





NN - LDA

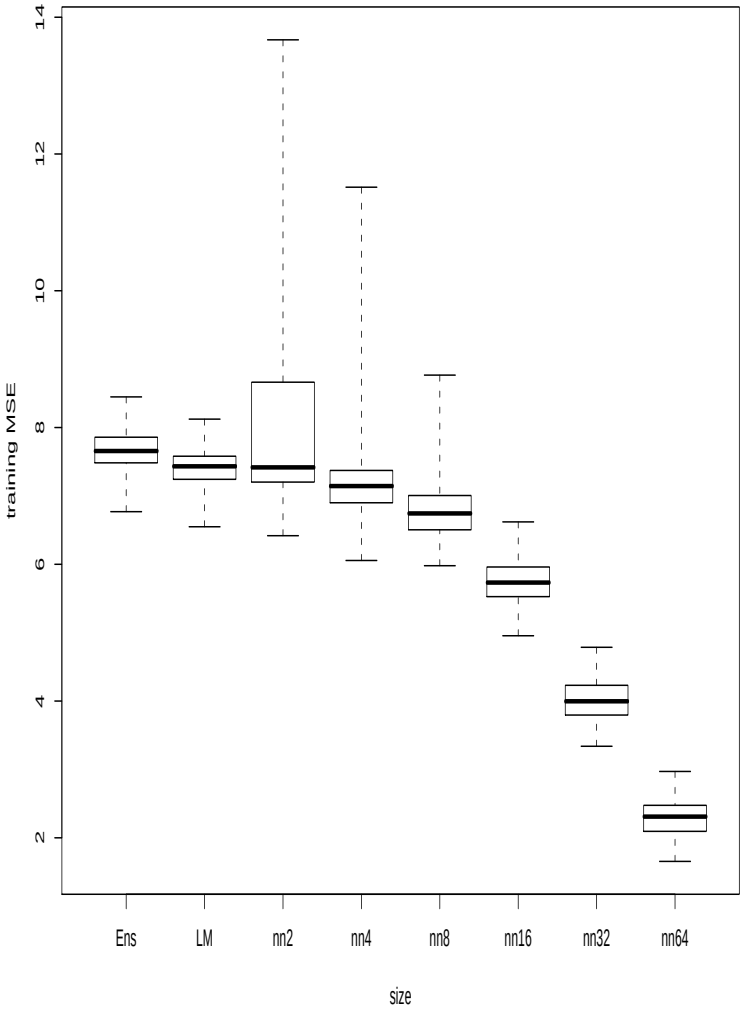




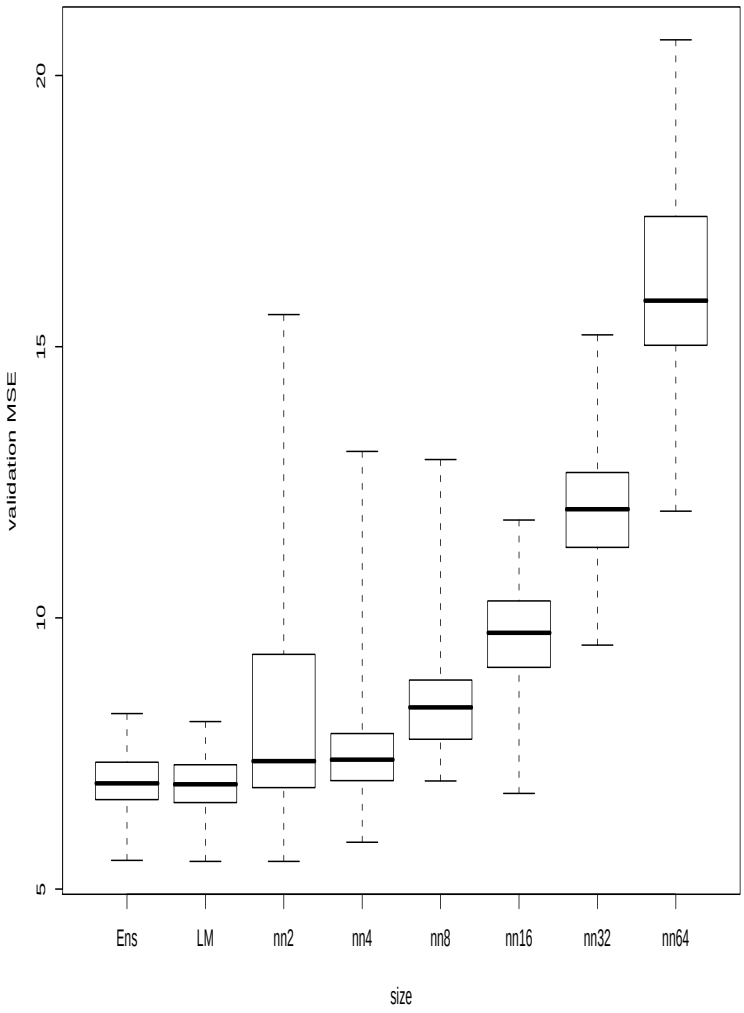
NNs Overfit!

KOKC MSE Temp vs. ensemble mean, LM, and NN with H = 2, 4, 8, 16, 32, 64 .

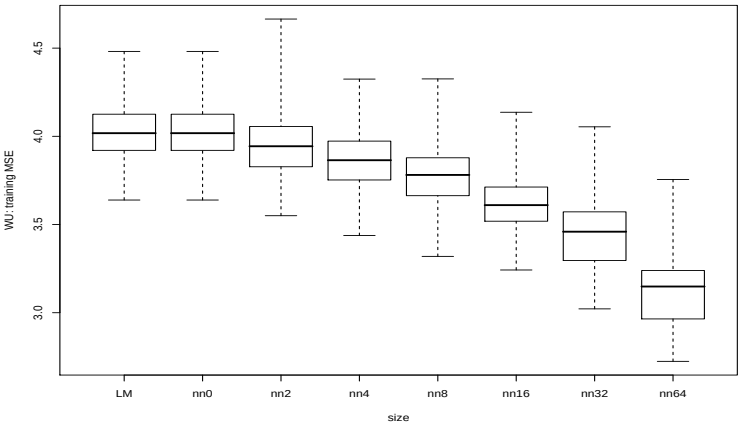
Training

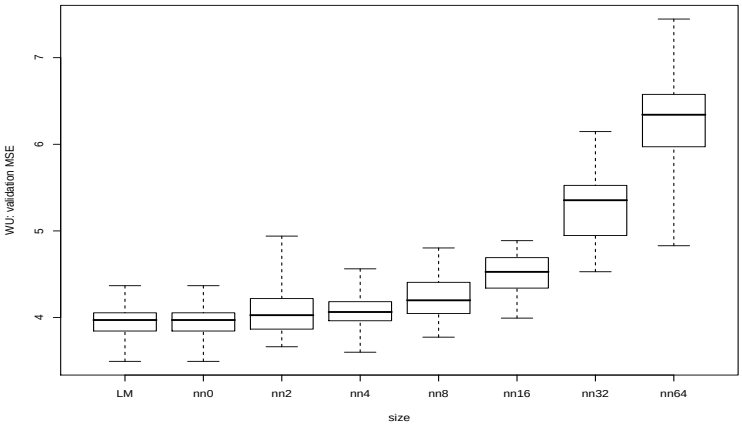


Validation

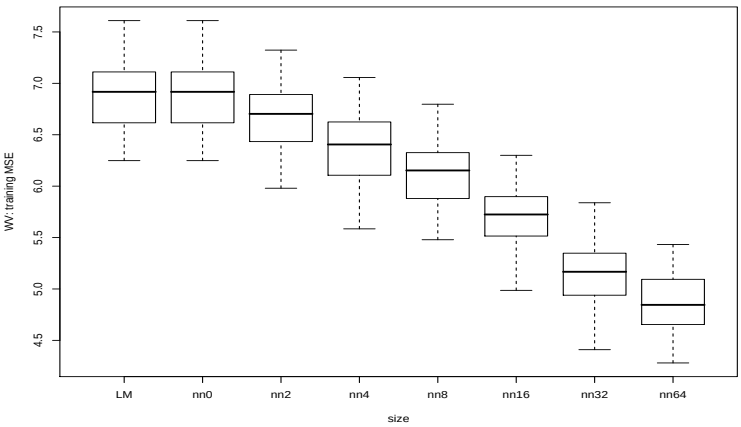


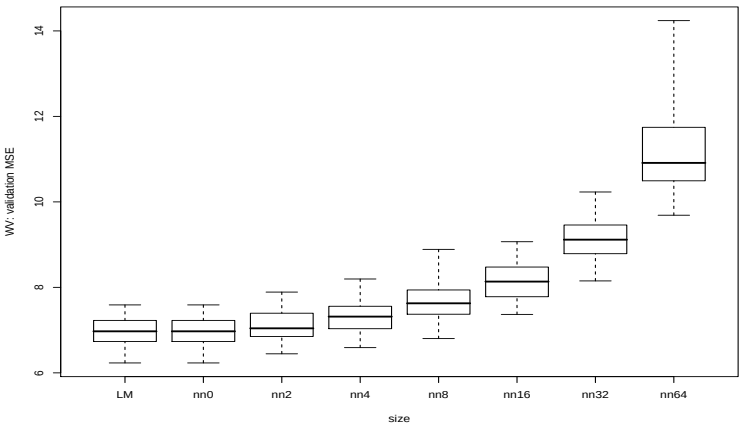
WU:


WU: training MSE
3.0
3.5
4.0
4.5
LM
nn0
nn2
nn4
nn8
nn16
nn32
nn64
size

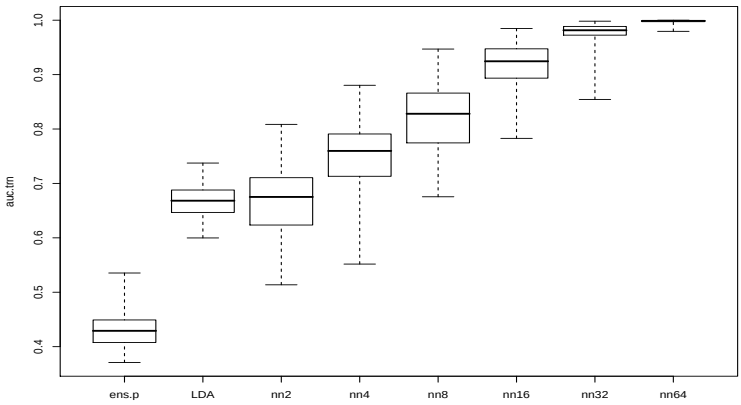

WU: validation MSE
4
5
6
7
LM
nn0
nn2
nn4
nn8
nn16
nn32
nn64
size

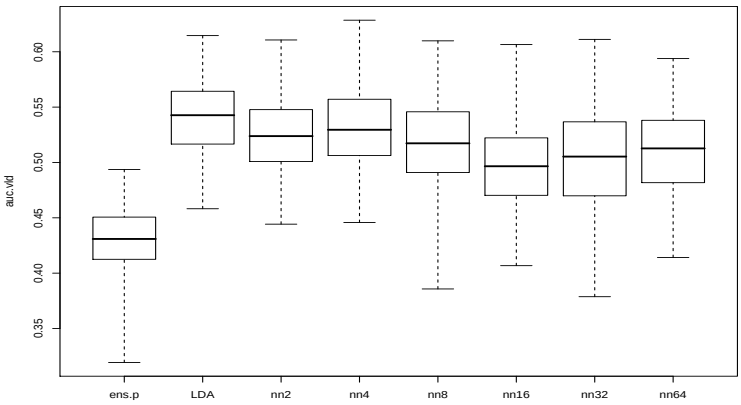
WV:


WV: training MSE
4.5
5.0
5.5
6.0
6.5
7.0
7.5
LM
nn0
nn2
nn4
nn8
nn16
nn32
nn64
size


WV: validation MSE
6
8
10
12
14
LM
nn0
nn2
nn4
nn8
nn16
nn32
nn64
size

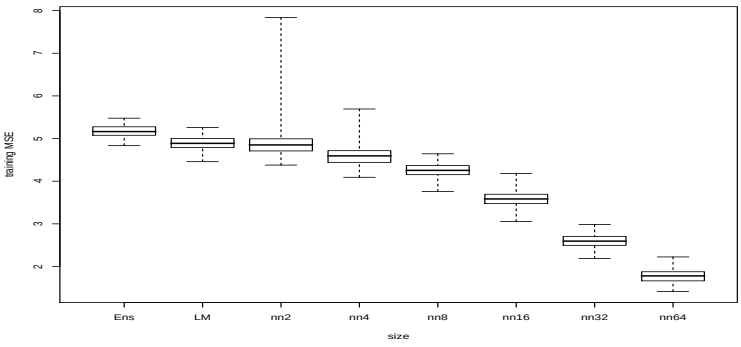
Prcp:

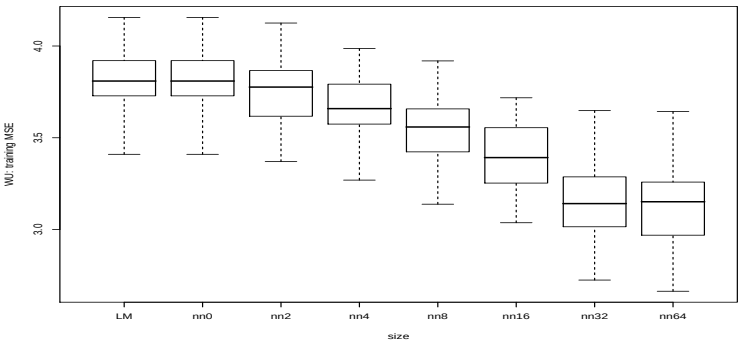

auc.tm
0.4
0.5
0.6
0.7
0.8
0.9
1.0
ens.p
LDA
nn2
nn4
nn8
nn16
nn32
nn64

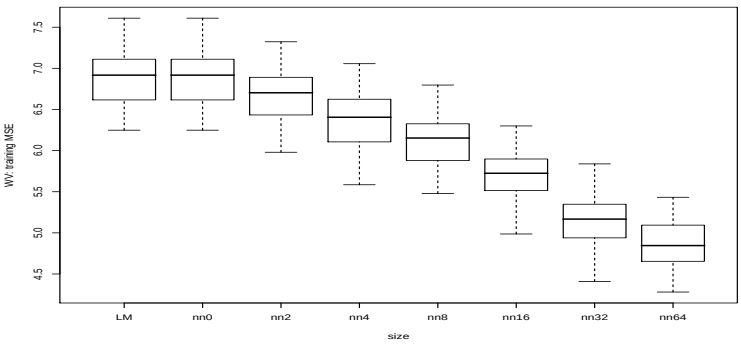

auc.vld
0.35
0.40
0.45
0.50
0.55
0.60
ens.p
LDA
nn2
nn4
nn8
nn16
nn32
nn64

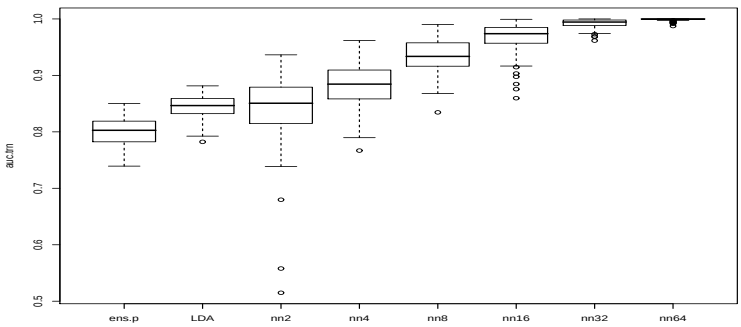
KSEA

Training

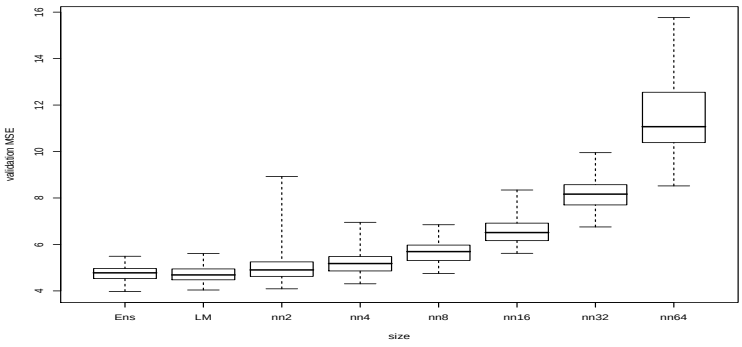


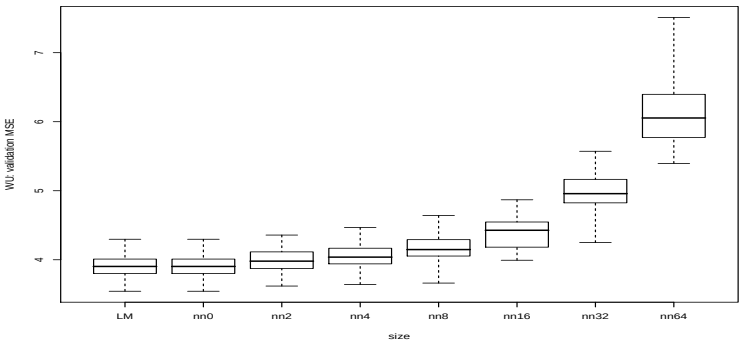


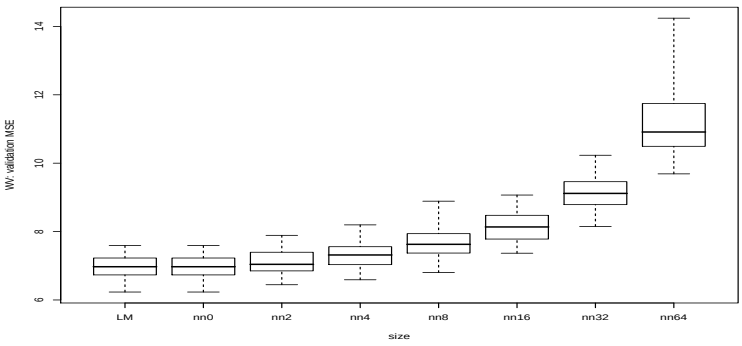


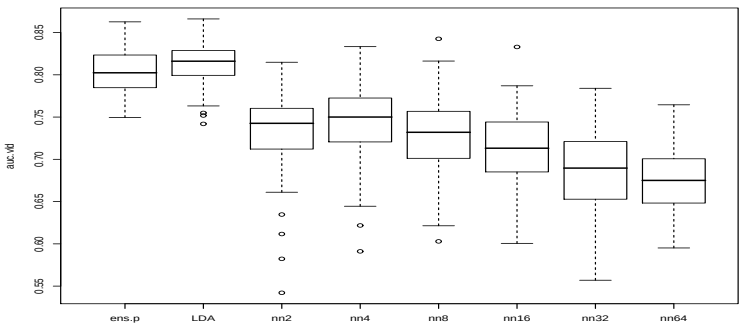


Validation









Conclusion I

Temperature, Wind U, Wind V:

- LM improves T, U and V predictions over ensemble means in terms of MSE.

- For wind U/V, adding T as additional covariate improves on MSE significantly for only a few stations. Forecasts of T are improved by adding wind U/V for most stations.

- For wind U/V, NN does not improve over LM in terms of MSE; forecasts of T from NN are worse than LM for all stations.

Conclusion II

Precipitation:

- LDA improves over ensemble proportion, except for a few stations in the Northwestern region.

- NN makes no improvement over LDA.
- A clear East-West spatial pattern observed when we compare LDA w/ additional covariates (LDA2) with LDA w/o additional covariates (LDA1).

We thank Fanyou Kong and Steve Leyton for the data.