Identify Potential to Improve Ensemble Sub-seasonal Precipitation and Temperature Forecasts With Machine Learning Technology

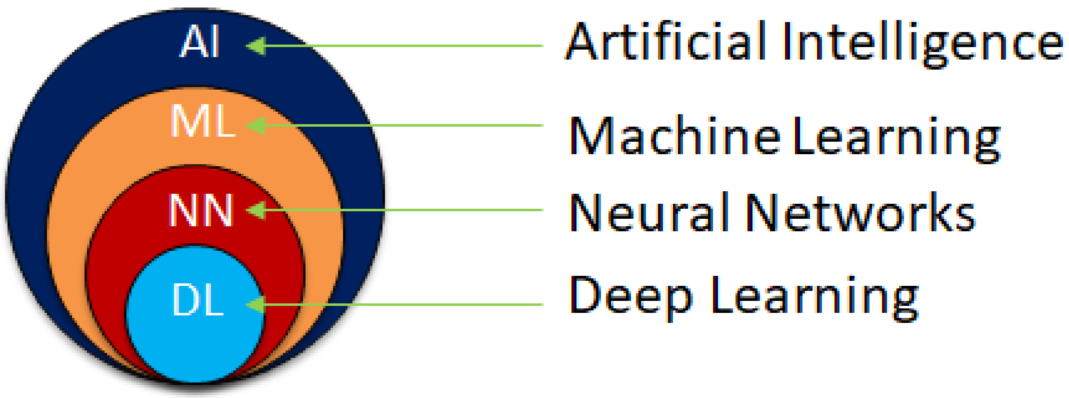
Yun Fan[1], Vladimir Krasnopolsky[2], Li Xu[3] and Jon Gottschalck[1]

[1]Climate Prediction Center
[2] Environmental Modeling Center
[3]ERT at Climate Prediction Center
NOAA Center for Weather and Climate Prediction

48th NOAA CDPW, Mar. 26-29, 2024 Tallahassee, FL

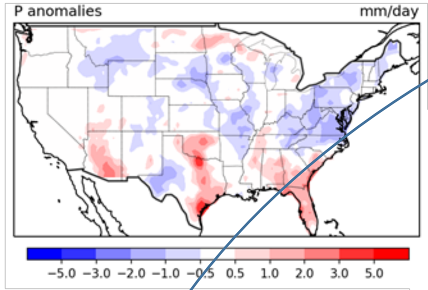
Outline

- Motivation & DL Basic
- Applications
- Summary

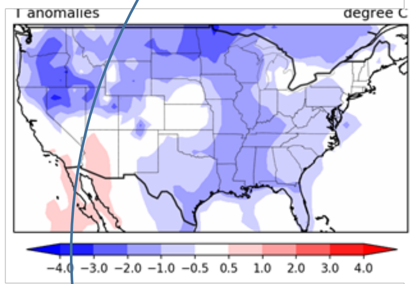


Deep Learning (Multi-layer neural network)

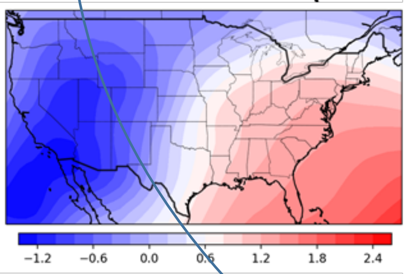
P forecast (z score)



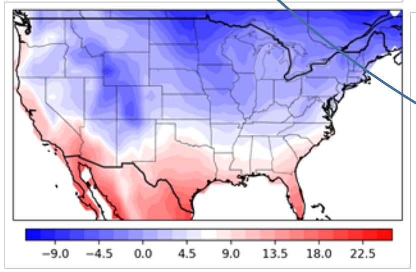
T forecast (z score)

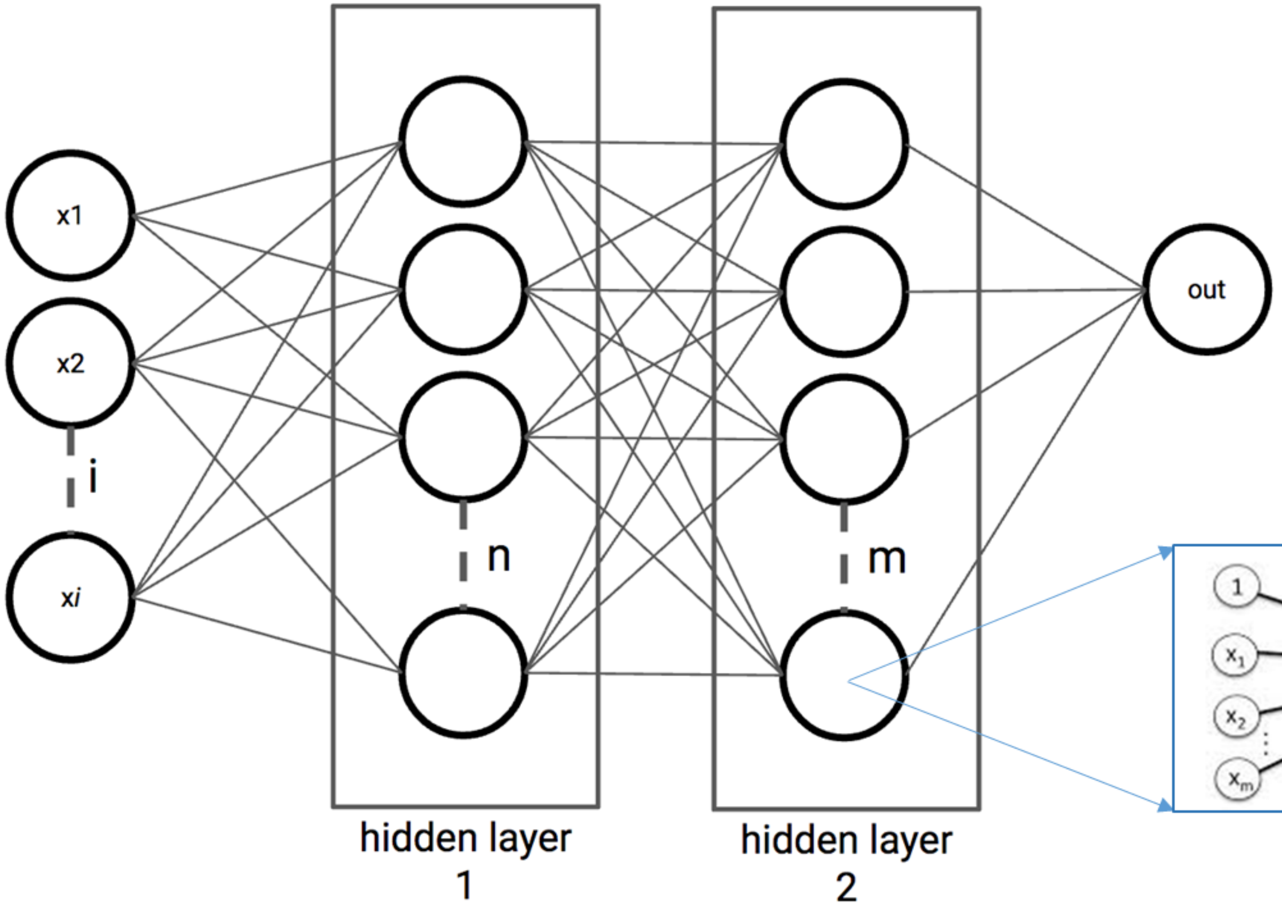


Z500 forecast (z score)



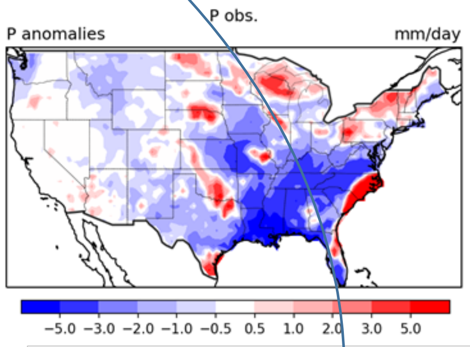
T/P climatology

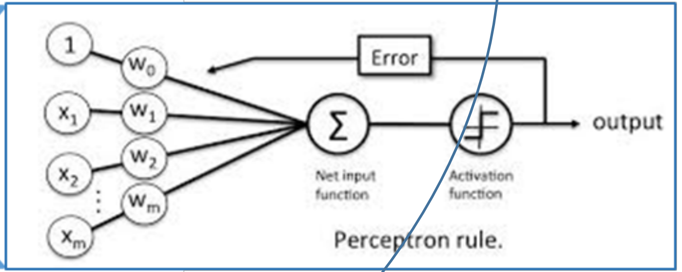




Big Data!!!

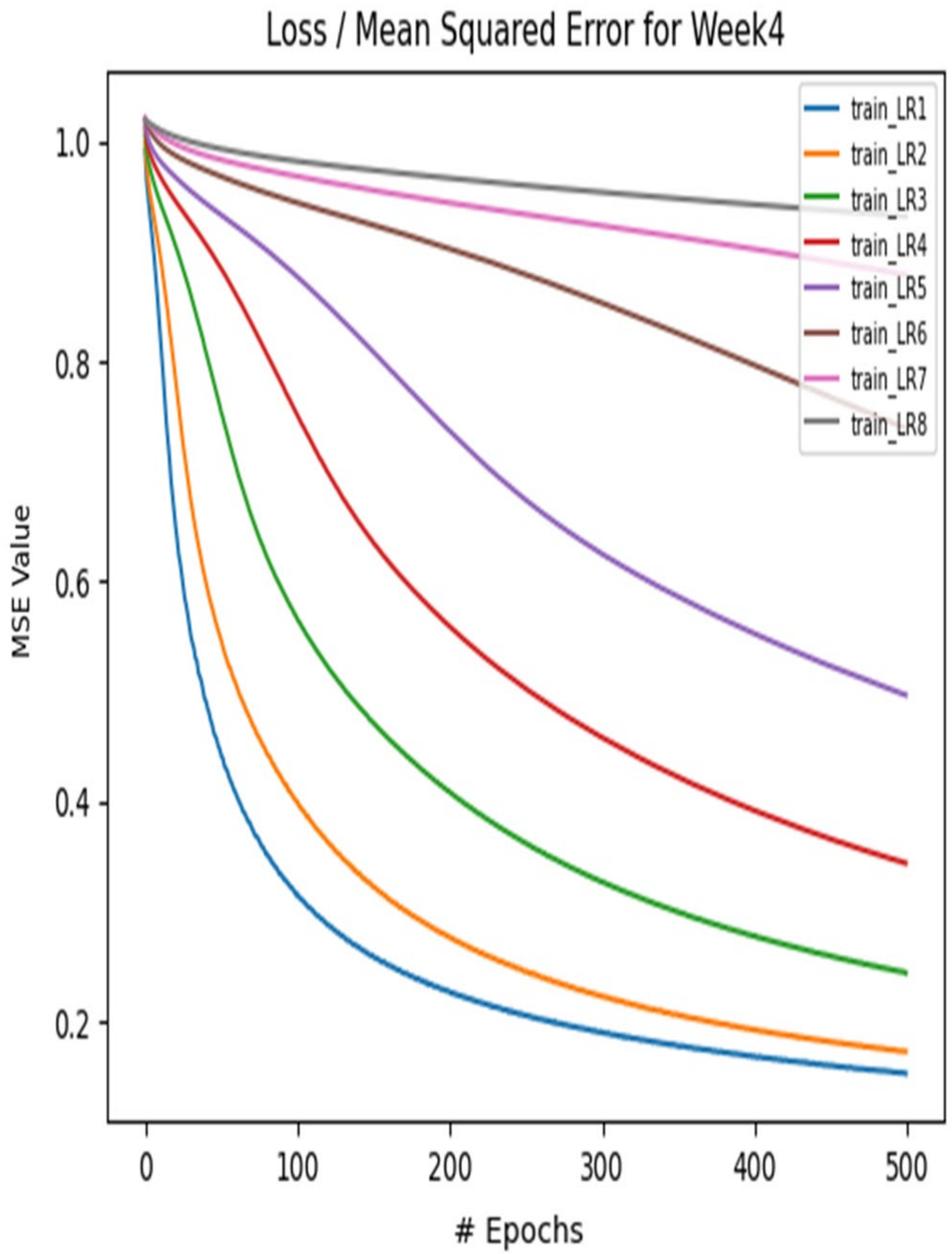
T/P obs (CPC unified)



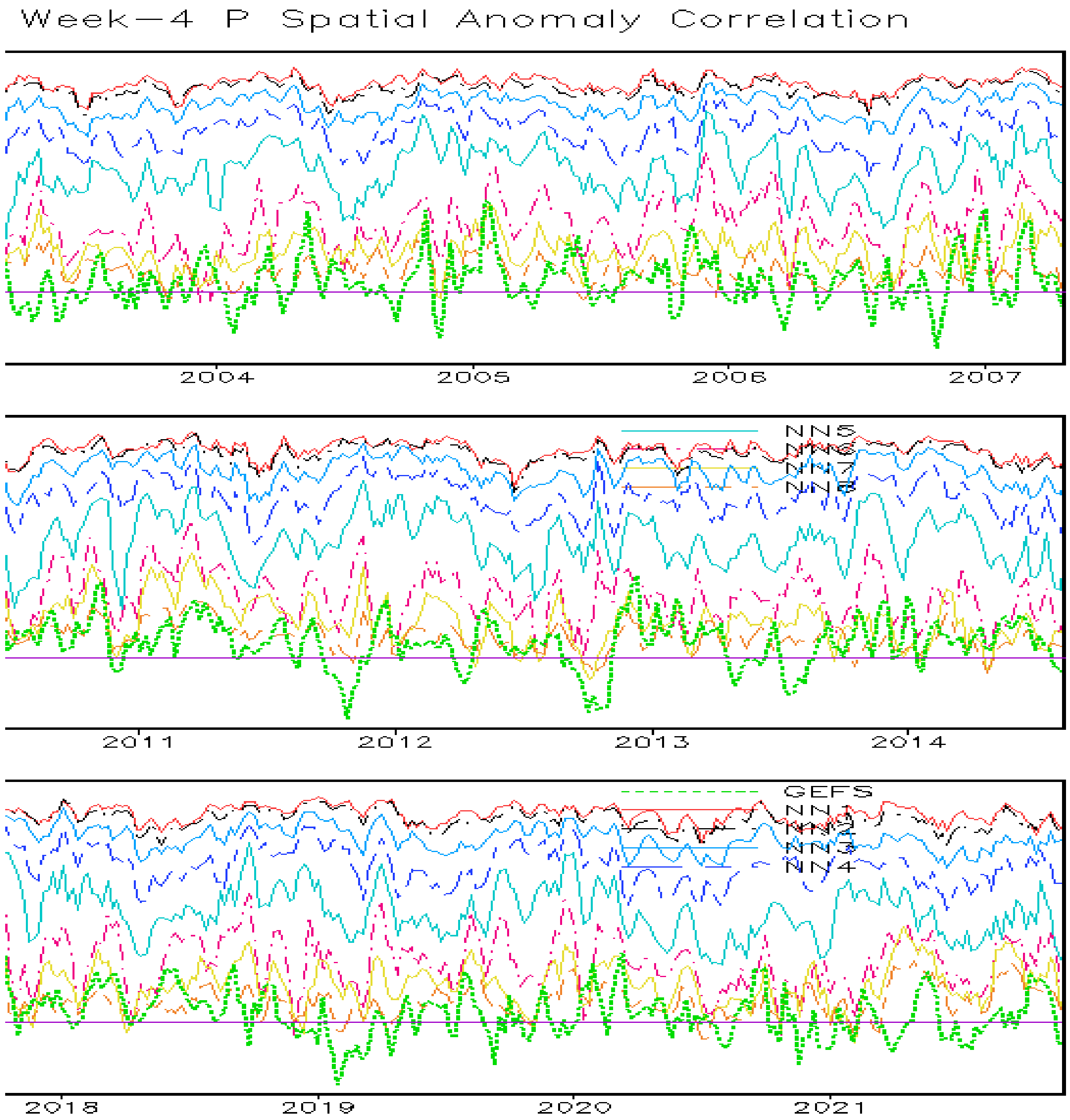


$$E = \frac{1}{N}\sum_{t=1}^{N}[Z_t - NN(X_t)]^2$$

Optimizer ： Root Mean Square Propagation (RMSProp)

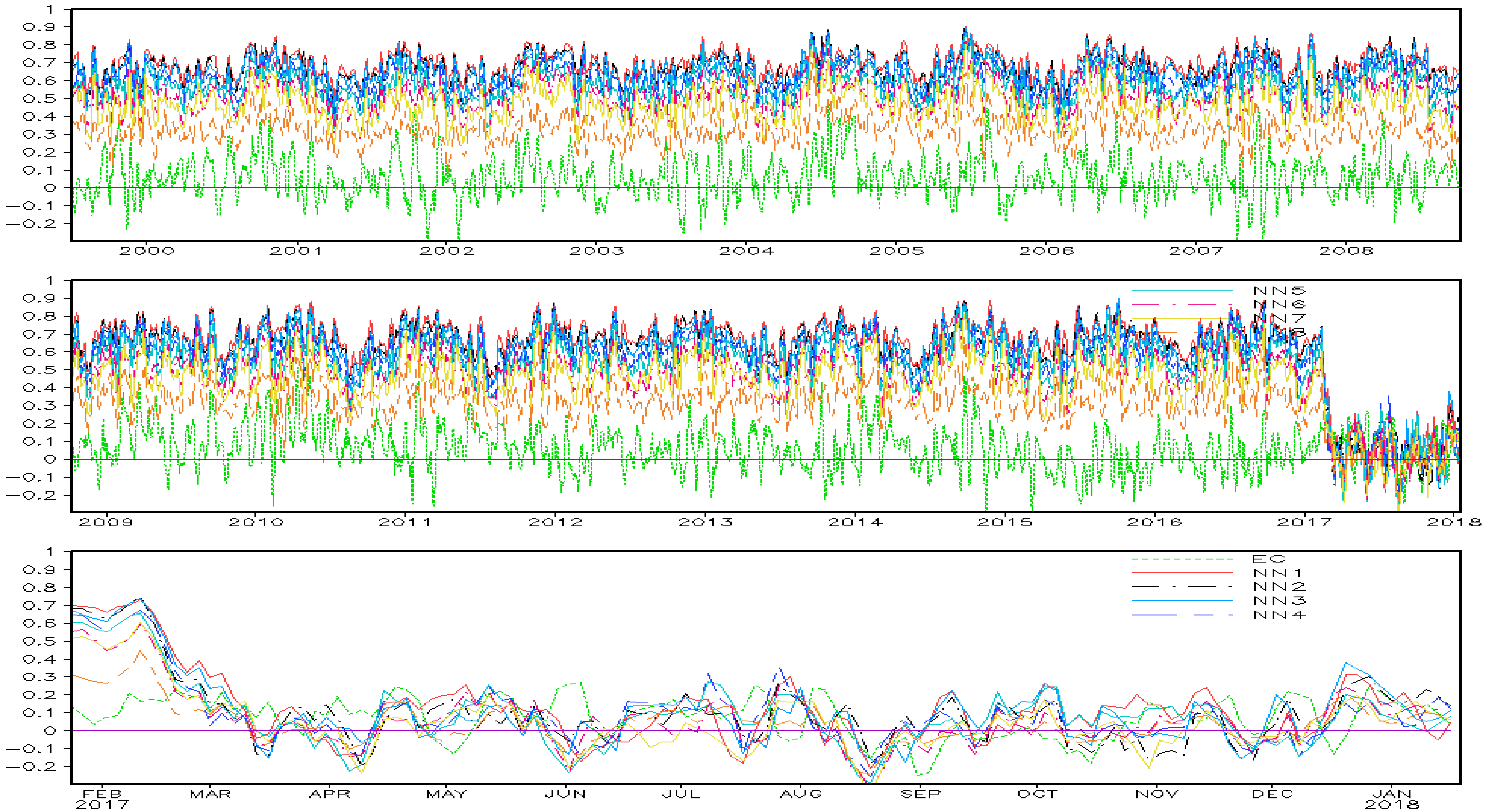


Loss / Mean Squared Error for Week4
train_LR1
train_LR2
train_LR3
train_LR4
train_LR5
train_LR6
train_LR7
train_LR8
MSE Value
Epochs

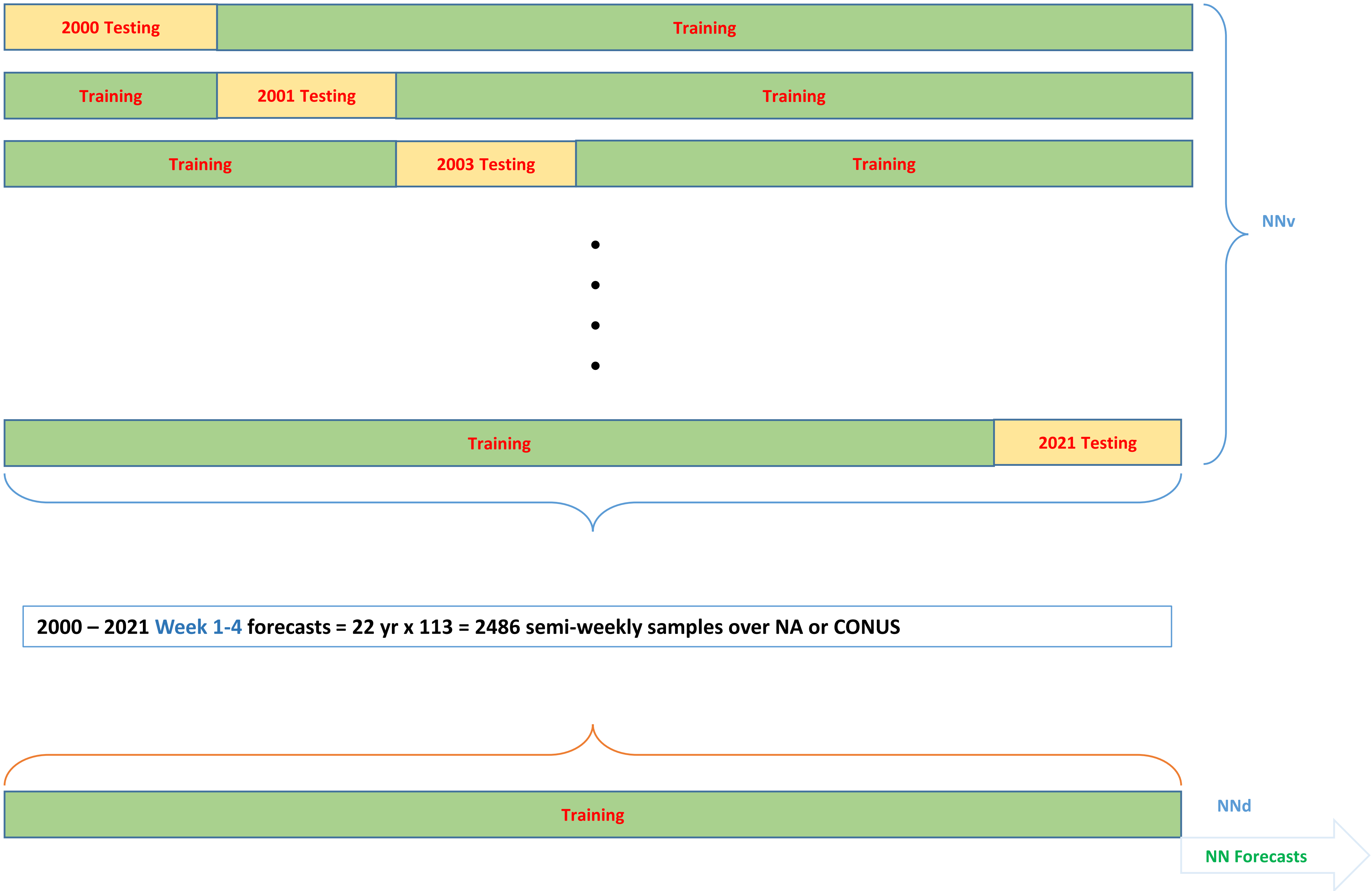


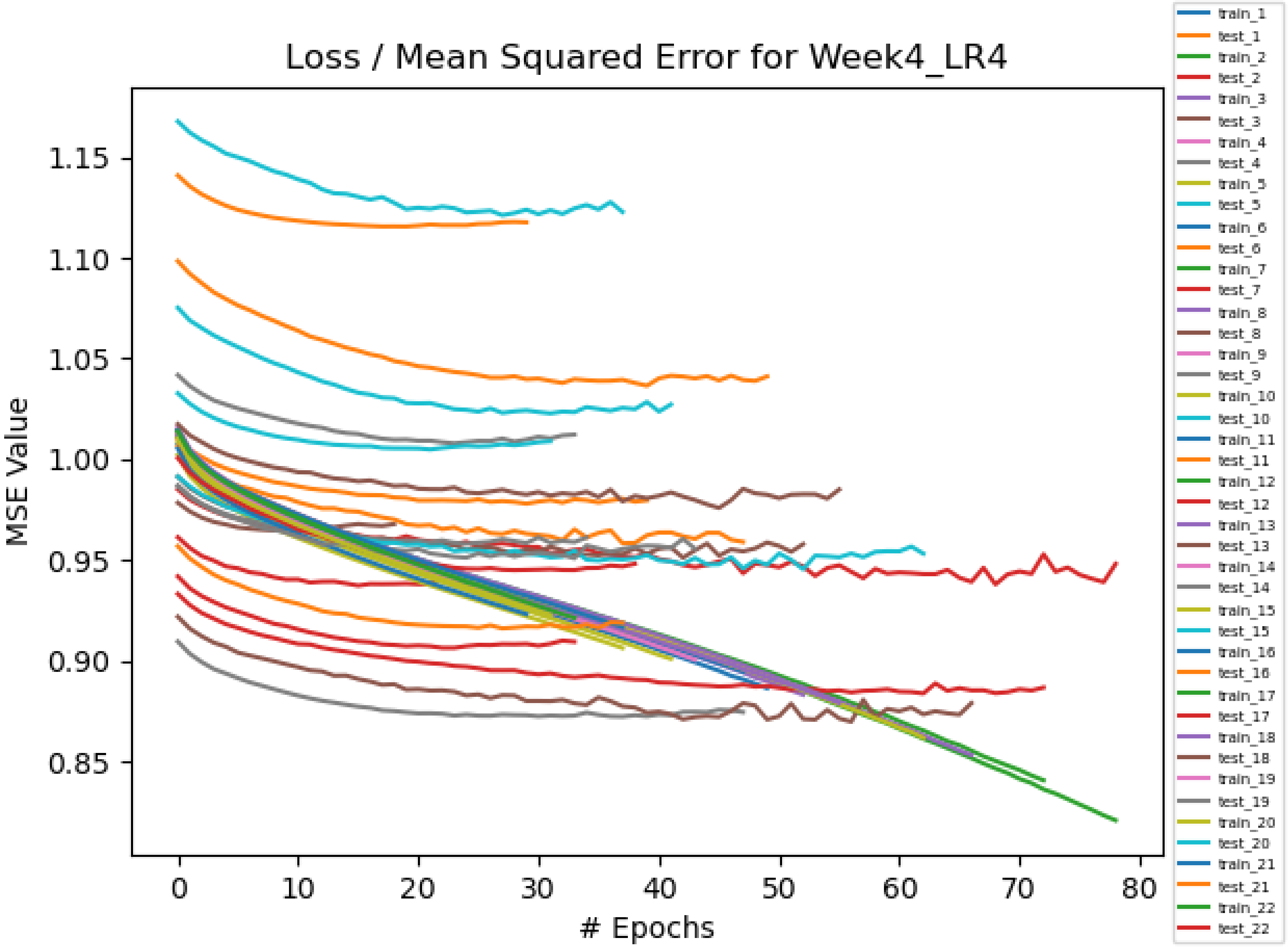
Week-4 P Spatial Anomaly Correlation
2004
2005
2006
2007
2011
2012
2013
2014
2018
2019
2020
2021
GEFS
NN1
NN2
NN3
NN4
NN5
NN6
NN7
NN8

Time Series of ECMWF Forecast Week-4 P Spatial Anomaly Correlation



Twenty-two leave one year out Cross-Validations for 2000-2021

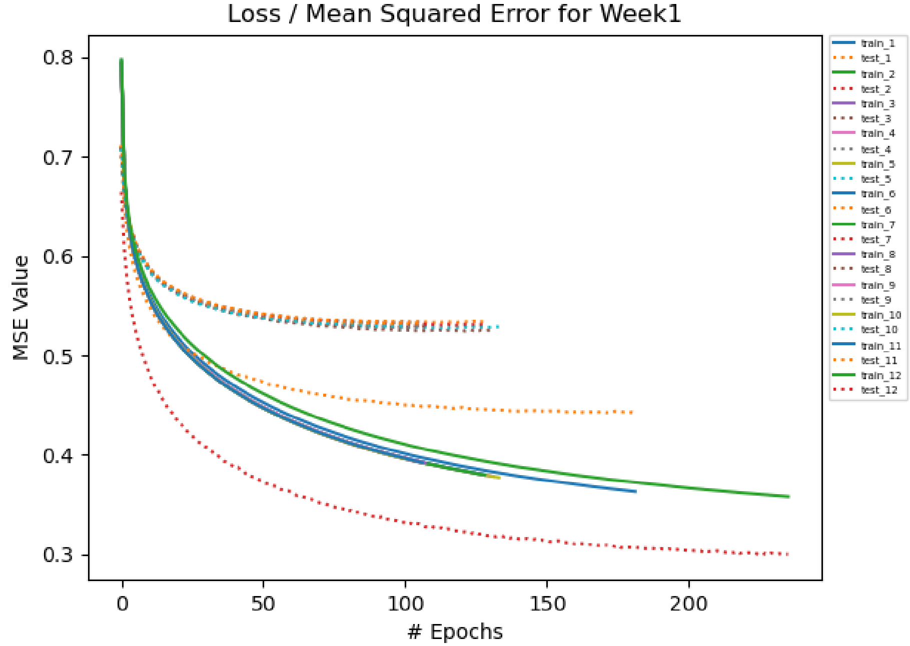




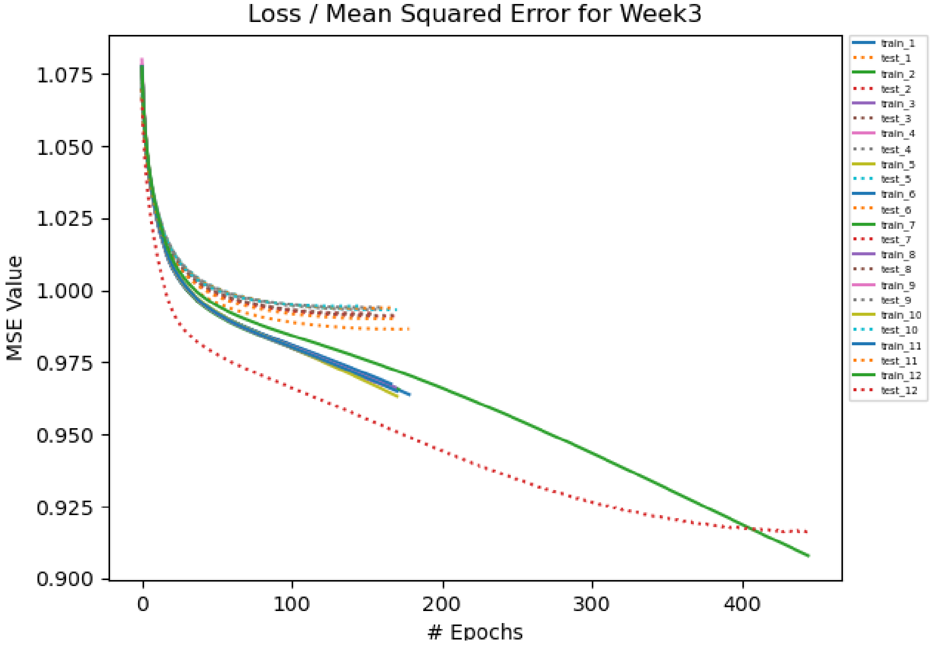
Loss / Mean Squared Error for Week4_LR4
MSE Value
Epochs
1.15
1.10
1.05
1.00
0.95
0.90
0.85
0
10
20
30
40
50
60
70
80
train_1
test_1
train_2
test_2
train_3
test_3
train_4
test_4
train_5
test_5
train_6
test_6
train_7
test_7
train_8
test_8
train_9
test_9
train_10
test_10
train_11
test_11
train_12
test_12
train_13
test_13
train_14
test_14
train_15
test_15
train_16
test_16
train_17
test_17
train_18
test_18
train_19
test_19
train_20
test_20
train_21
test_21
train_22
test_22

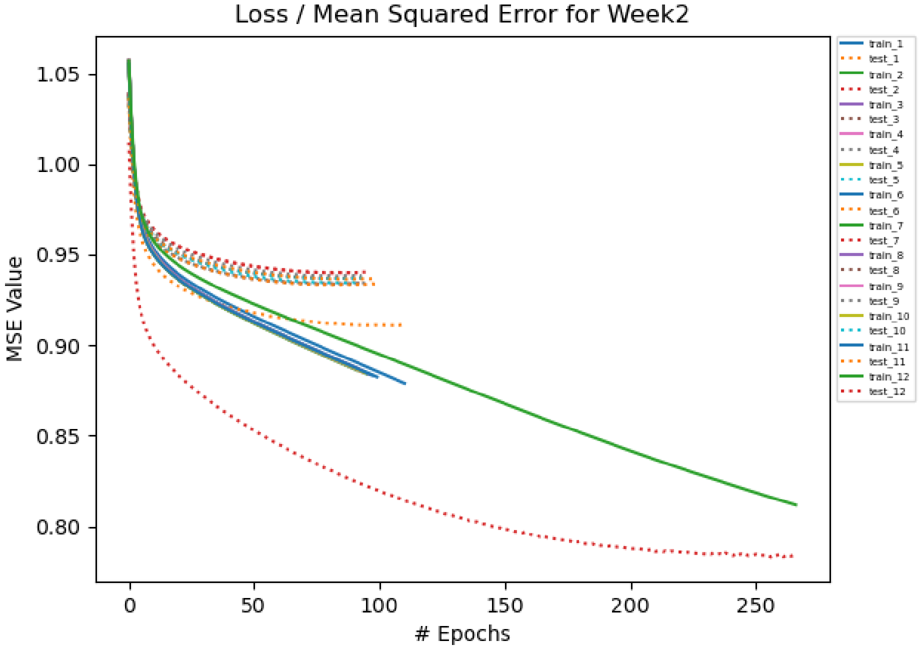
How about
Validation Across Individual Members?

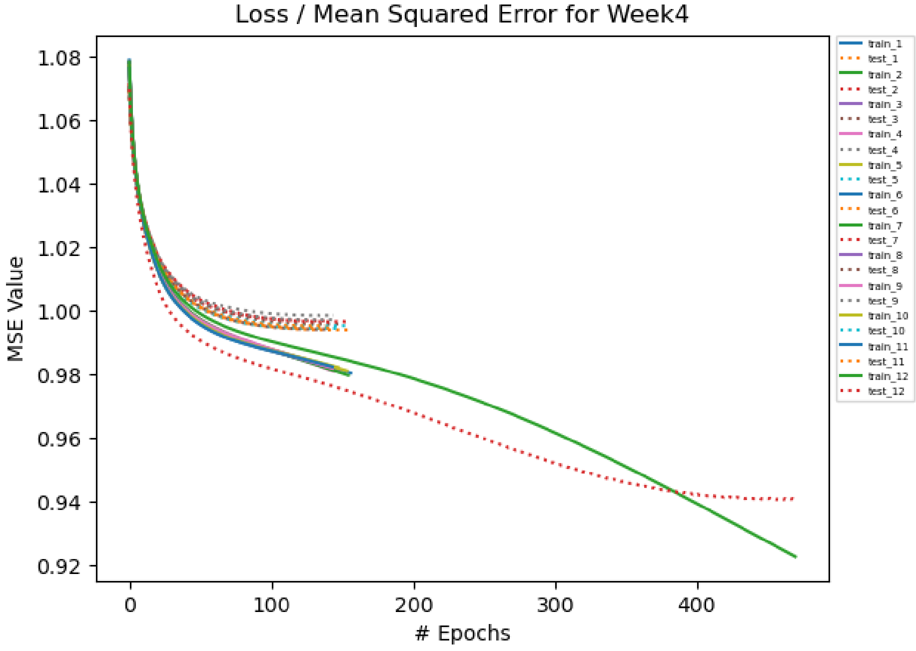
Precipitation



Data: 2000-2021 ECMWF Reforecasts.
1: control run, 2-11: perturbation run, 12: ensemble mean

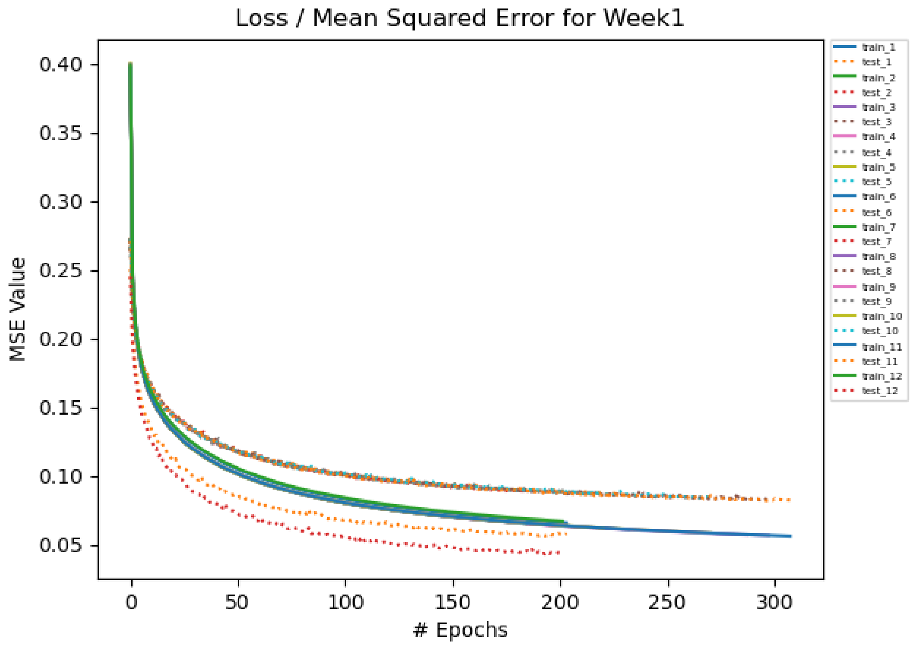


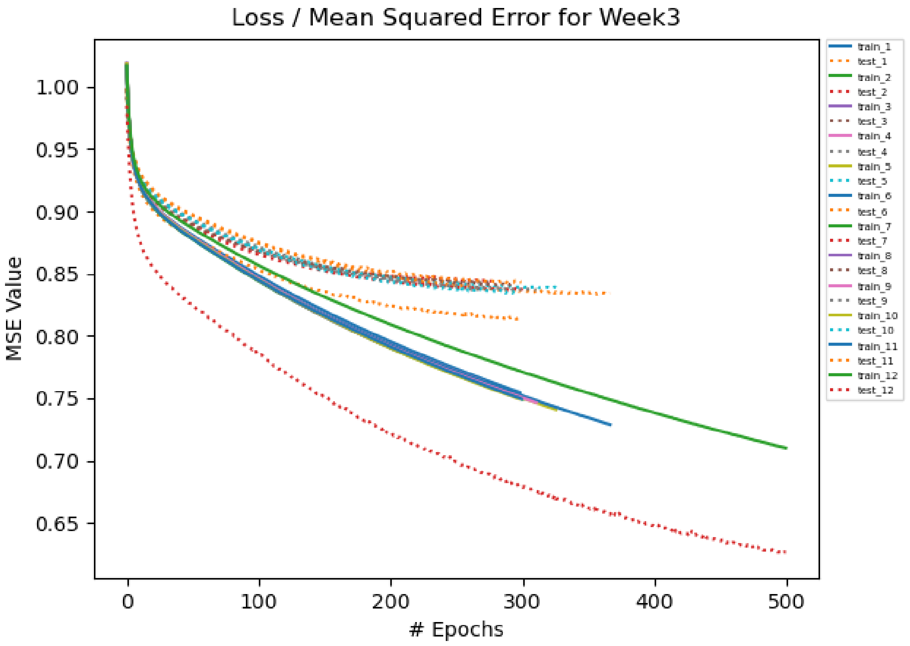


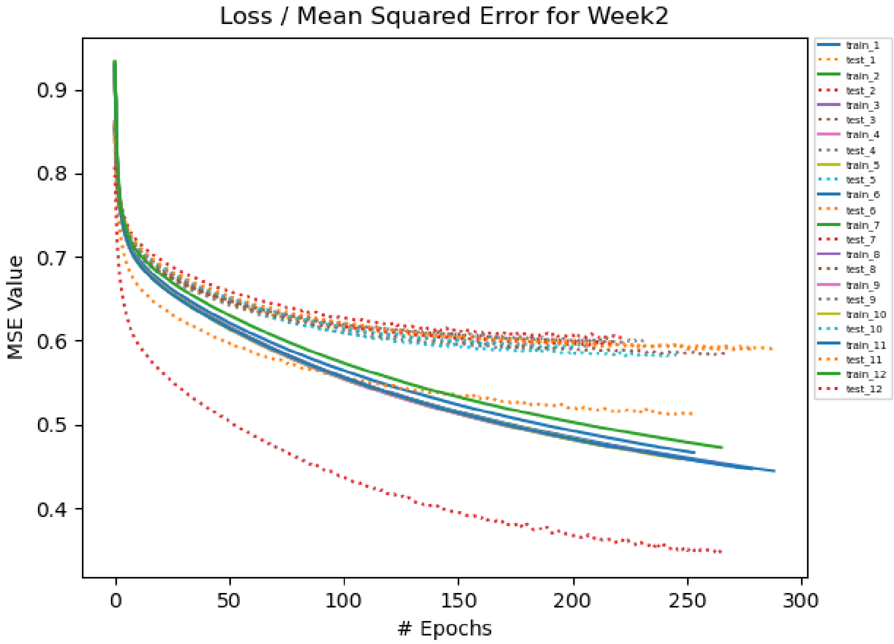


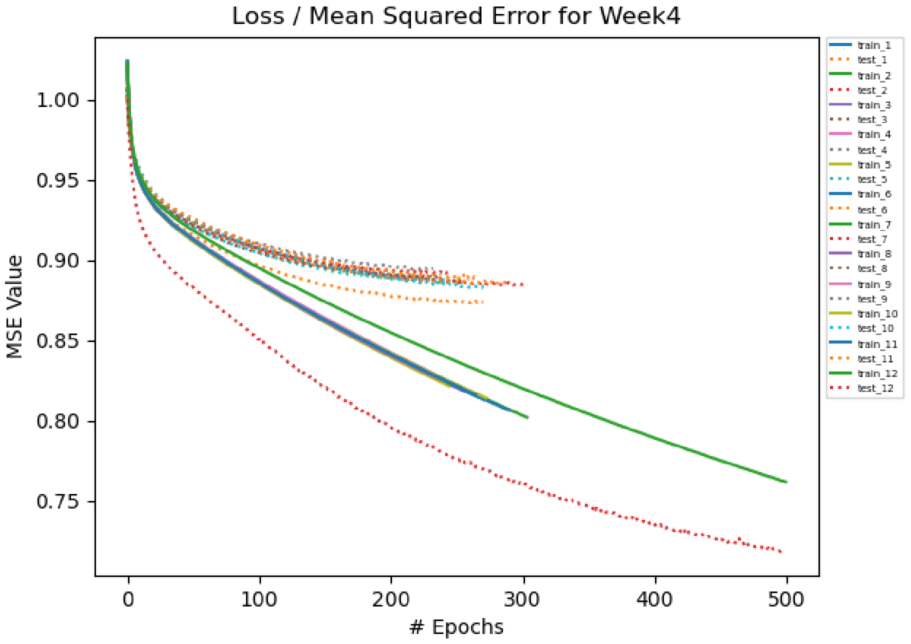
2m Temperature

Data: 2000-2021 ECMWF Reforecasts.
1: control run, 2-11: perturbation run, 12: ensemble mean

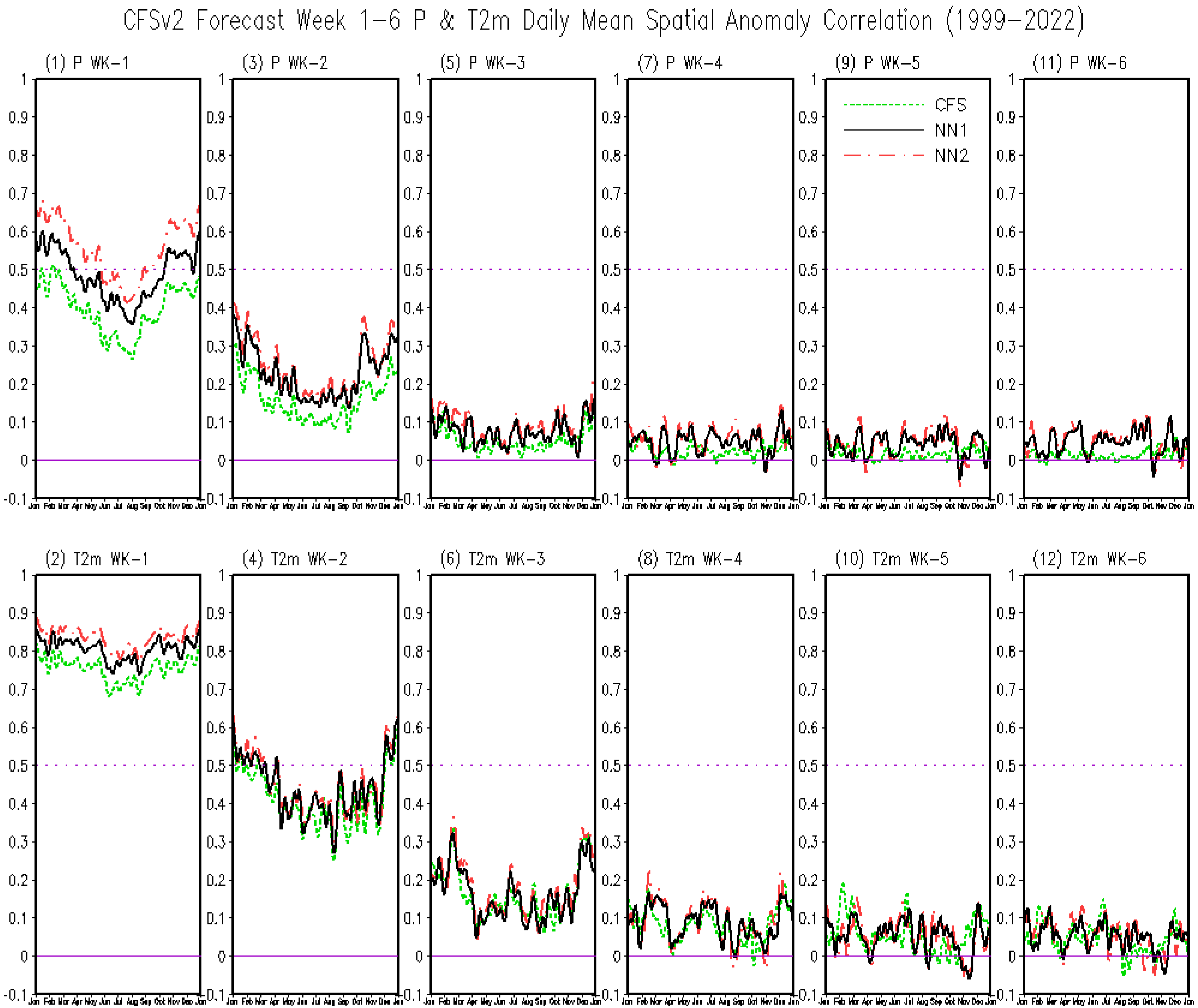






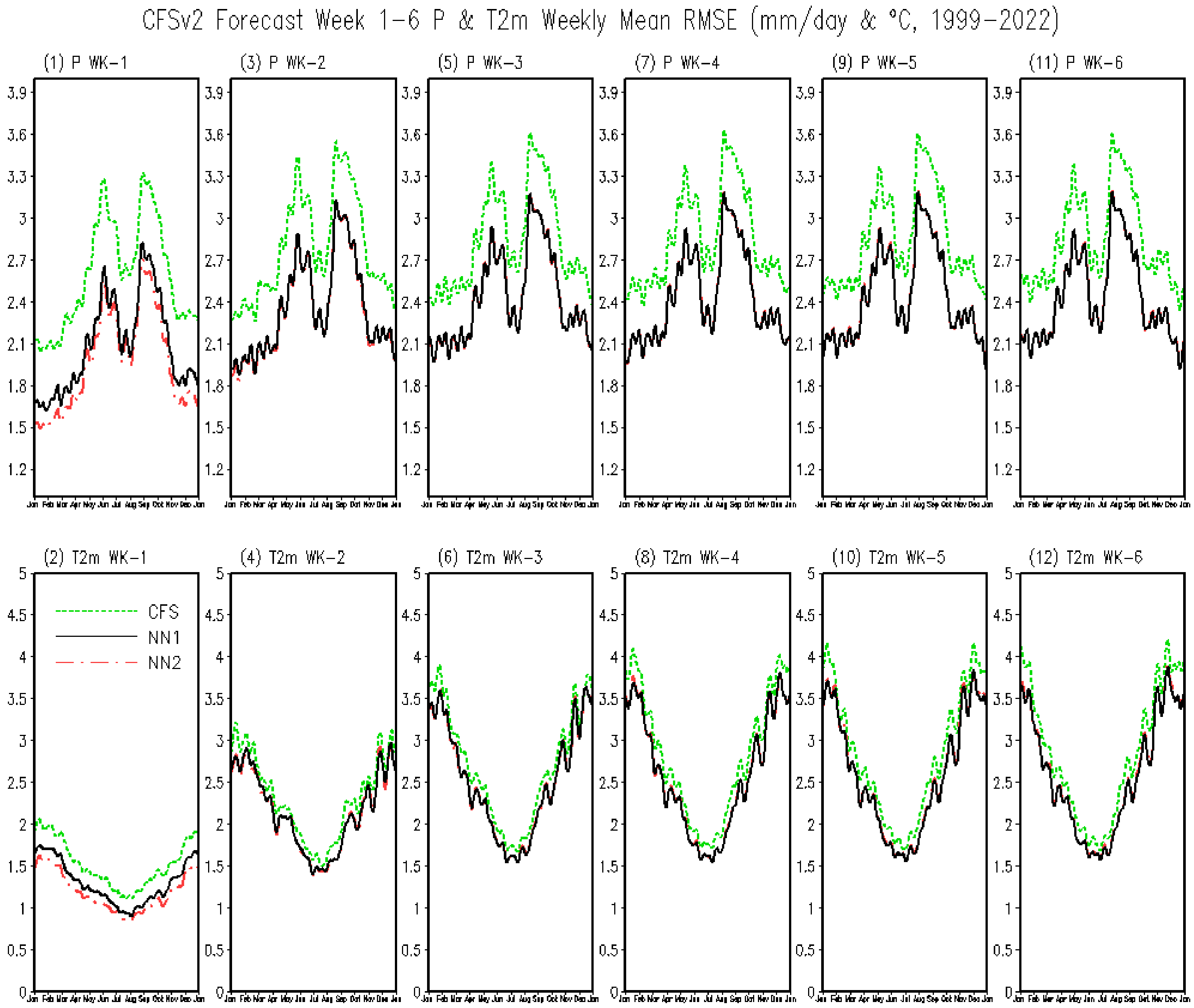


Using Multiple Days
Initial Conditions Help?

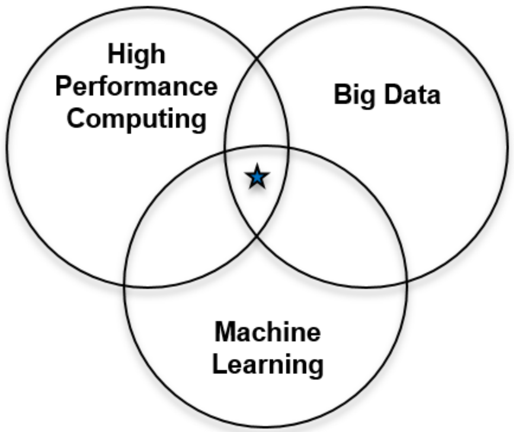


Dataset: 1999-2022 CFSv2
NN1 – Using one days ICs-FCSTs
NN2 – Using 3 day ICs-FCSTs

Dataset: 1999-2022 CFSv2
NN1 – Using one days ICs-FCSTs
NN2 – Using 3 day ICs-FCSTs



Summary



1. DL advantages

   Flexible nonlinear tool & Easy to handle BIG DATA

2. Unique & beneficial NN architectures

   extract more sophisticated info hidden behind multiple dimensional big data
   improve subseasonal P & T2m FCSTs

3. ML as a diagnostic tool – identify potential to improve S2S ensemble FCSTs:

   use better model for perturbation runs (e.g. control run model)
   use more ICs (Daily CFSv2, ~~Semiweekly ECMWF~~, ~~Weekly GEFSv12~~)

4. ML applications & limitations

   Weather-climate modeling, data-assimilation, post-processing & diagnosing etc.