Ensemble forecasting:

Error bars and beyond

Jim Hansen, NRL

Walter Sessions, NRL

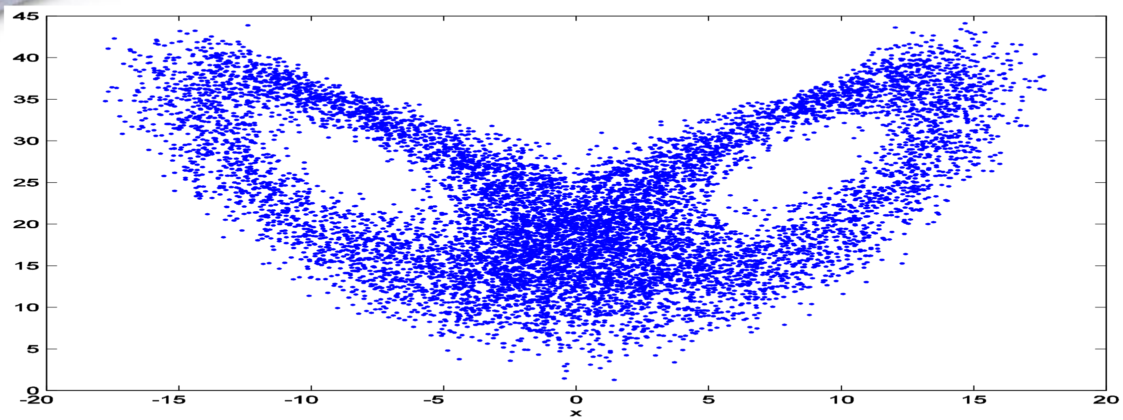
Jeff Reid,NRL

May, 2011

Why ensembles

- Traditional justification
  - Predict expected error
- (Perhaps) more valuable justification
  - Improve observation systems
  - Improve data assimilation
  - Statistically probe system dynamics for improved understanding

Lorenz '63 attractor



z

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

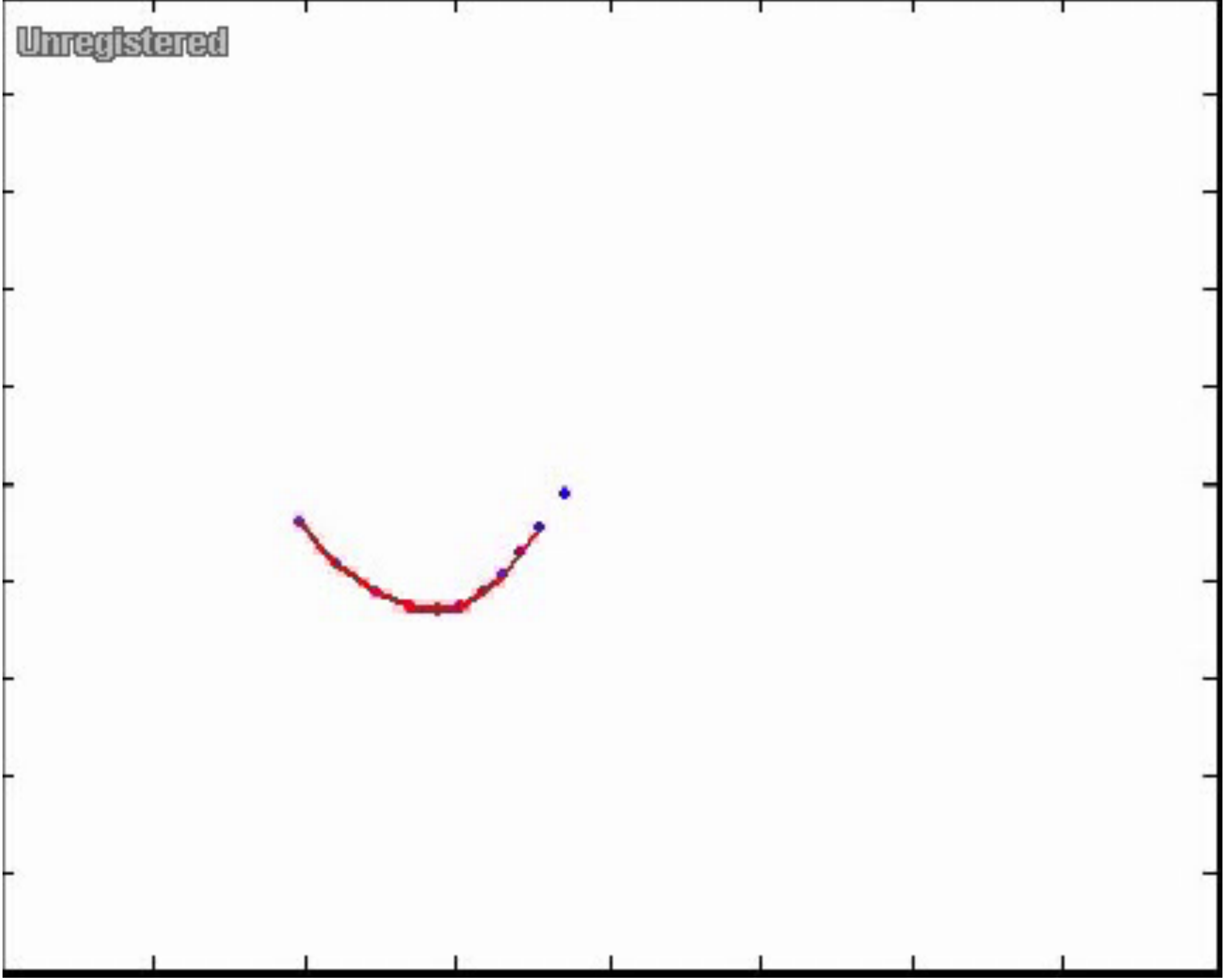
$$\frac{dz}{dt} = xy - \beta z$$

Blue dots represent solutions to the equations

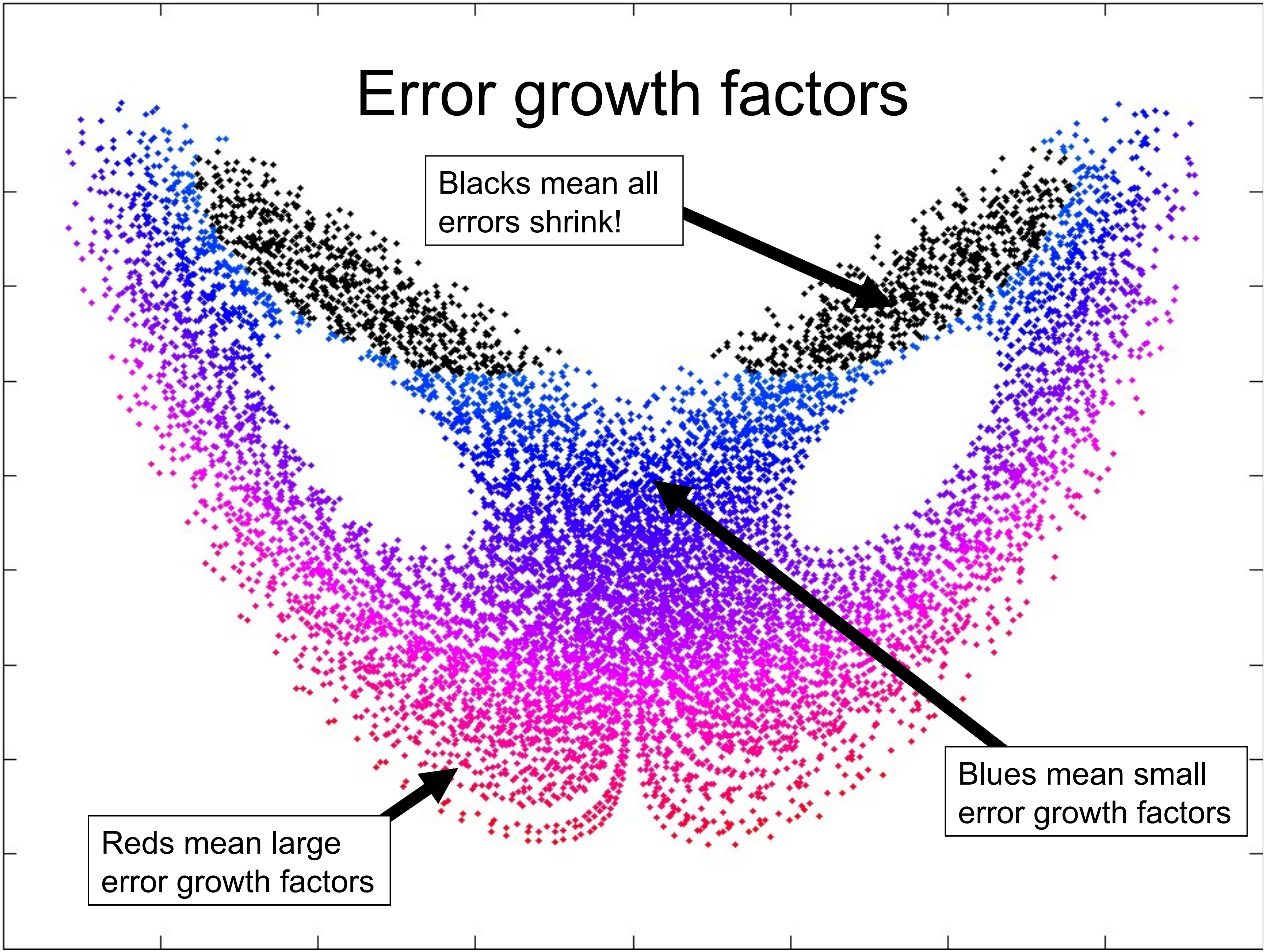
The system lives here

It does not live here

x



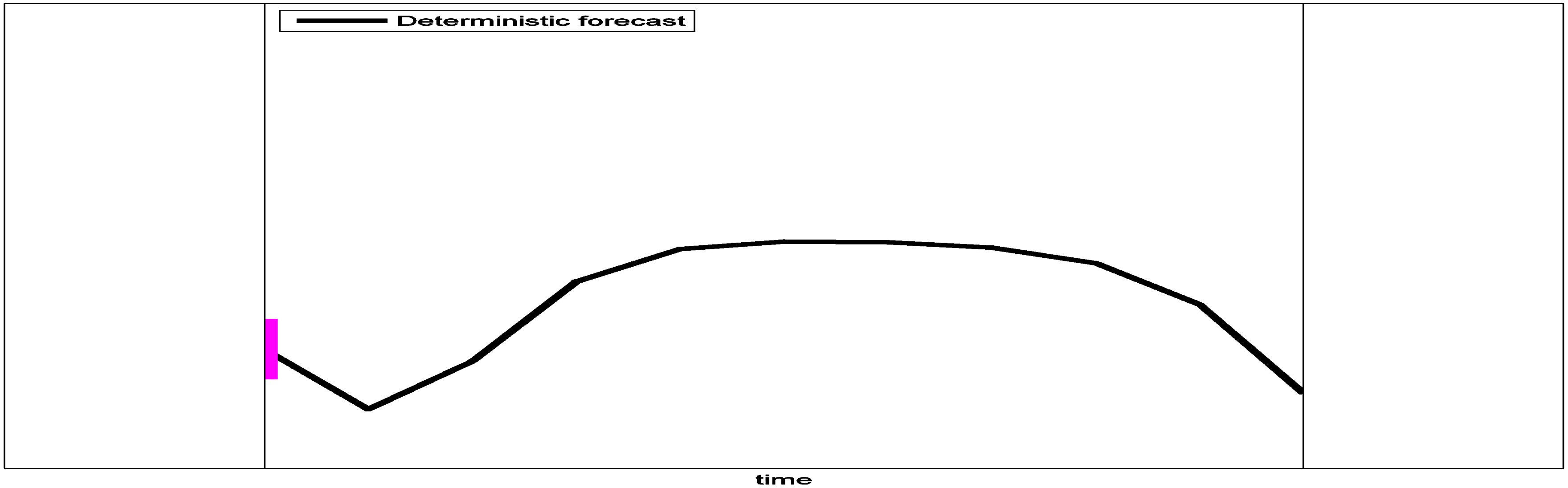
Unregistered



Error growth factors
Blacks mean all errors shrink!
Blues mean small error growth factors
Reds mean large error growth factors

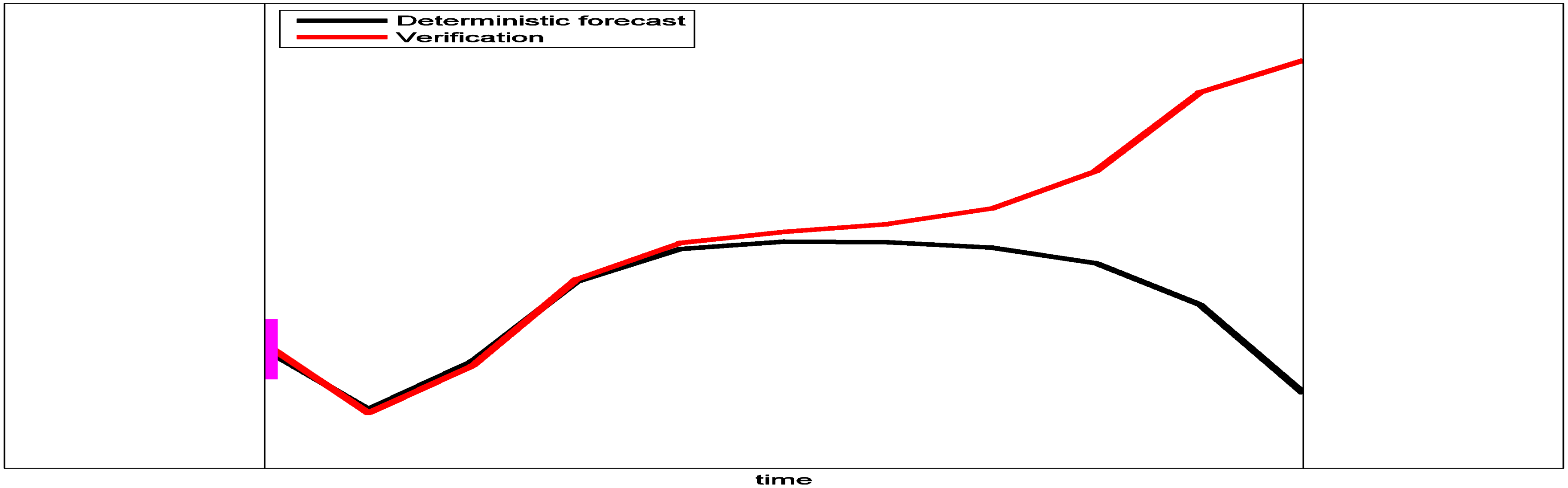
t=0

t=3days



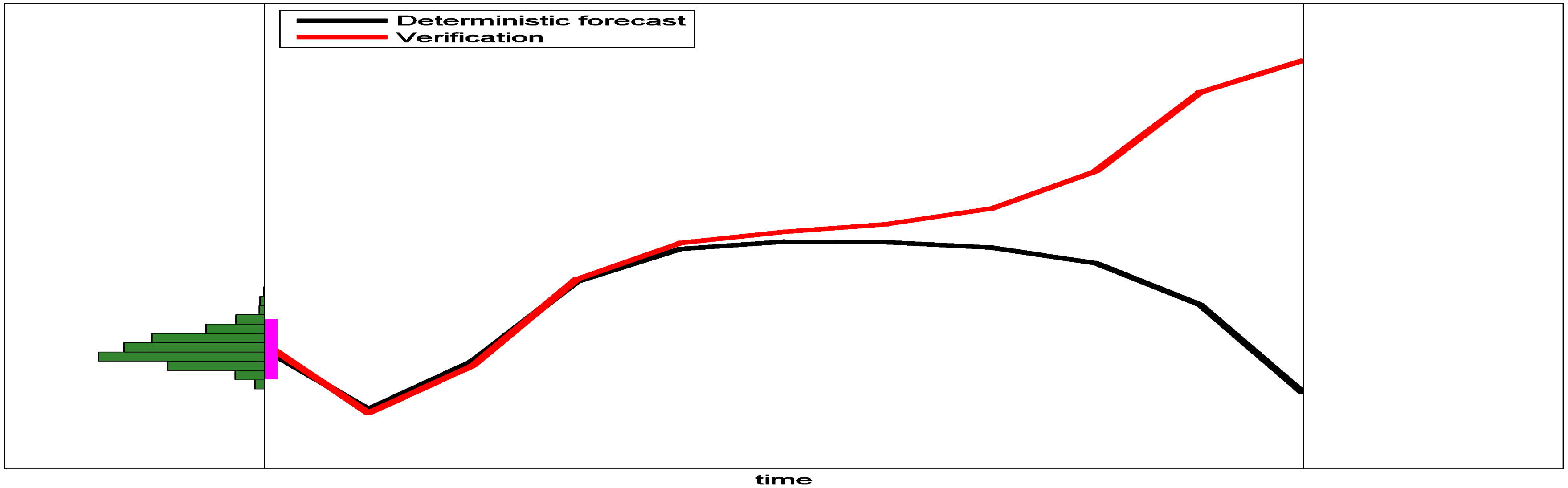
t=0

t=3days



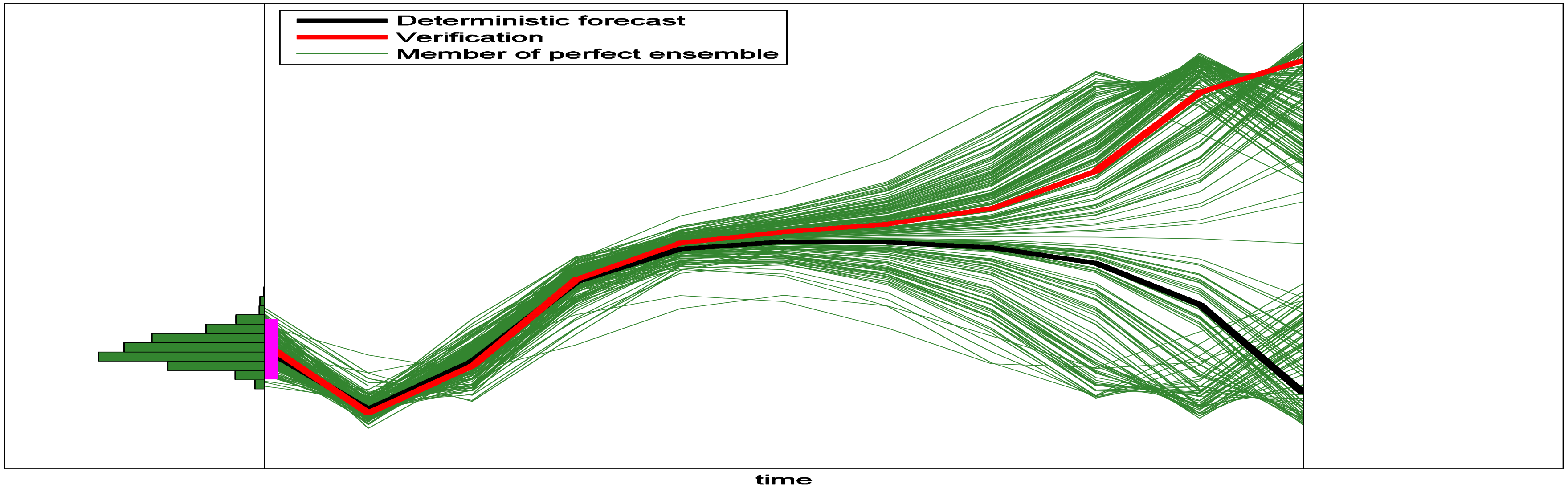
t=0

t=3days



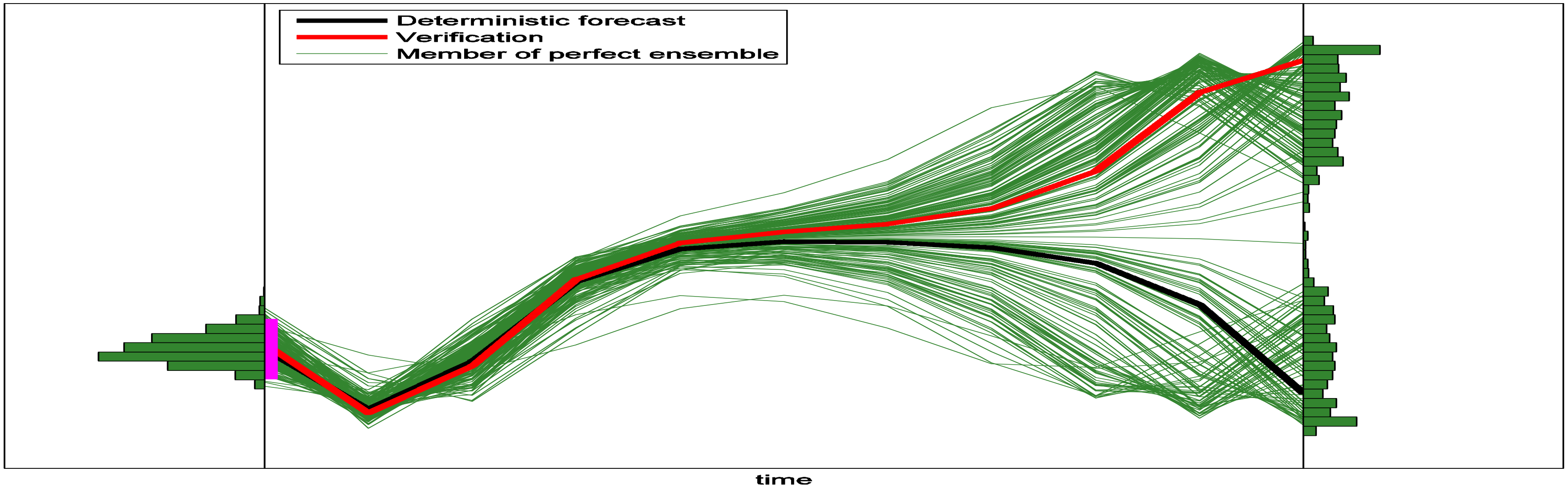
t=0

t=3days


Deterministic forecast
Verification
Member of perfect ensemble
time

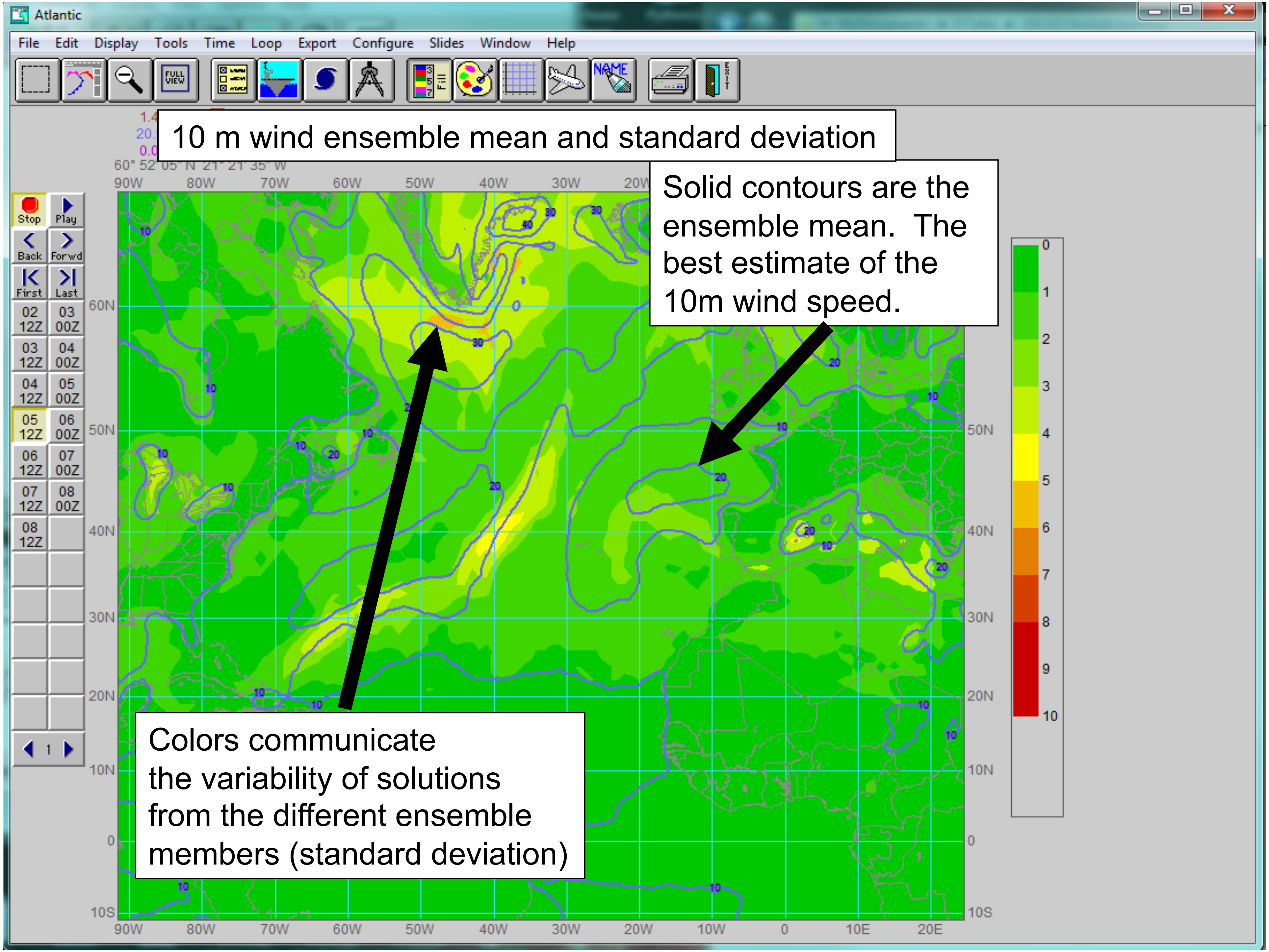
t=0

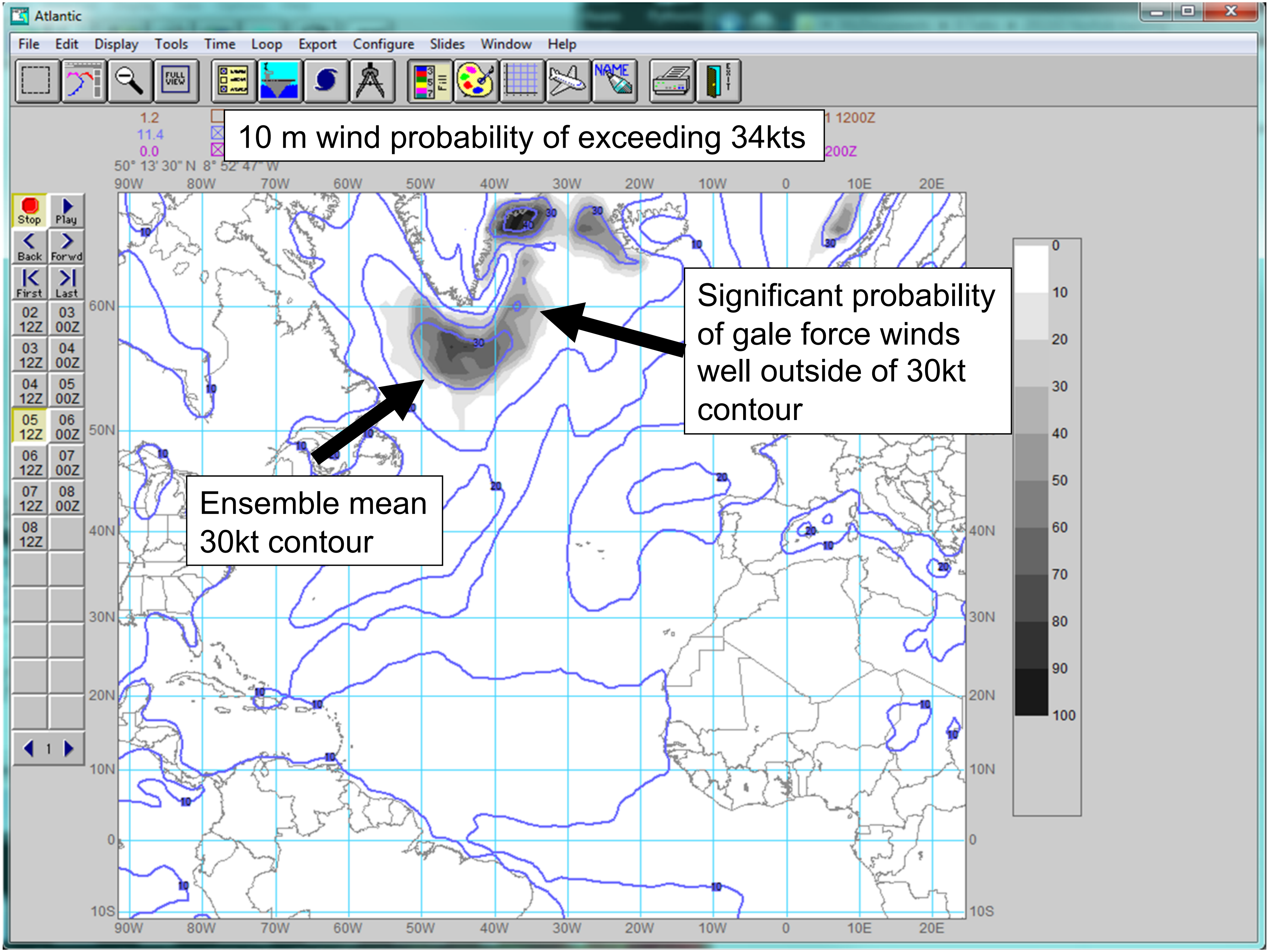
t=3days


Deterministic forecast
Verification
Member of perfect ensemble
time

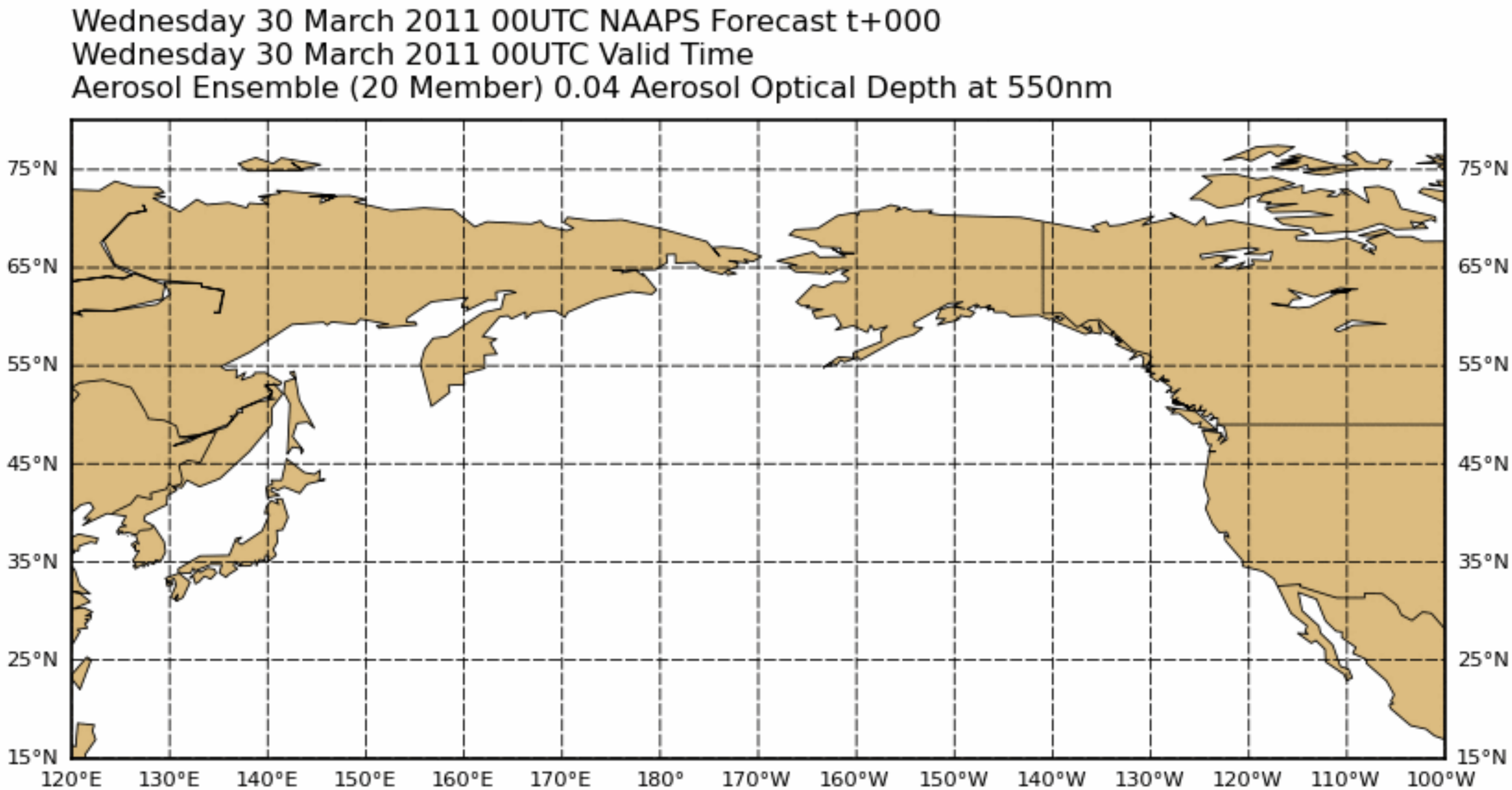
t=0

t=3days


Atlantic
File Edit Display Tools Time Loop Export Configure Slides Window Help
10 m wind ensemble mean and standard deviation
Solid contours are the ensemble mean. The best estimate of the 10m wind speed.
Colors communicate the variability of solutions from the different ensemble members (standard deviation)
Stop
Play
Back
Forwd
First
Last
90W 80W 70W 60W 50W 40W 30W 20W 10W 0 10E 20E
60N 50N 40N 30N 20N 10N 0 10S
0 1 2 3 4 5 6 7 8 9 10


Atlantic
File Edit Display Tools Time Loop Export Configure Slides Window Help
10 m wind probability of exceeding 34kts
Significant probability of gale force winds well outside of 30kt contour
Ensemble mean 30kt contour
90W 80W 70W 60W 50W 40W 30W 20W 10W 0 10E 20E
60N 50N 40N 30N 20N 10N 0 10S
0 10 20 30 40 50 60 70 80 90 100

FUKUSHIMA



Plots Generated Wednesday 30 March 2011 20UTCNRL/Monterey Aerosol Modeling
NOT OFFICIAL FNMOC NAAPS RUN

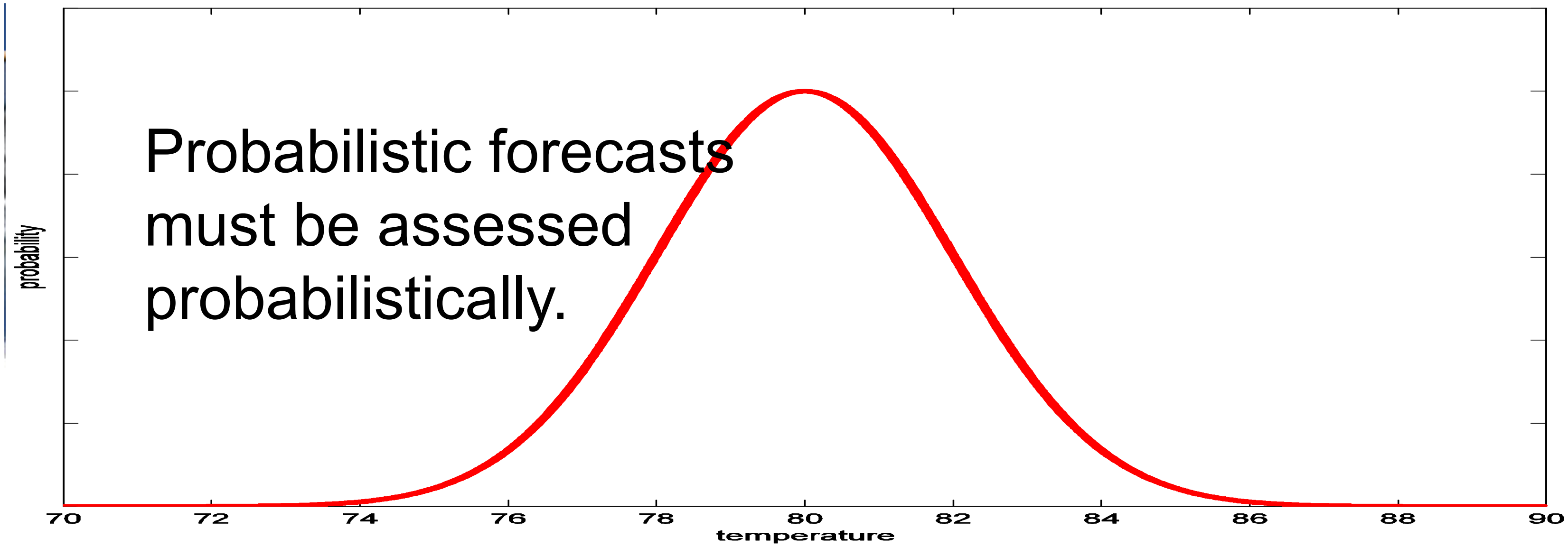
Sessions, et al

Menagerie of ensembles

- Single model ensemble
- Multi-parameterization ensemble
- Multi-parameter ensemble
- Multi-resolution ensemble
- Additive stochastic ensemble
- Multiplicative stochastic ensemble
- Poor-man's ensemble
- Ensemble of ensembles
- Hybrid ensemble

Ensemble verification

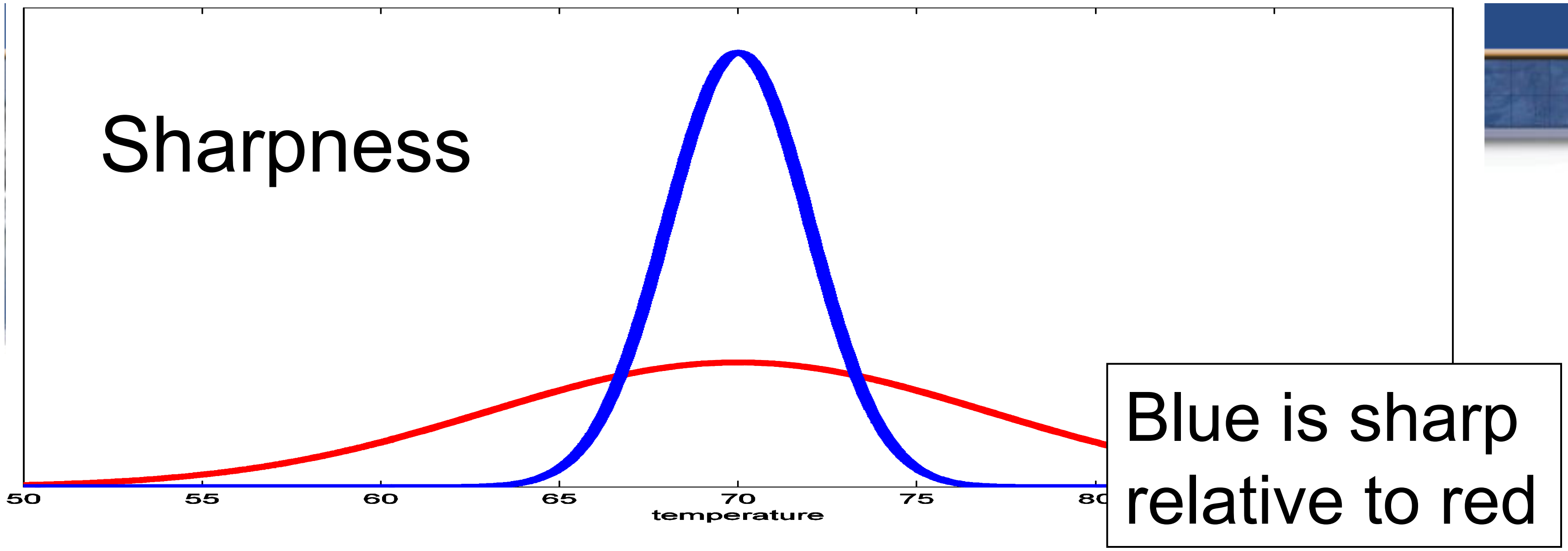
Probabilistic forecasts must be assessed probabilistically.



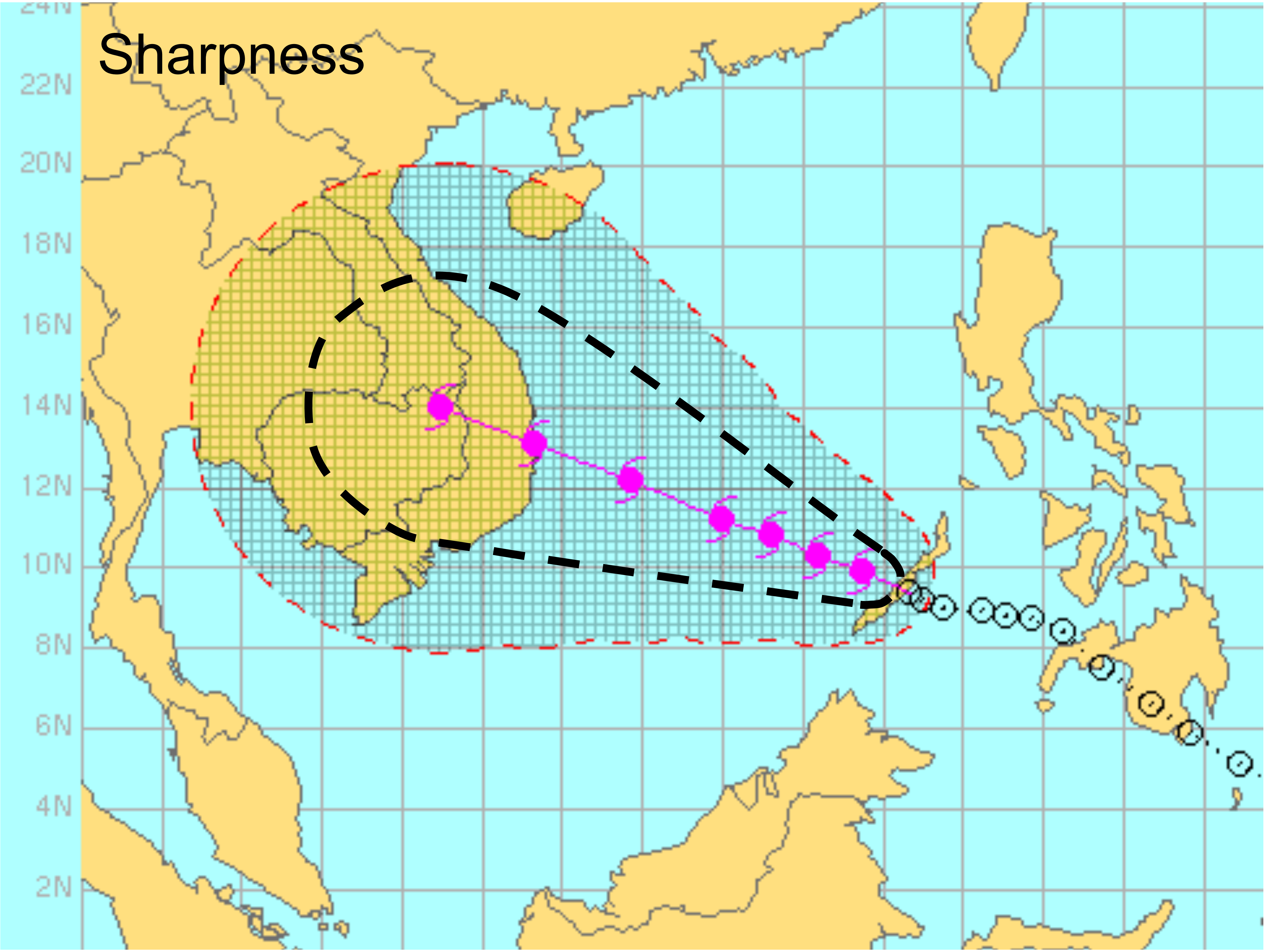
100*AOD

Measures of probabilistic quality

- Sharpness
  - Probabilistic forecasts are different from climatology.
- Reliability
  - 30% probability events happen 30% of the time.
- Goal is to produce as sharp a probabilistic forecast as possible subject to the constraint of reliability.

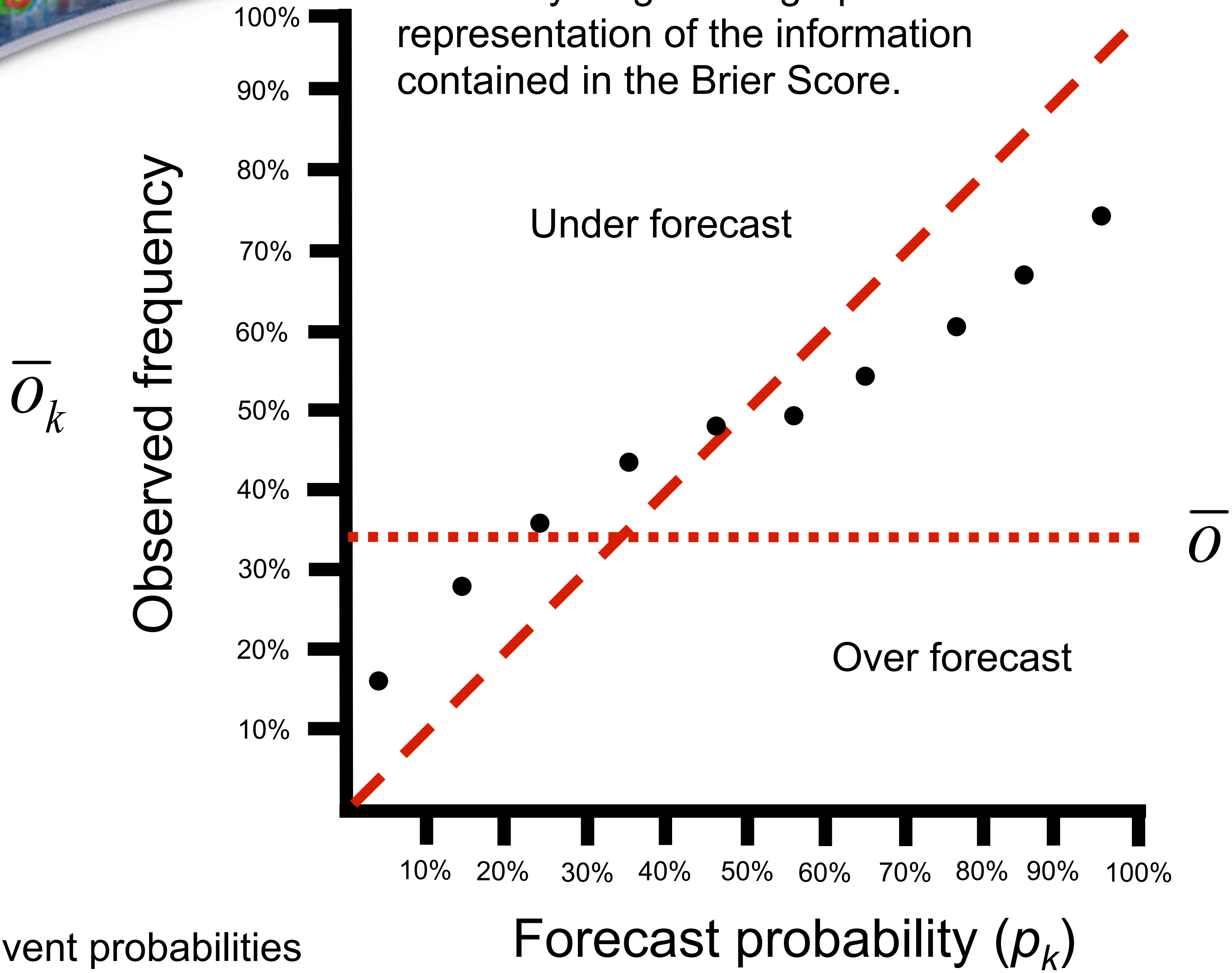

Sharpness
50
55
60
65
70
75
temperature
Blue is sharp relative to red

100*AOD



Sharpness
24N
22N
20N
18N
16N
14N
12N
10N
8N
6N
4N
2N

Reliability diagram: A graphical representation of the information contained in the Brier Score.



Event probabilities

Post processing

What’s in a name

- “Post-processing” has many connotations:
  - Statistical correction of a forecast
  - Generation of diagnostic quantities (e.g. AOT)
  - Processing that is done on transmitted data to ships and shore facilities (this has caused me no end of grief)

Note: you could post-process a message containing a post-processed quantity from post-processed fields!

Post-processing

SuperEnsembles

Analogs

Downscaling

Calibration

Rank histogram

Logistic regression

Kalman filter

Bias correction

Bayesian processor of forecasts

Non-homogeneous Gaussian regression

Dressing

Linear regression

Kernel density

CDF

Bayesian model averaging

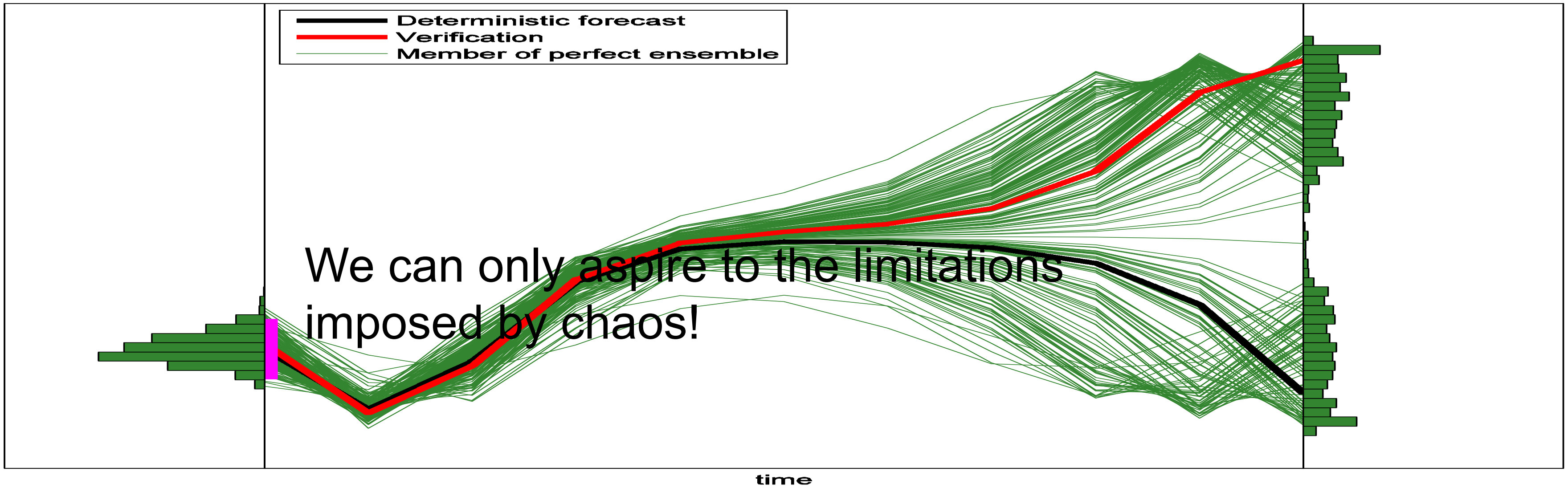
Suggested vocabulary

- Calibration instead of post-processing
  - **Adjusting** the output of a model to agree with the value of the **applied standard** (the observation), within a specified accuracy.
- Adjusting
  - Methods for calibrating the mean (or deterministic forecast)
    - Bias correction, Kalman filter, linear regression, etc.
  - Methods for calibrating the distribution
    - Analogs, BMA, rank histogram, etc.
- Applied standard
  - Analyses
  - Observations

Old: “Downscaling”
New: “Our Kalman filter technique is a method for calibrating the ensemble mean using observations as the applied standard”

Ensembles aren’t a silver bullet

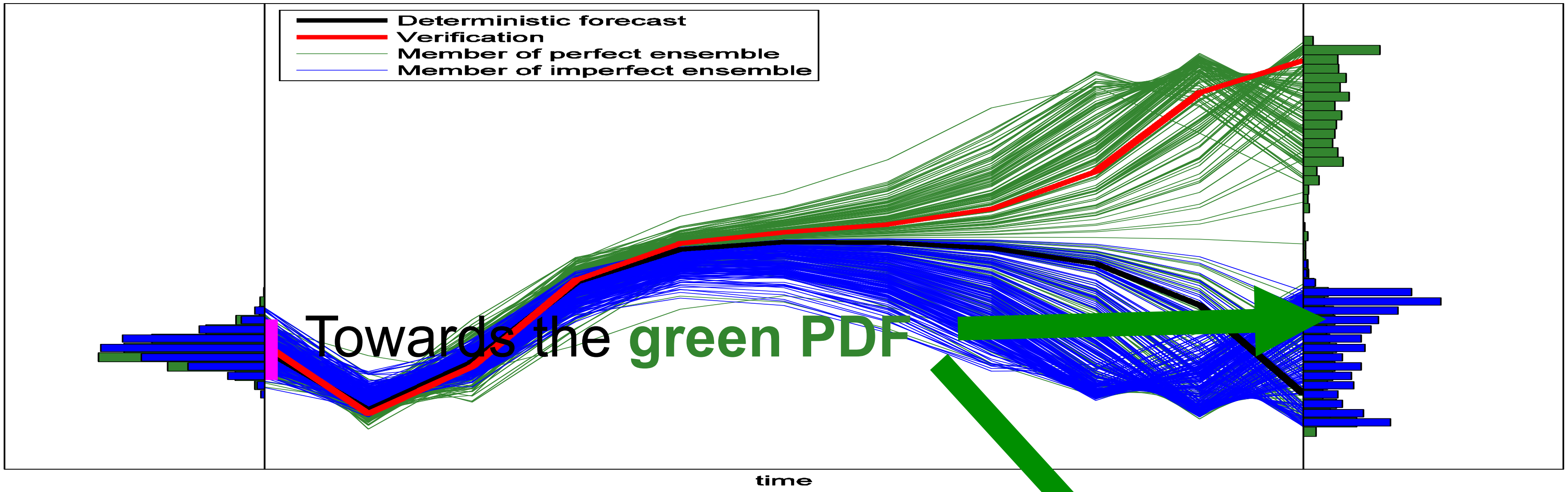
- Running a crap model 100 times doesn’t make it any less crap
  - An ensemble is only as good as the model(s) that go into it
- Multi-model ensemble forecasts cannot give you correct probabilistic forecasts (in fact, we probably shouldn’t even call them probability forecasts!)
  - But they can provide better probabilistic forecasts than single model ensembles



We can only aspire to the limitations imposed by chaos!

t=0

t=3days

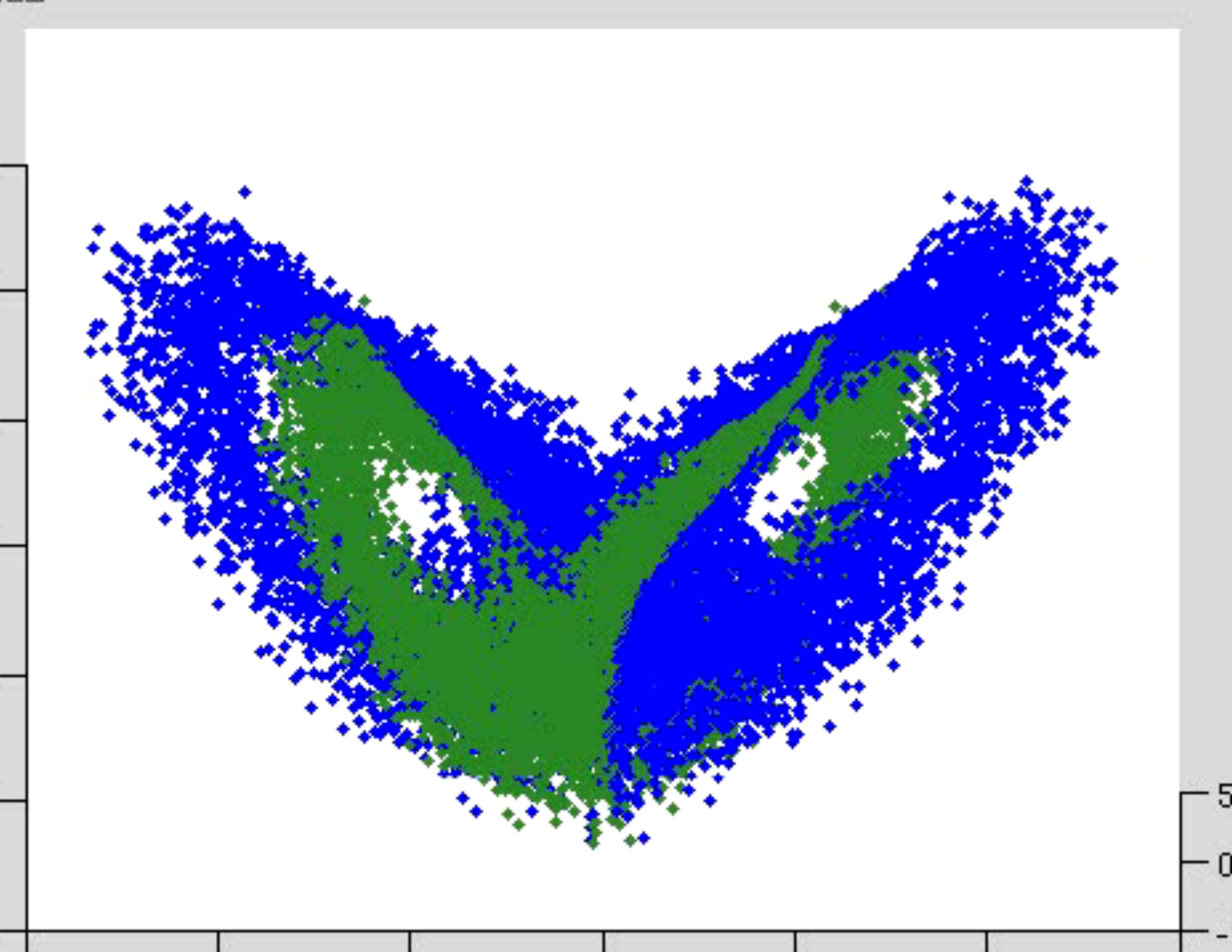


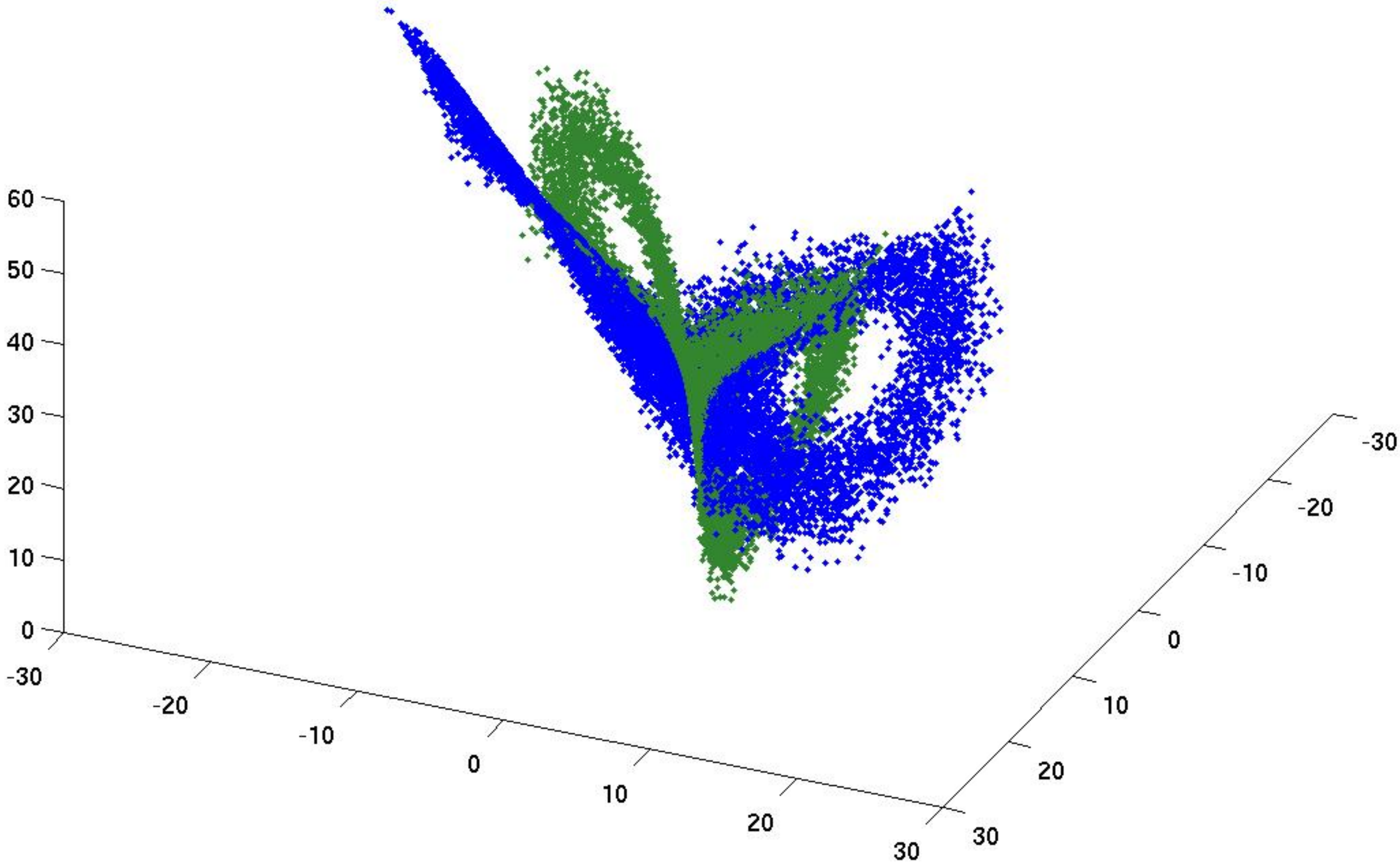
Towards the **green PDF**

Goal is to move the **blue PDF**

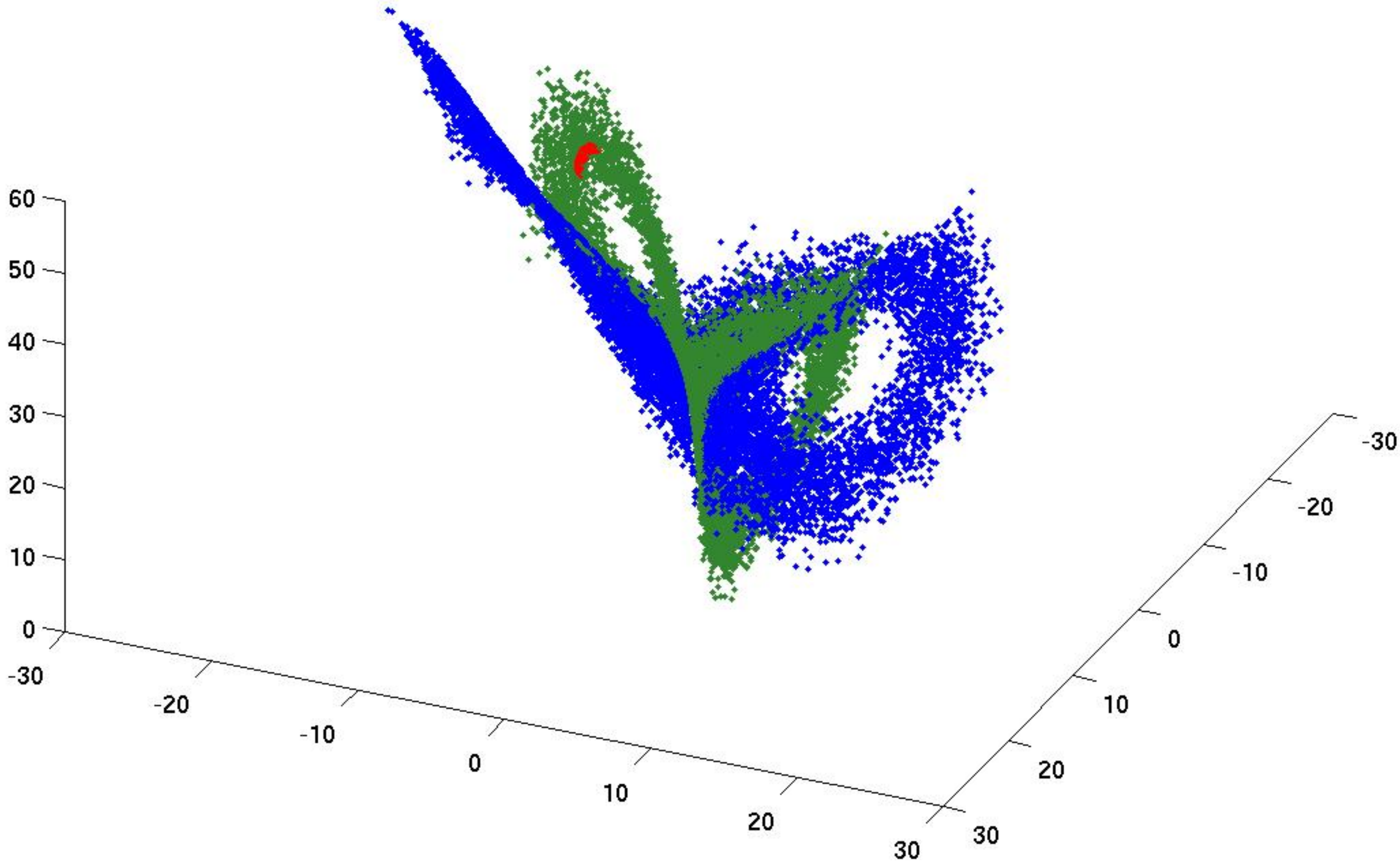
t=0

t=3days

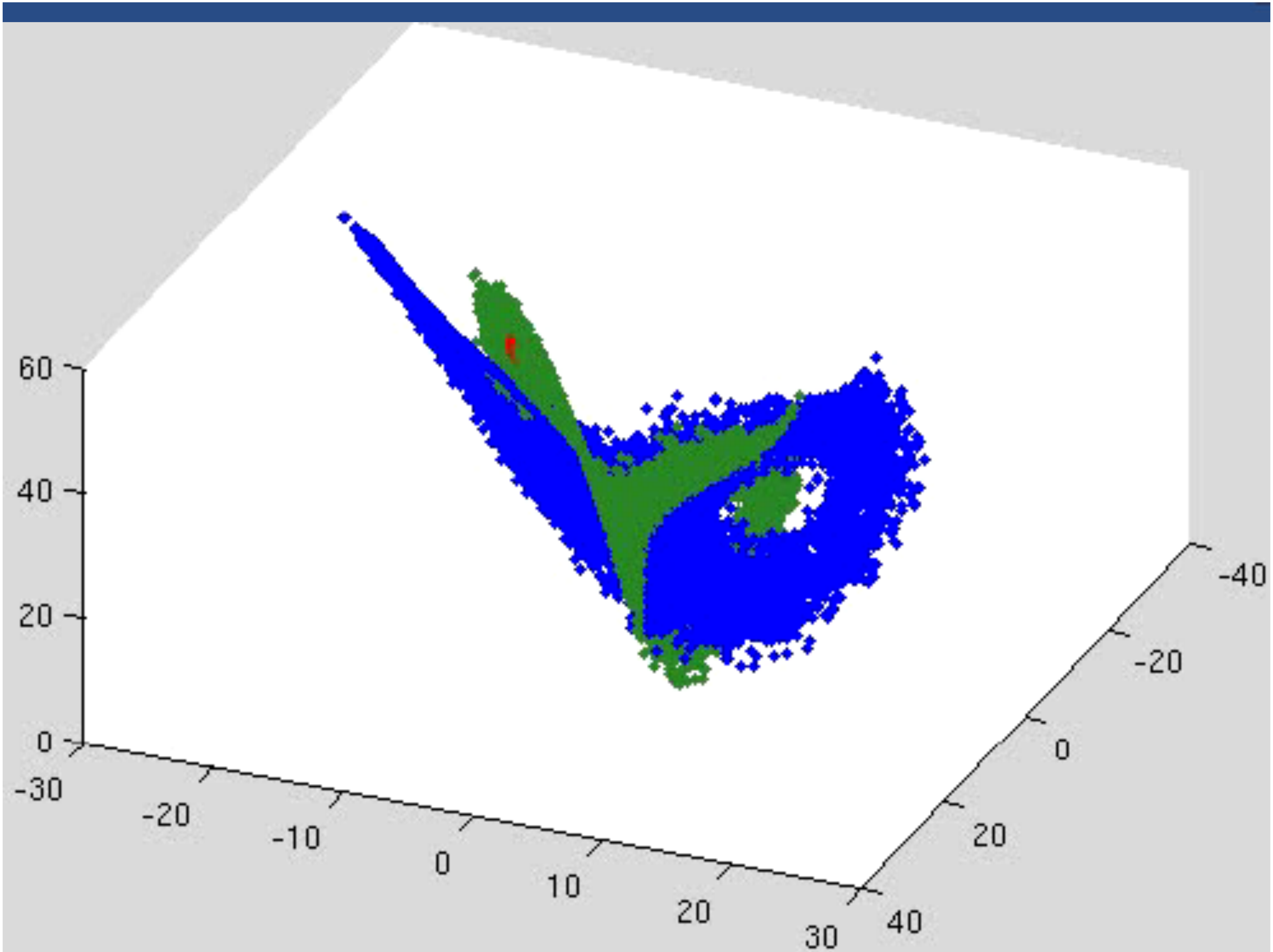




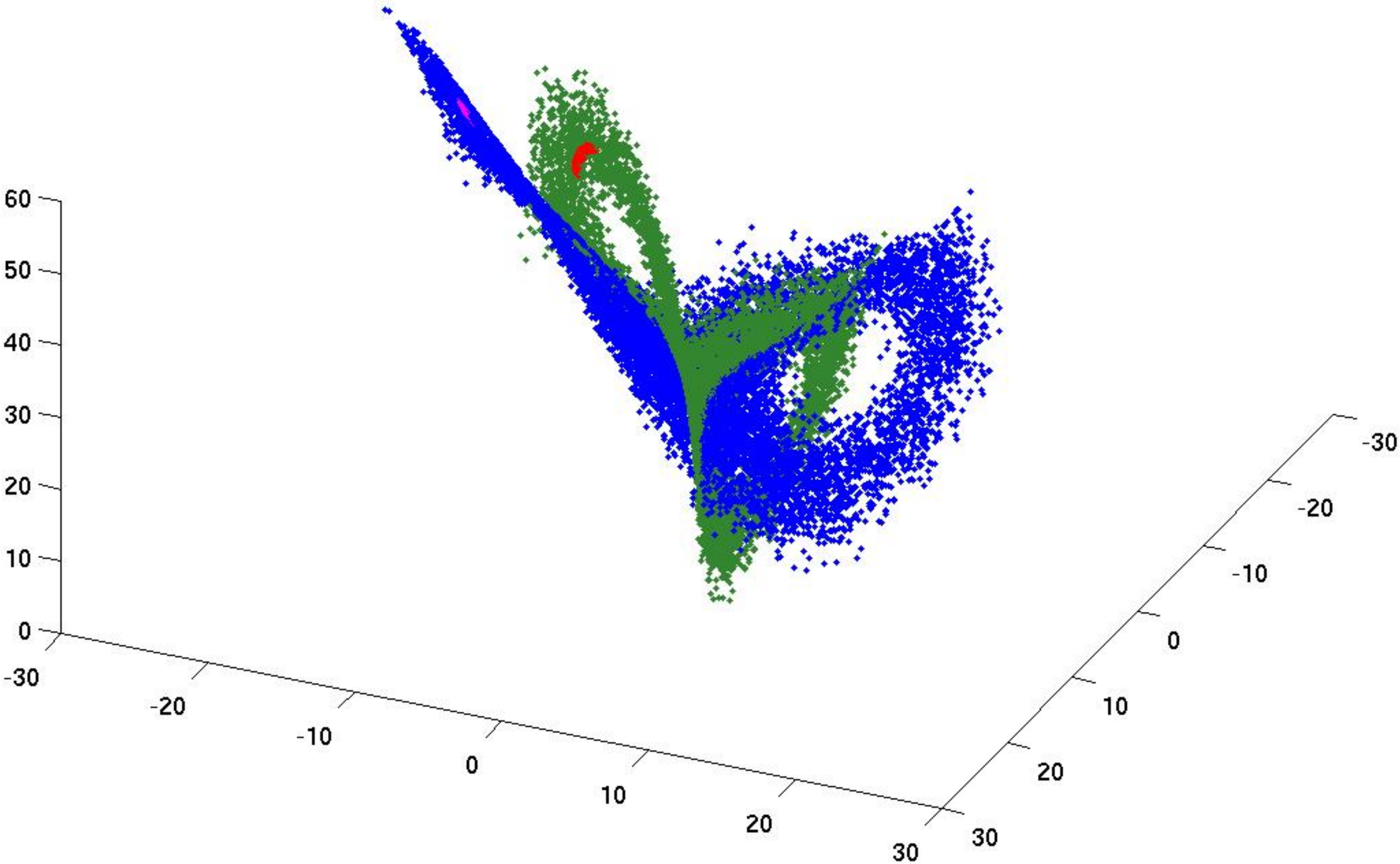
60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
-30
-20
-10
0
10
20
30



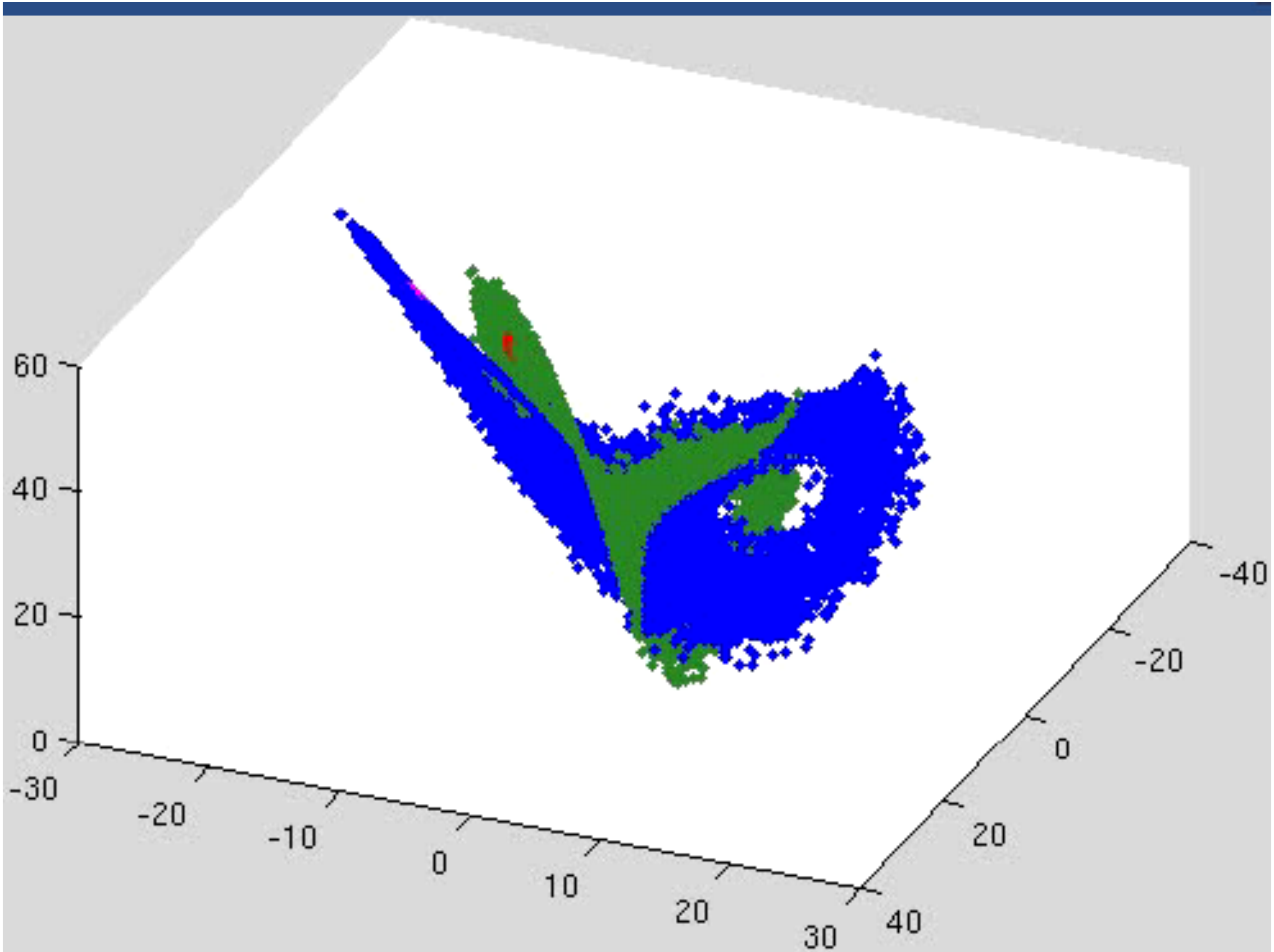
60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
-30
-20
-10
0
10
20
30

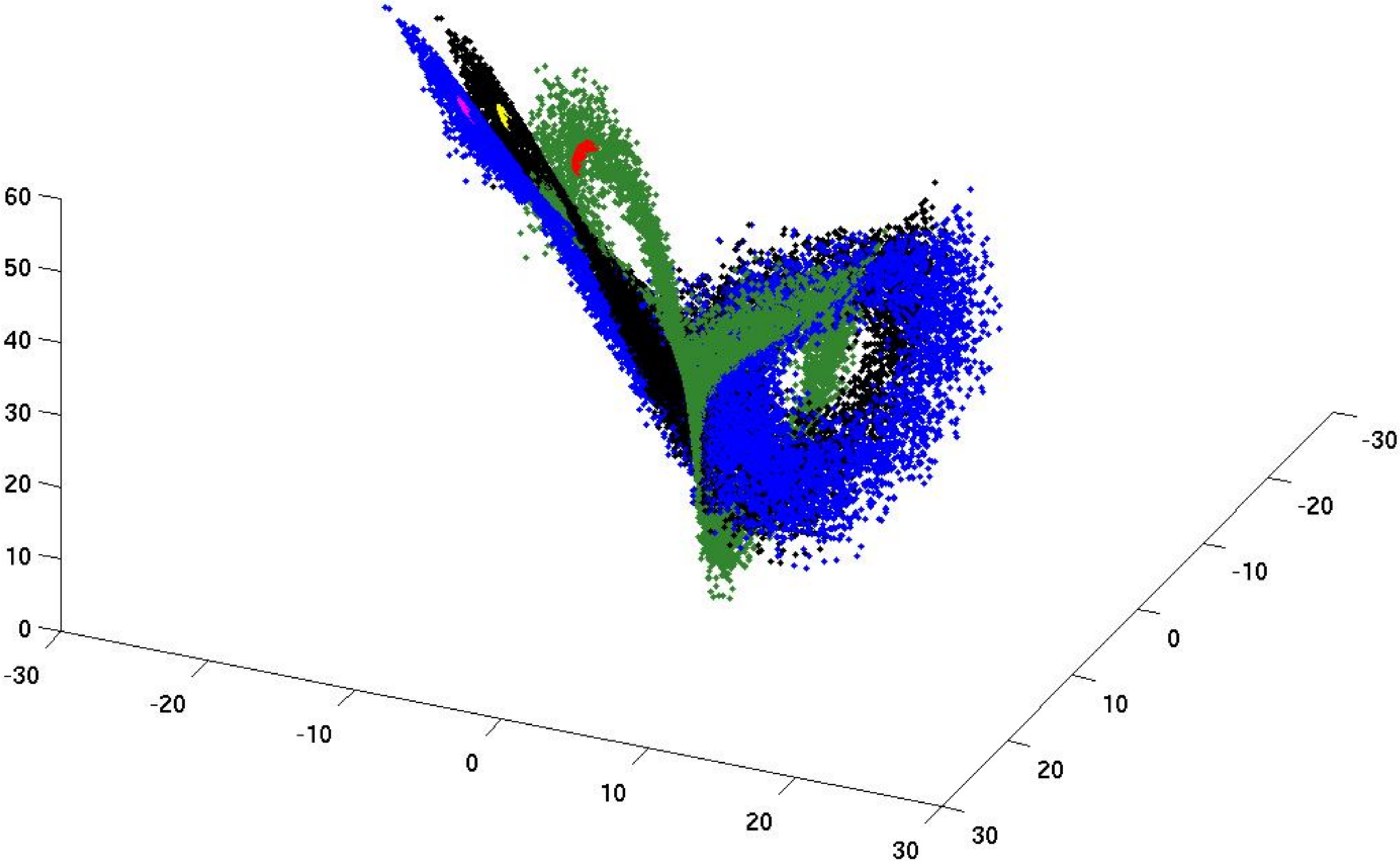


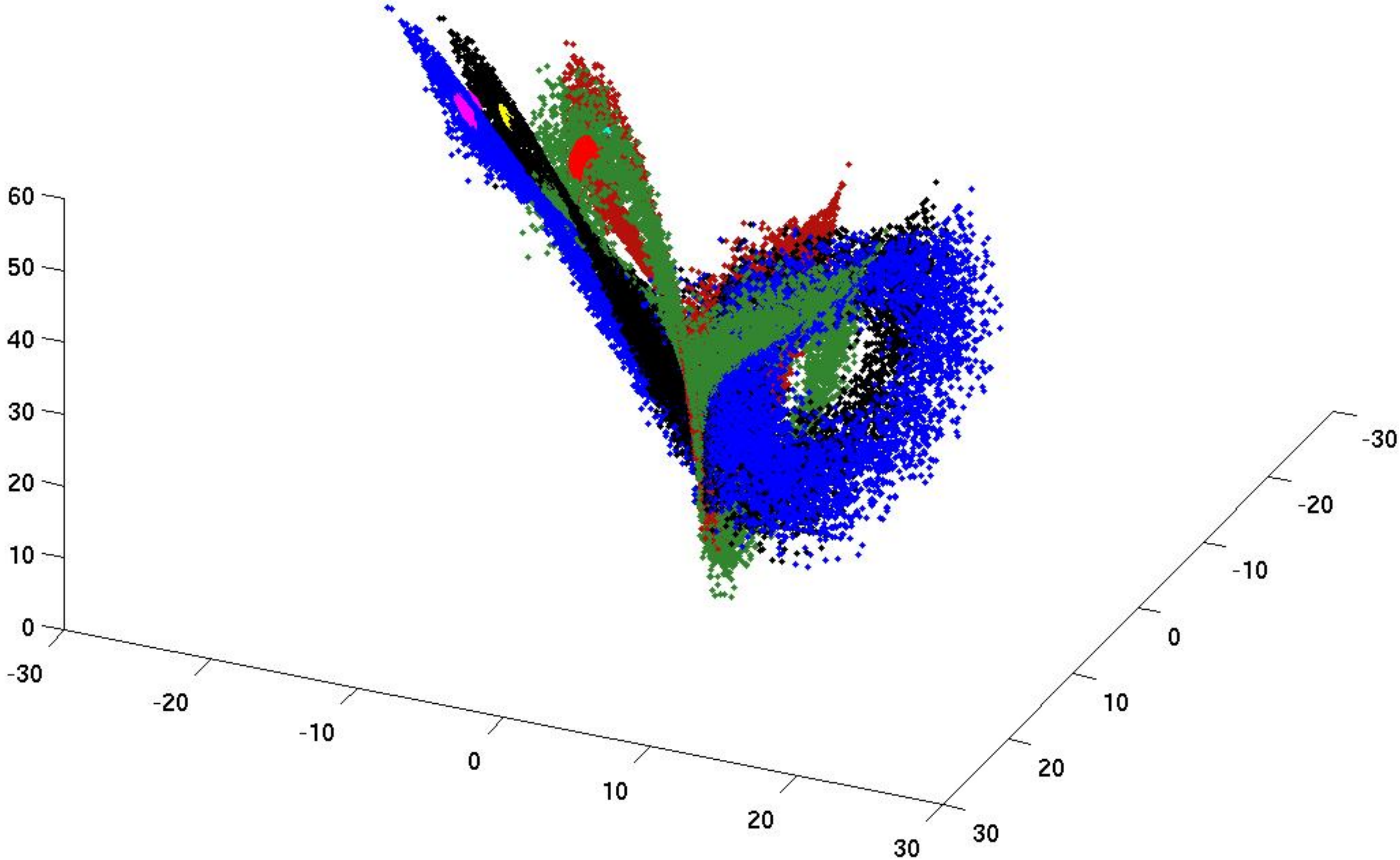
60
40
20
0
-30
-20
-10
0
10
20
30
40
20
0
-20
-40

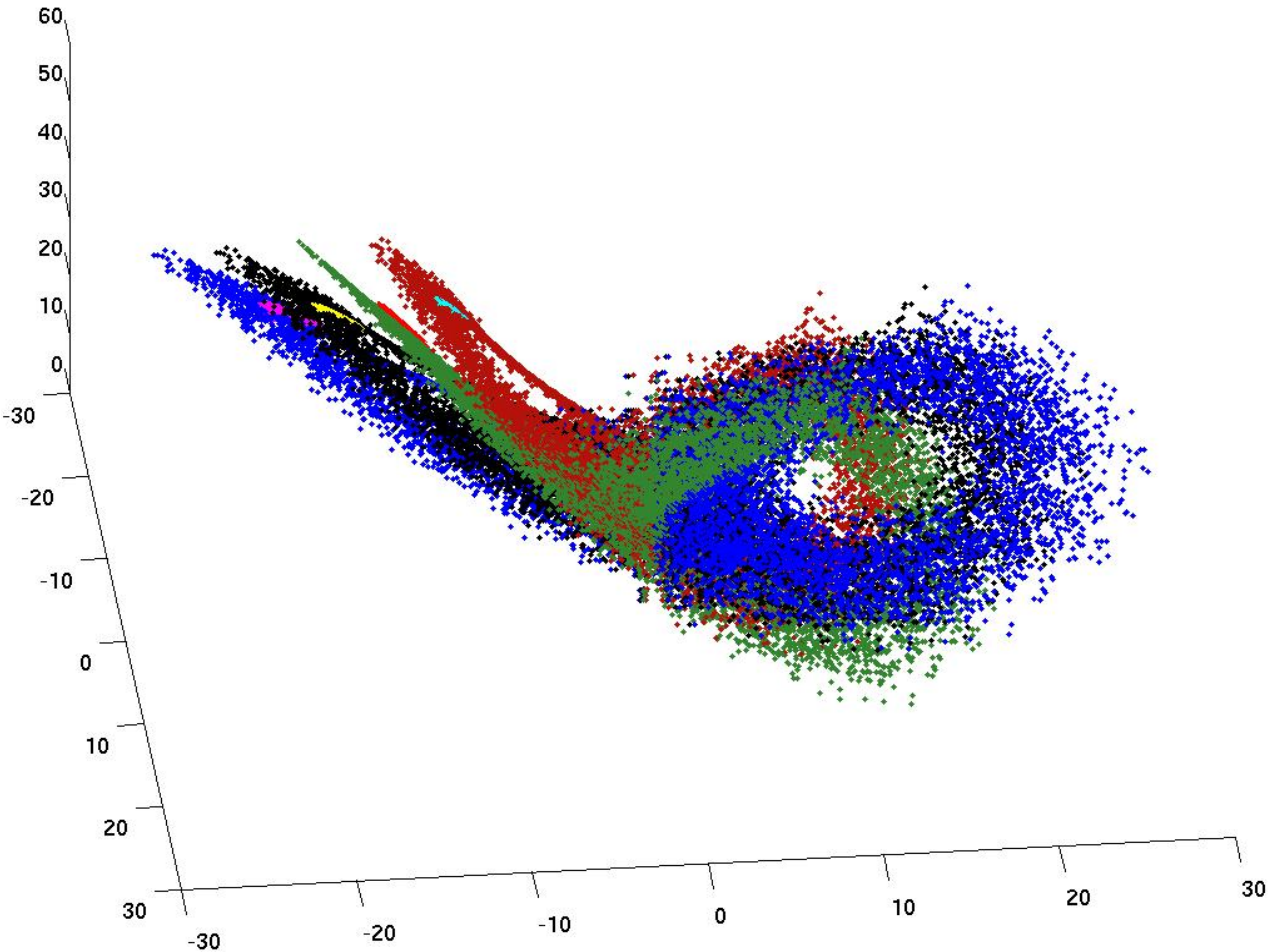


60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
-30
-20
-10
0
10
20
30


60
40
20
0
-30
-20
-10
0
10
20
30
-40
-20
0
20
40


60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
30
20
10
0
-10
-20
-30


60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
-30
-20
-10
0
10
20
30


60
50
40
30
20
10
0
-30
-20
-10
0
10
20
30
-30
-20
-10
0
10
20
30

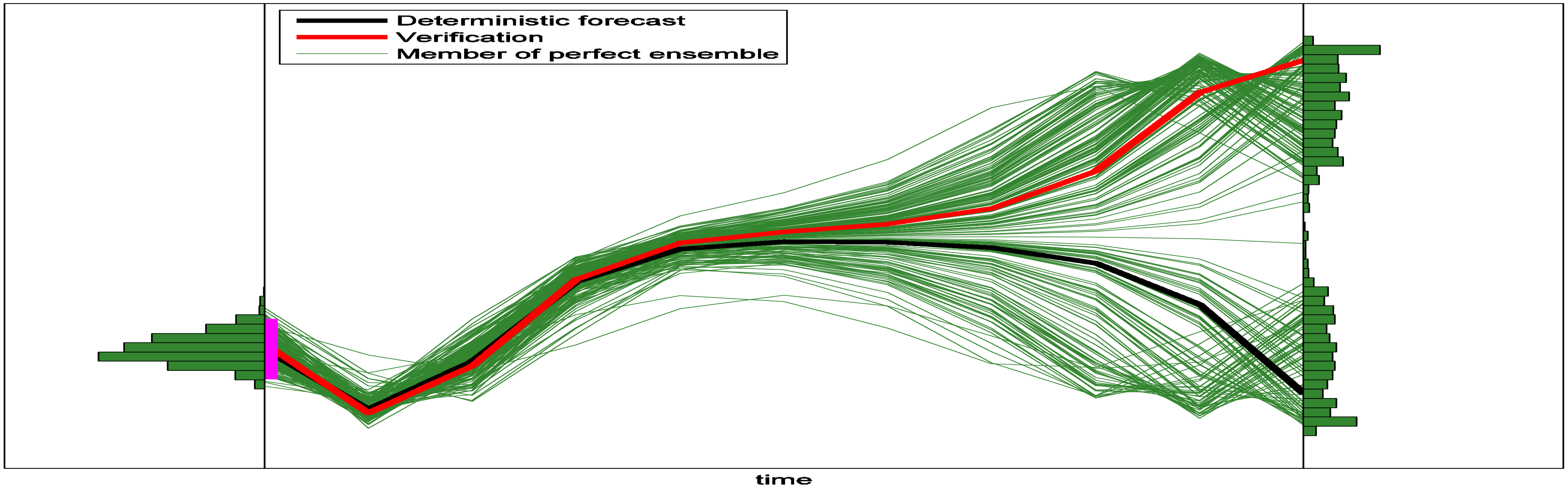
Impact of model inadequacy

- We can only aspire to the limitations imposed by chaos.
- In the same way that initial condition uncertainty guarantees we will never have perfect deterministic forecasts, model uncertainty guarantees we will never have perfect probabilistic forecasts.
- In the same way that deterministic forecasts in the face of initial condition uncertainty are still useful, so too are “probabilistic” (distribution? odds?) forecasts in the face of model uncertainty.

Beyond error bars

- We want to leverage the ensemble to
  - Improve our initial conditions
  - Improve our understanding of dynamics
  - Help hedge forecasts
  - Target observations

- This information is obtained by exploiting the state-dependent, space-time statistical relationships within the ensemble forecast.


Deterministic forecast
Verification
Member of perfect ensemble
time

t=0

t=3days

Ensemble sensitivity

Relationship between different variables at different times

$$\bar{\mathbf{y}}(\tau_u) = f(\bar{\mathbf{x}}(\tau_s))$$

Add perturbations

$$\bar{\mathbf{y}} + \mathbf{y}' = f(\bar{\mathbf{x}} + \mathbf{x}')$$

Taylor expansion

$$f(\bar{\mathbf{x}} + \mathbf{x}') = f(\bar{\mathbf{x}}) + \frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}}\mathbf{x}'$$

Substitute

$$\bar{\mathbf{y}} + \mathbf{y}' = f(\bar{\mathbf{x}}) + \frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}}\mathbf{x}'$$

Cancel terms

$$\mathbf{y}' = \frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}}\mathbf{x}'$$

Multiply both sides by $\mathbf{x}'^T$

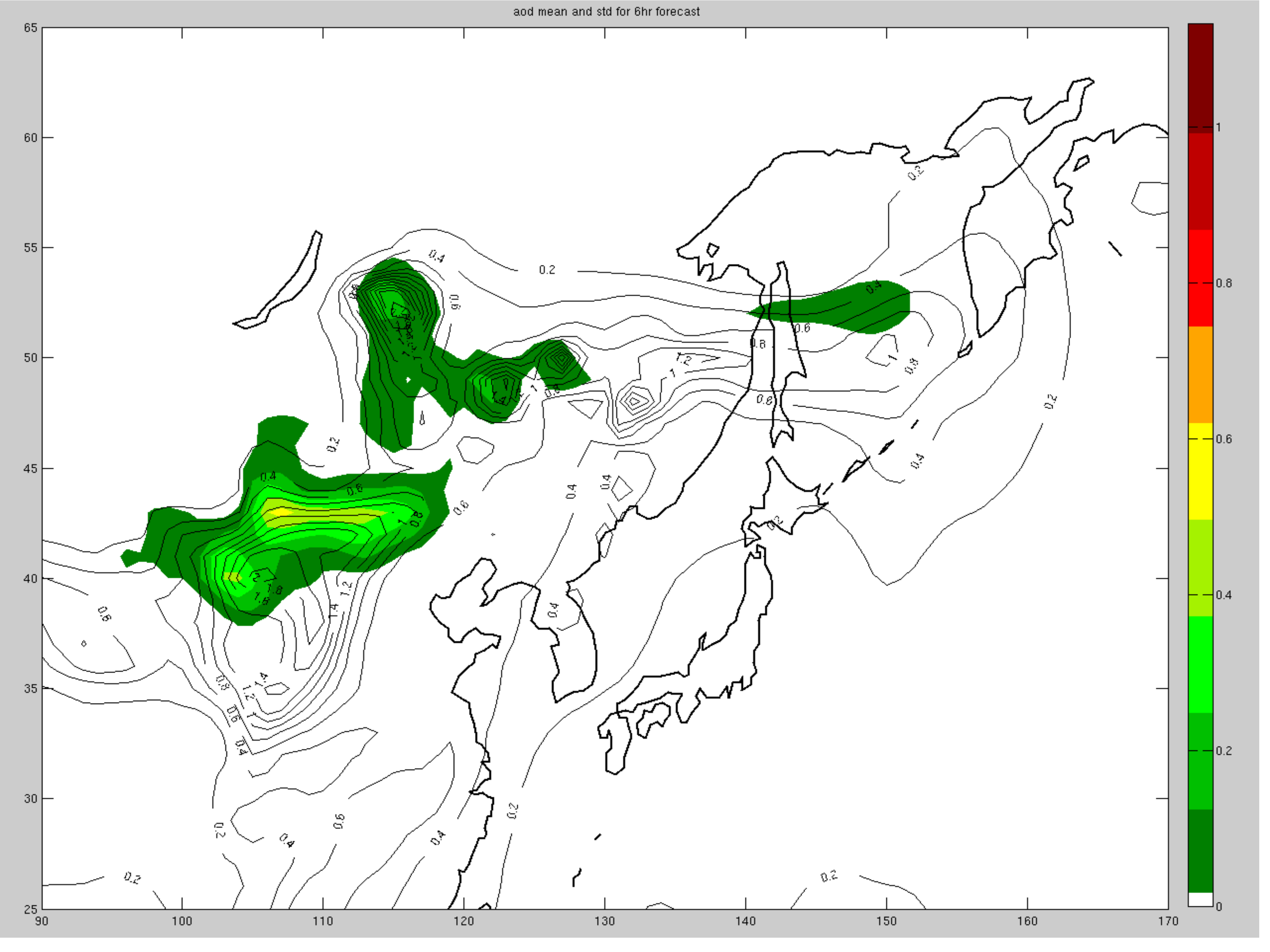
$$\mathbf{y}'\mathbf{x}'^T = \frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}}\mathbf{x}'\mathbf{x}'^T$$

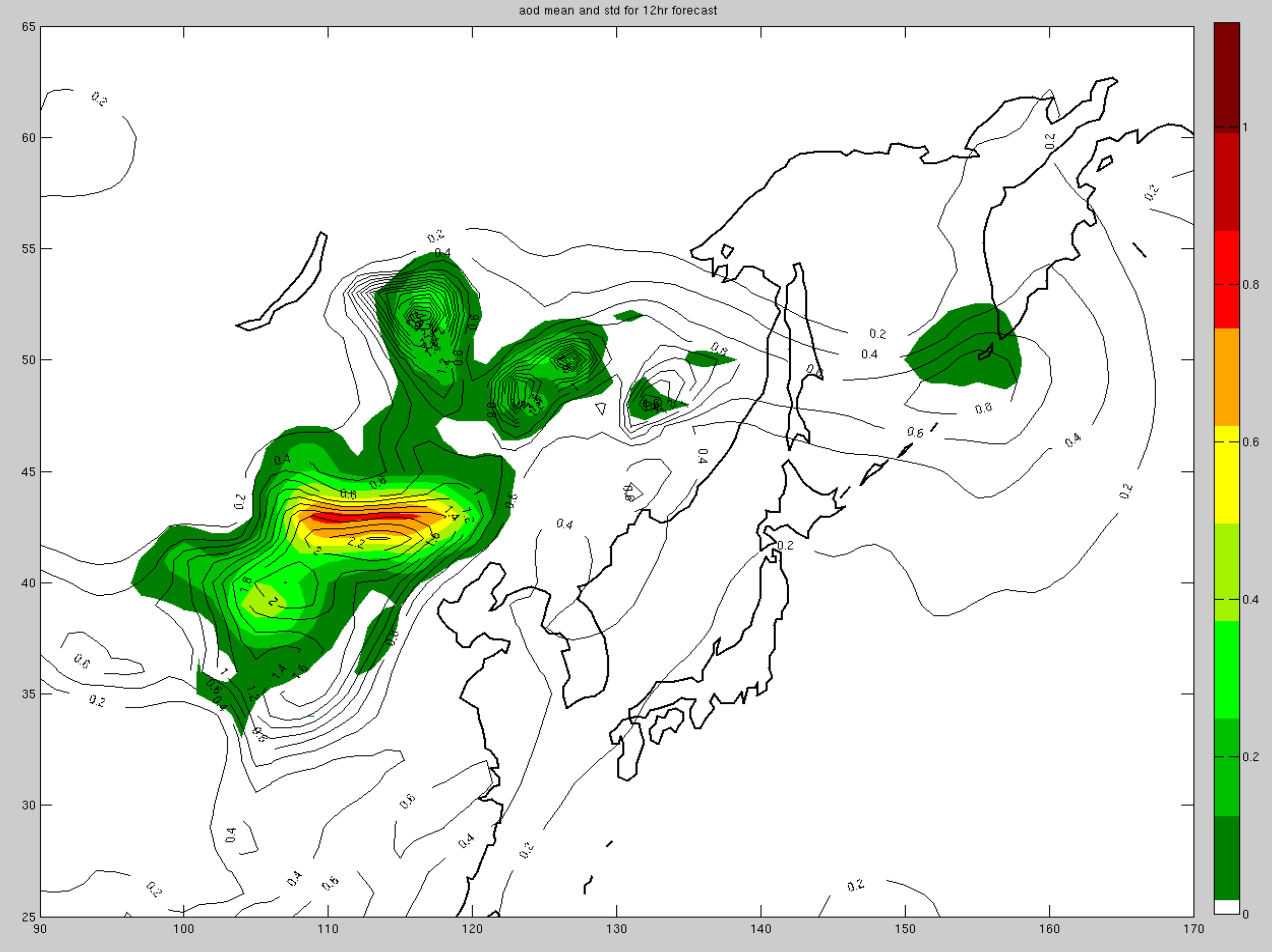
Apply an expectation operator

$$\mathrm{cov}(\mathbf{x},\mathbf{y}) = \frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}}\mathrm{cov}(\mathbf{x})$$

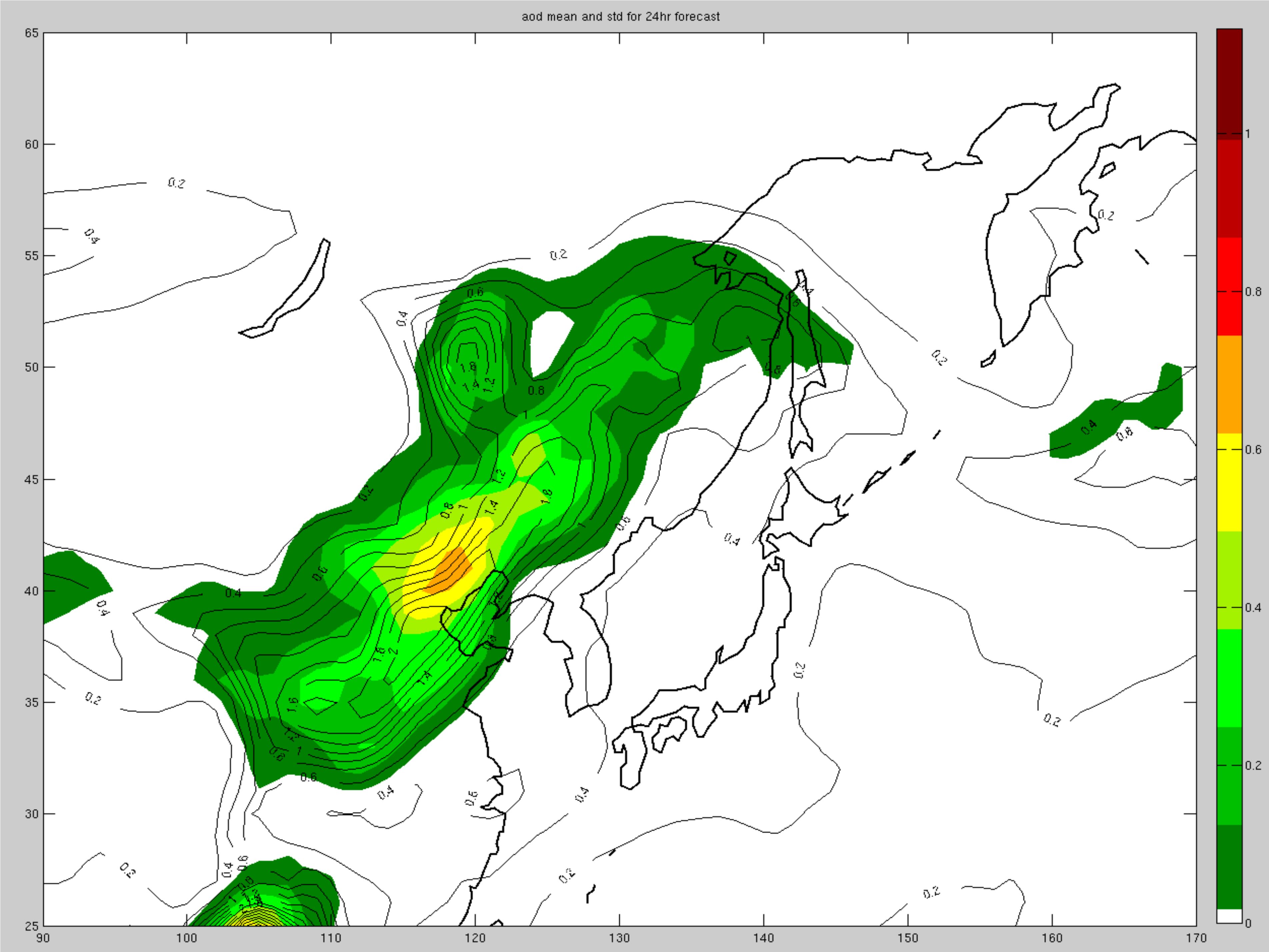
Rearrange

$$\frac{\partial f(\bar{\mathbf{x}})}{\partial \mathbf{x}} = \mathrm{cov}(\mathbf{x}(\tau_s),\mathbf{y}(\tau_u))\,\mathrm{cov}(\mathbf{x}(\tau_s))^{-1}$$

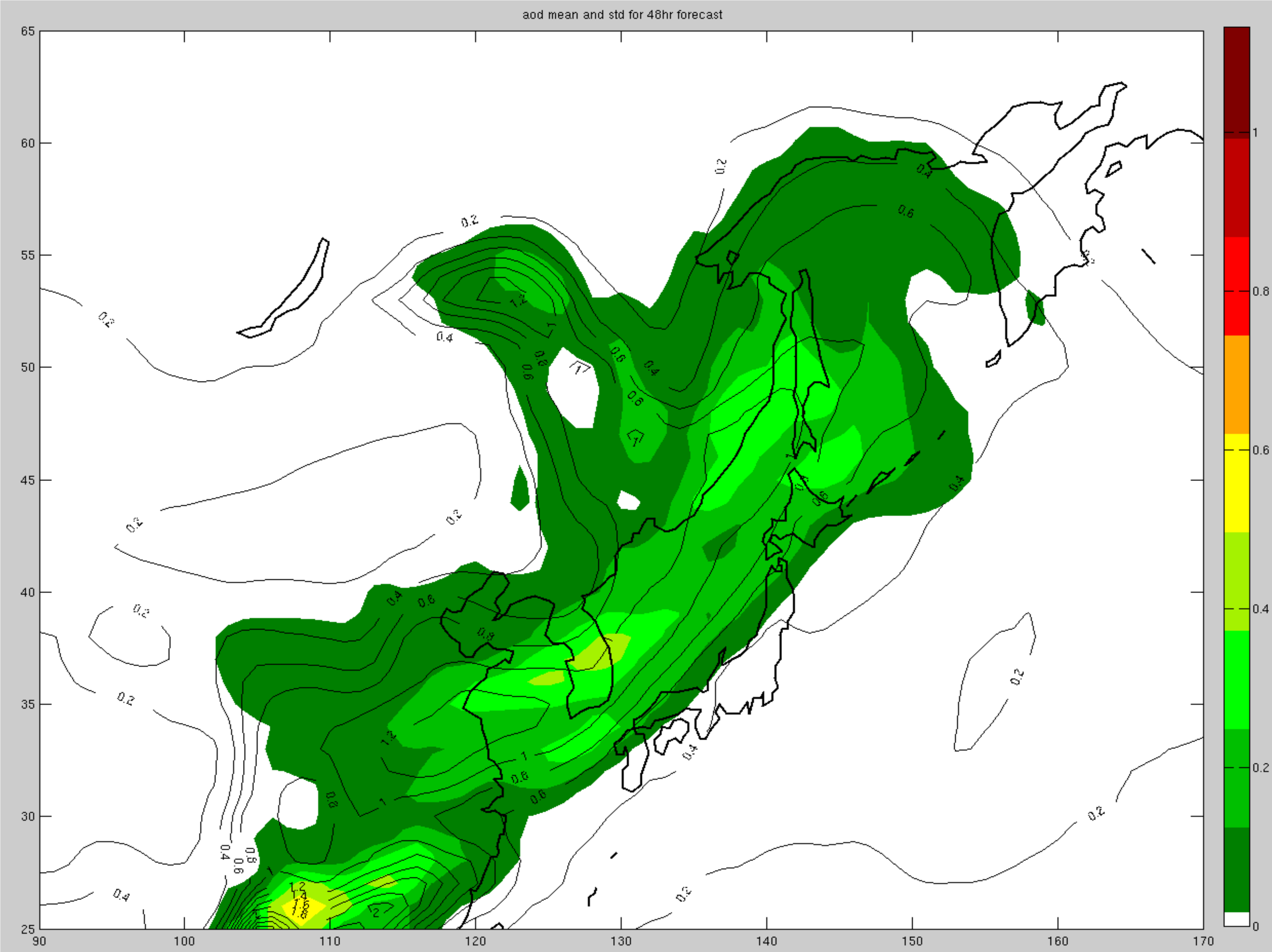

aod mean and std for 6hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0



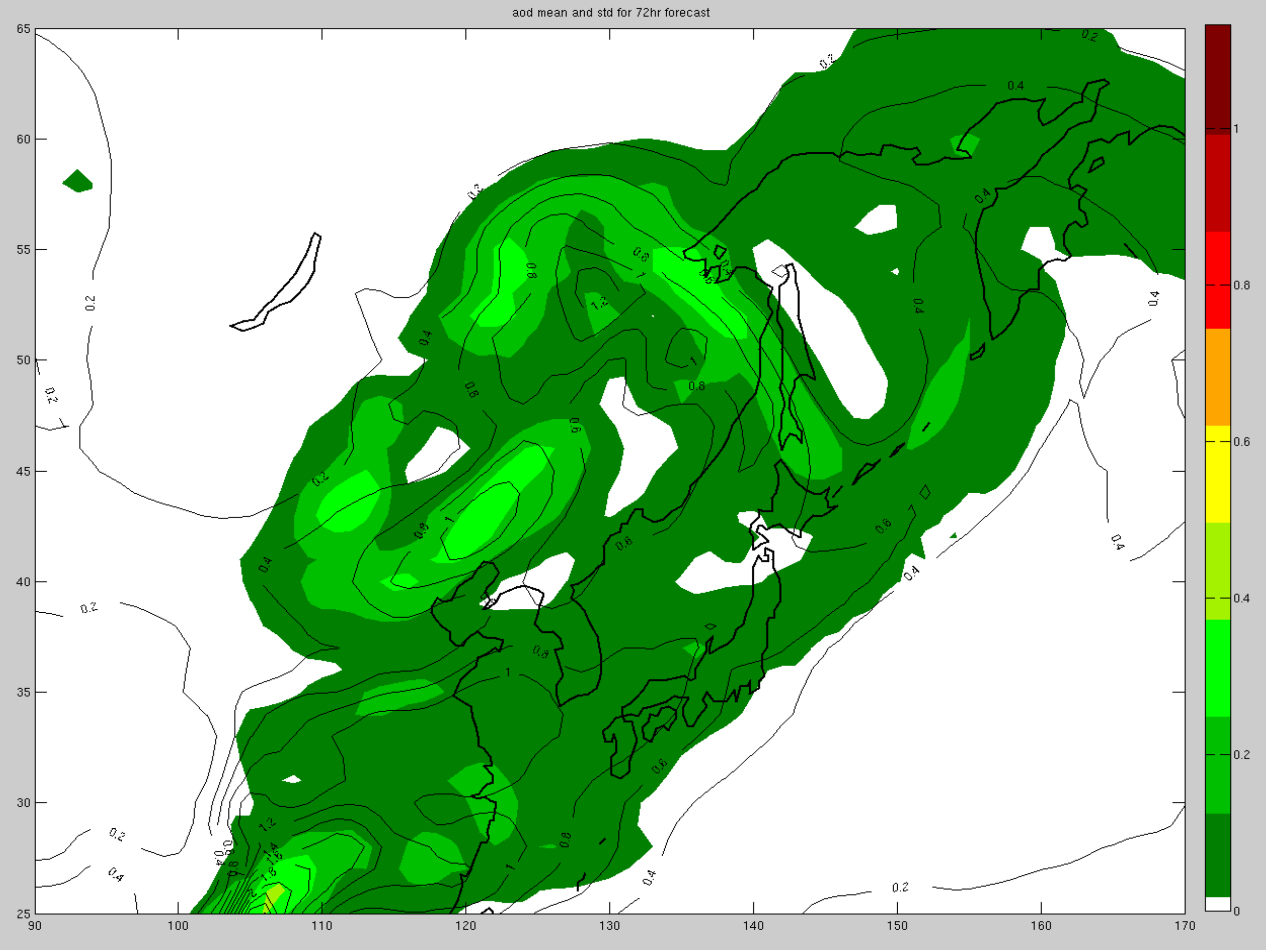
aod mean and std for 12hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0

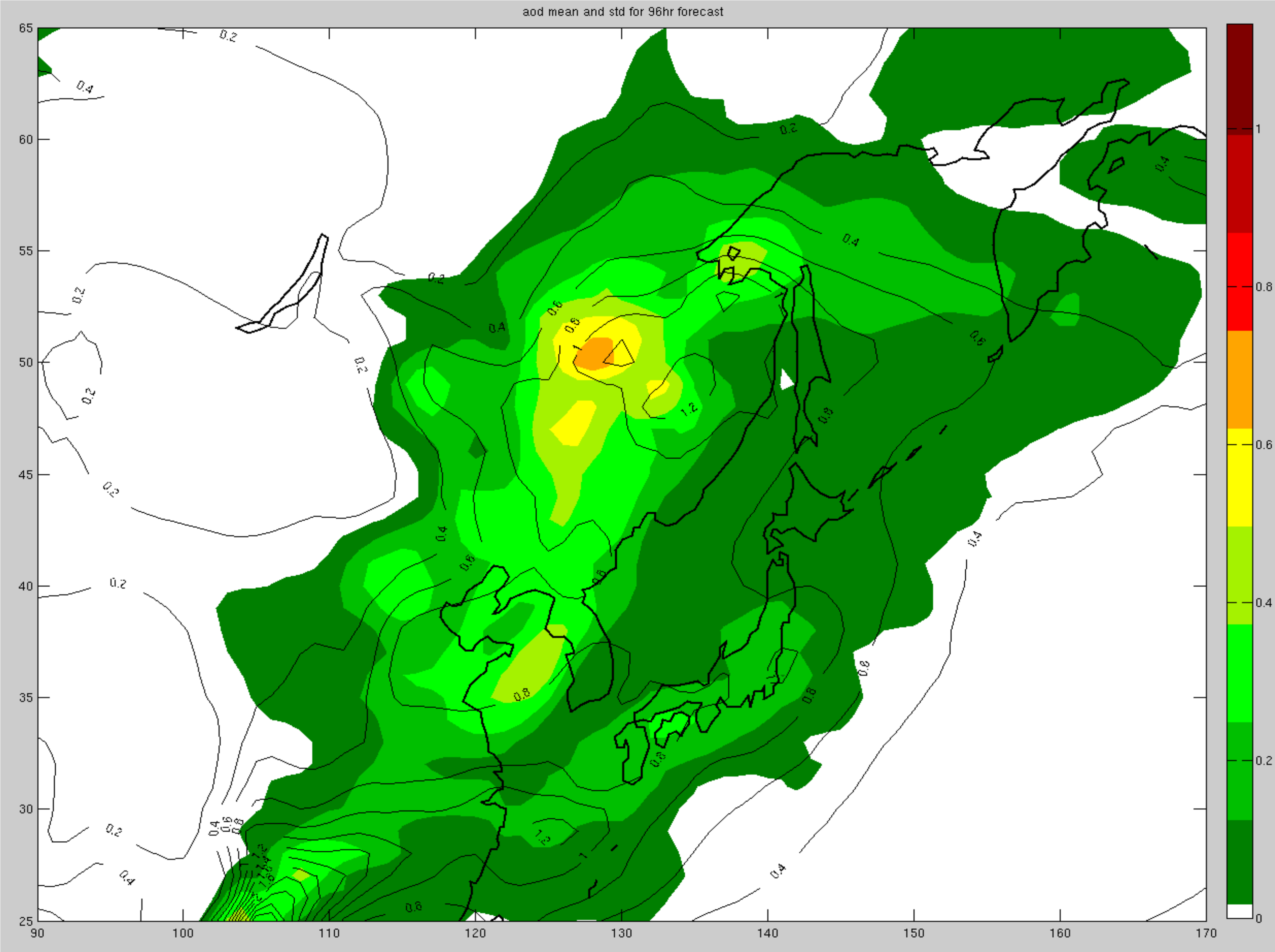


aod mean and std for 24hr forecast

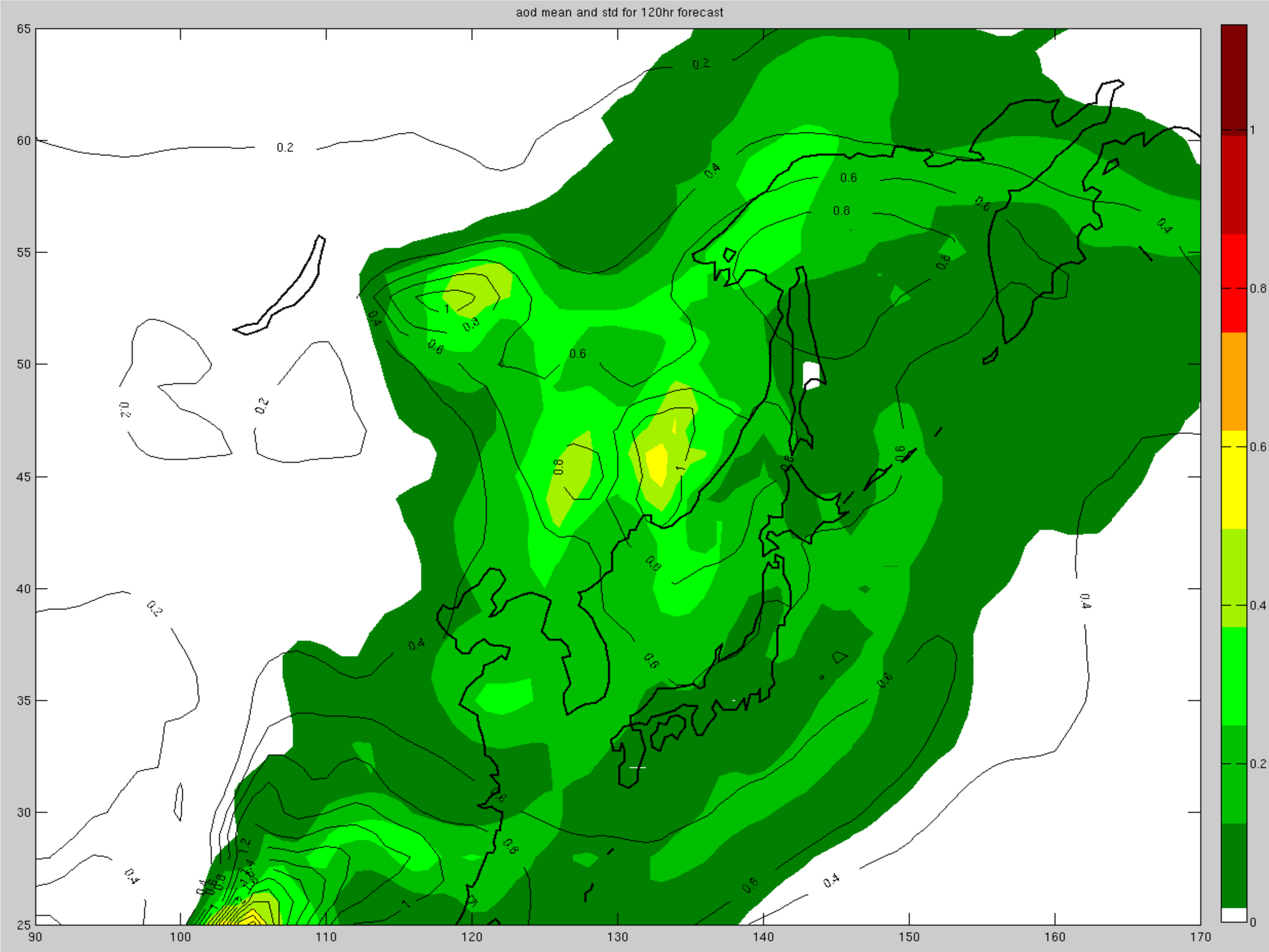


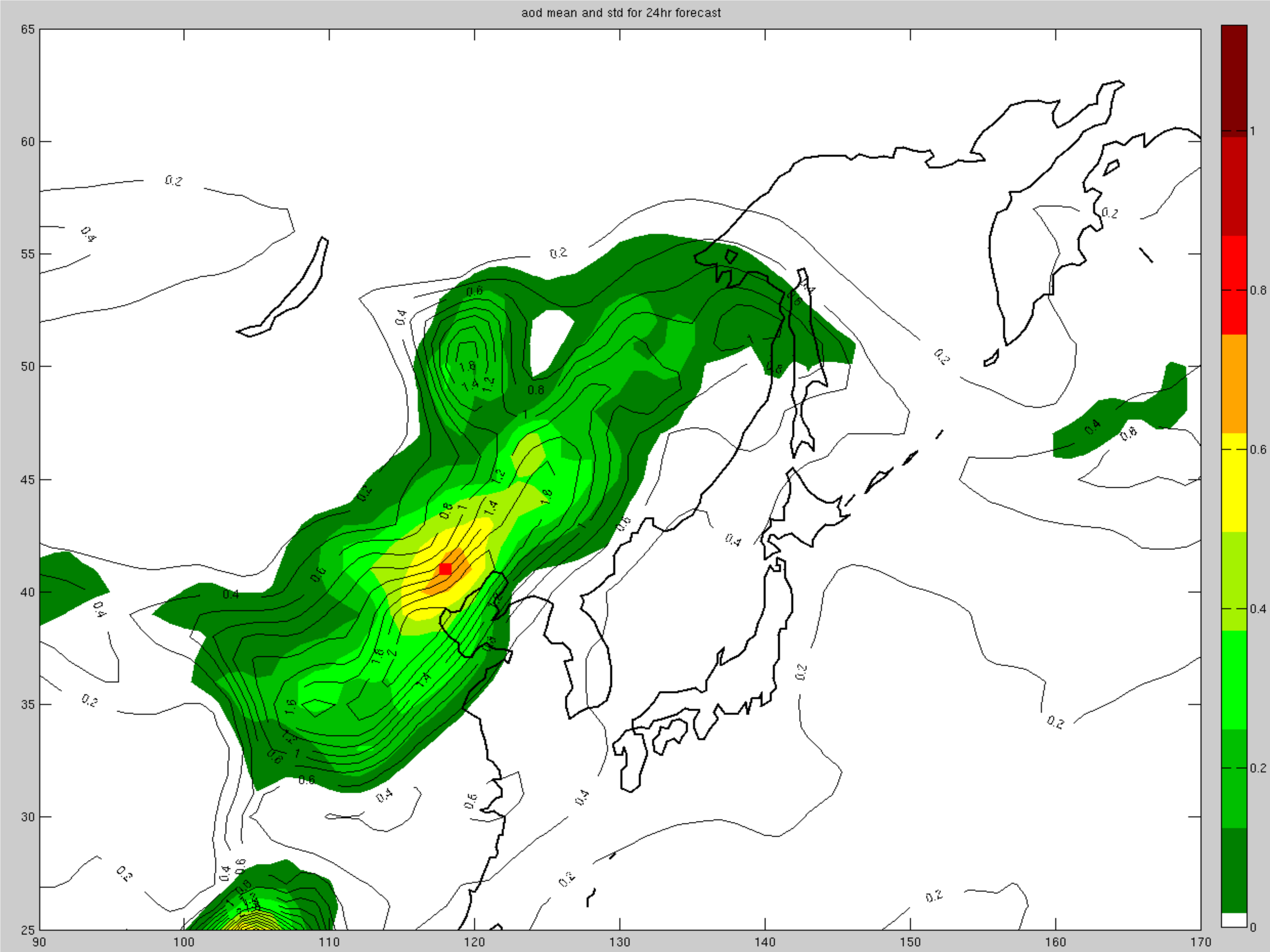
aod mean and std for 48hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0

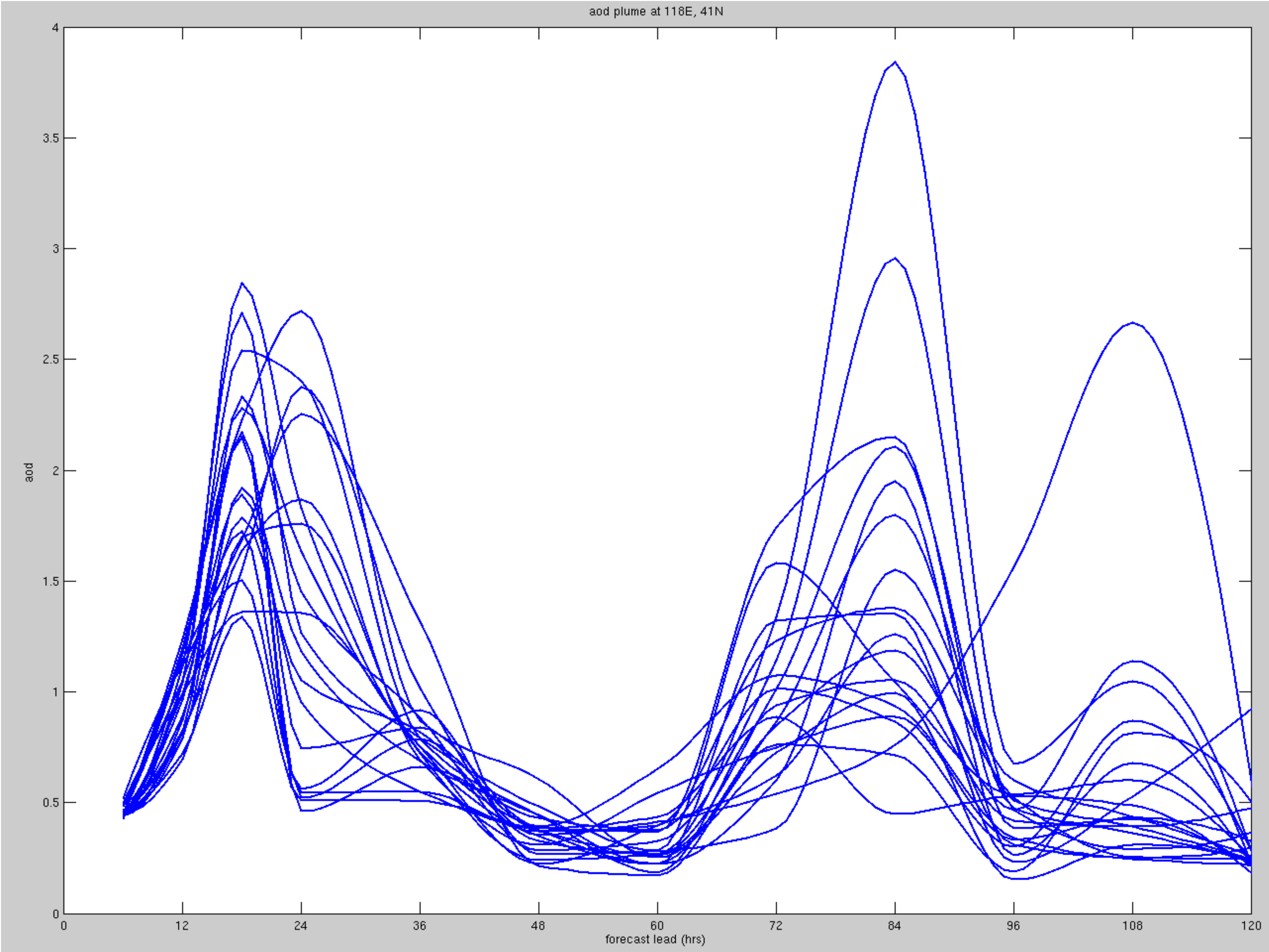

aod mean and std for 72hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0

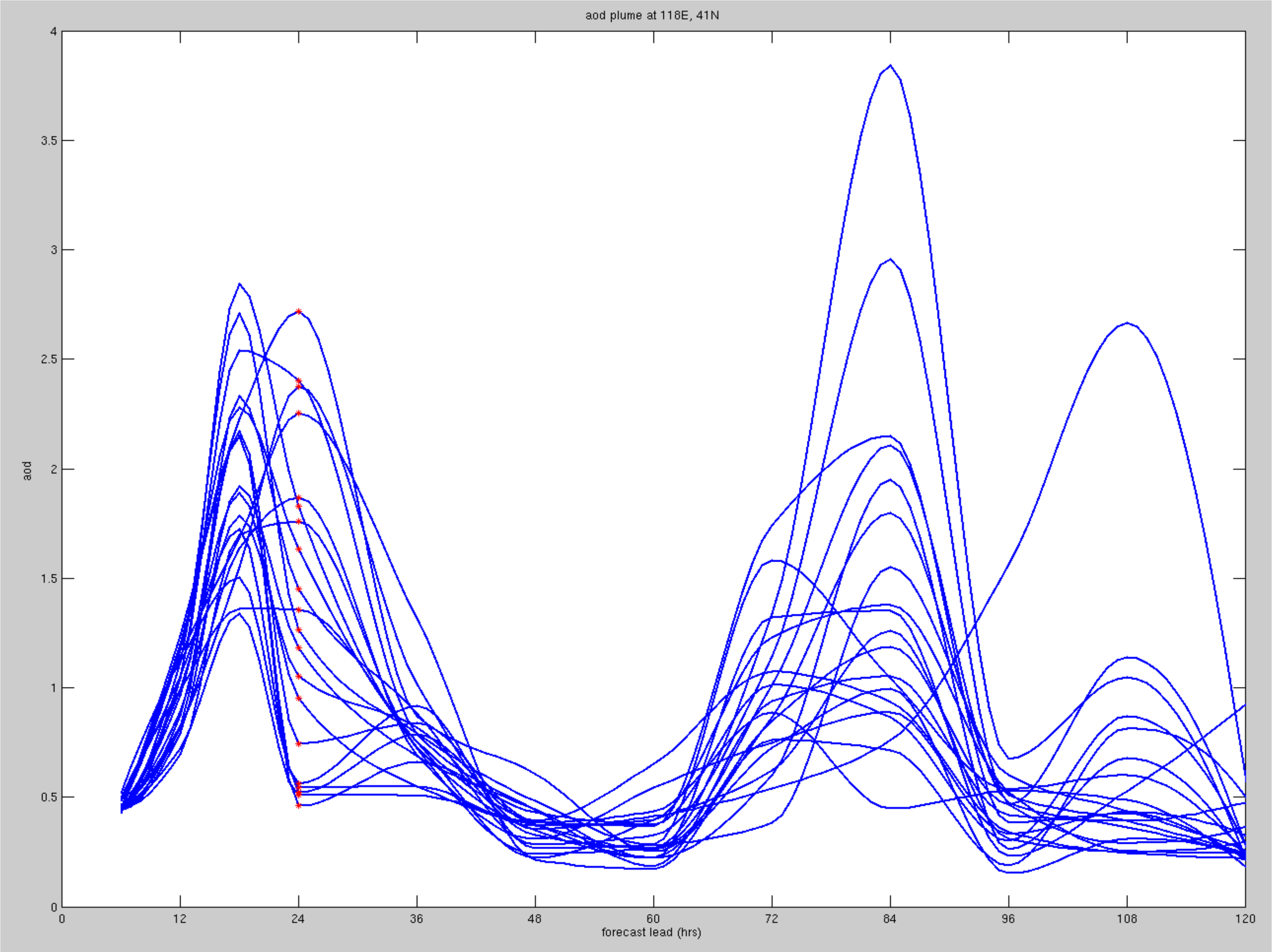


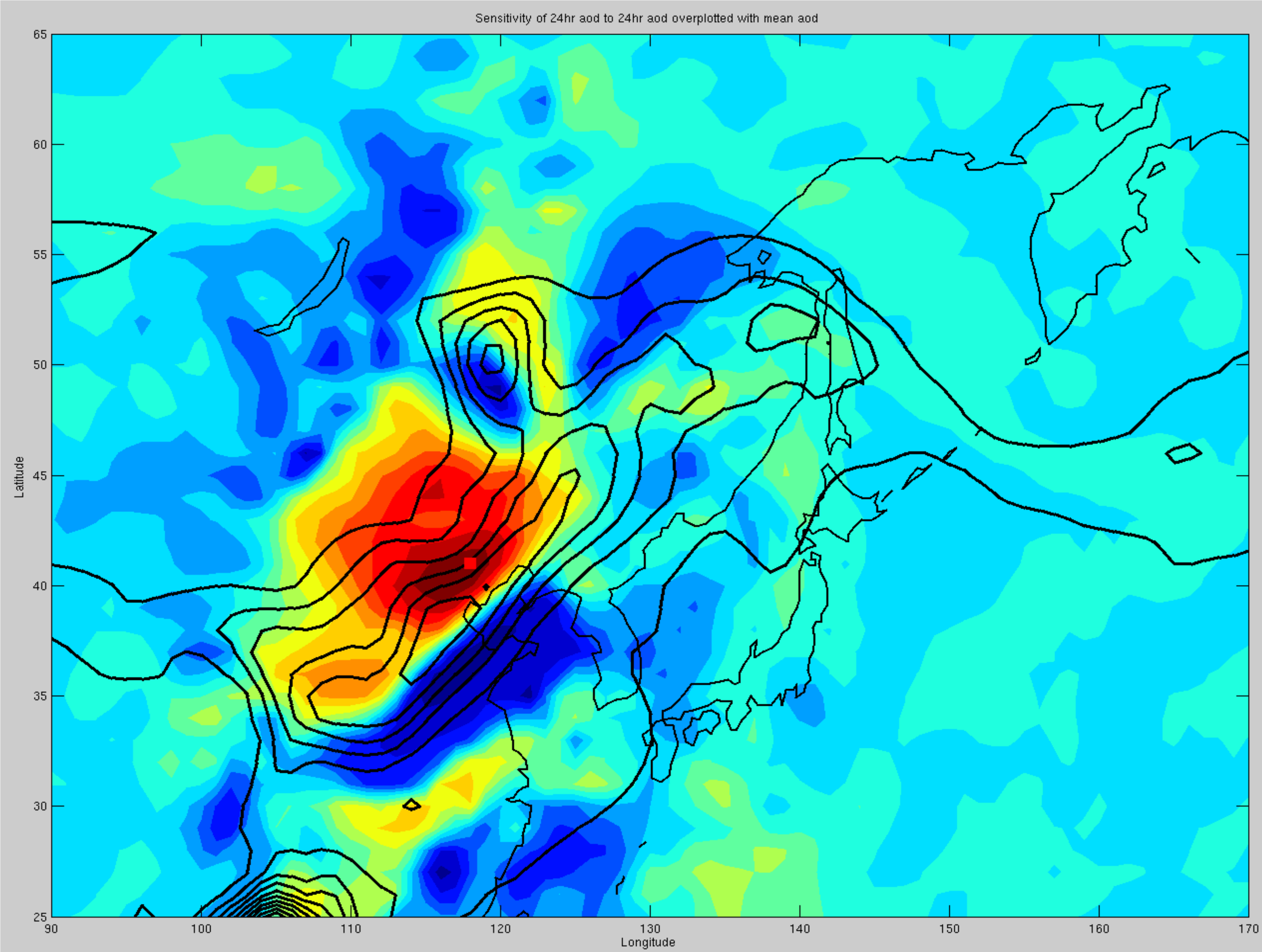
aod mean and std for 96hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0


aod mean and std for 120hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
0
0.2
0.4
0.6
0.8
1

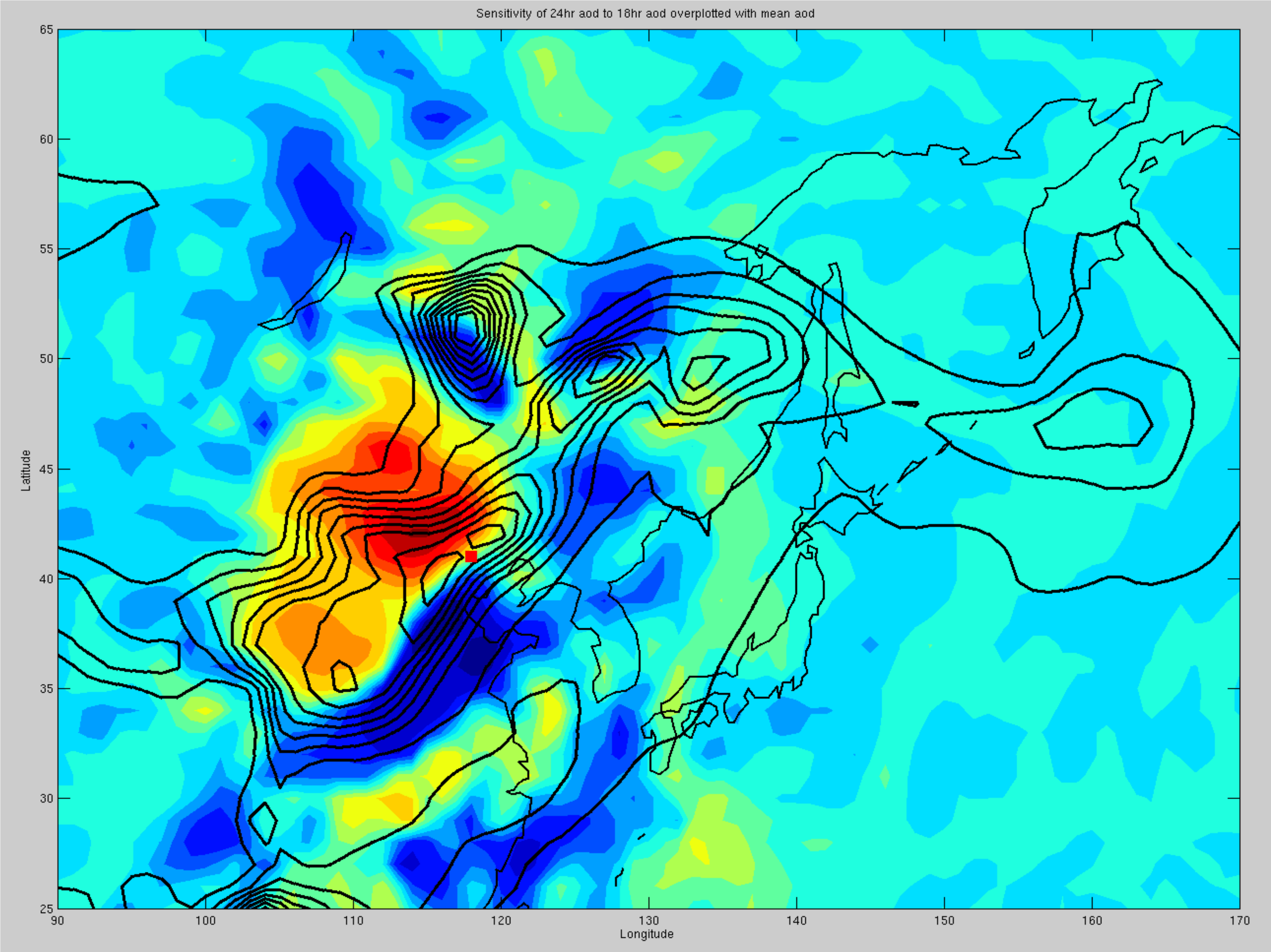

aod mean and std for 24hr forecast
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
1
0.8
0.6
0.4
0.2
0


aod plume at 118E, 41N
aod
forecast lead (hrs)
0
0.5
1
1.5
2
2.5
3
3.5
4
0
12
24
36
48
60
72
84
96
108
120

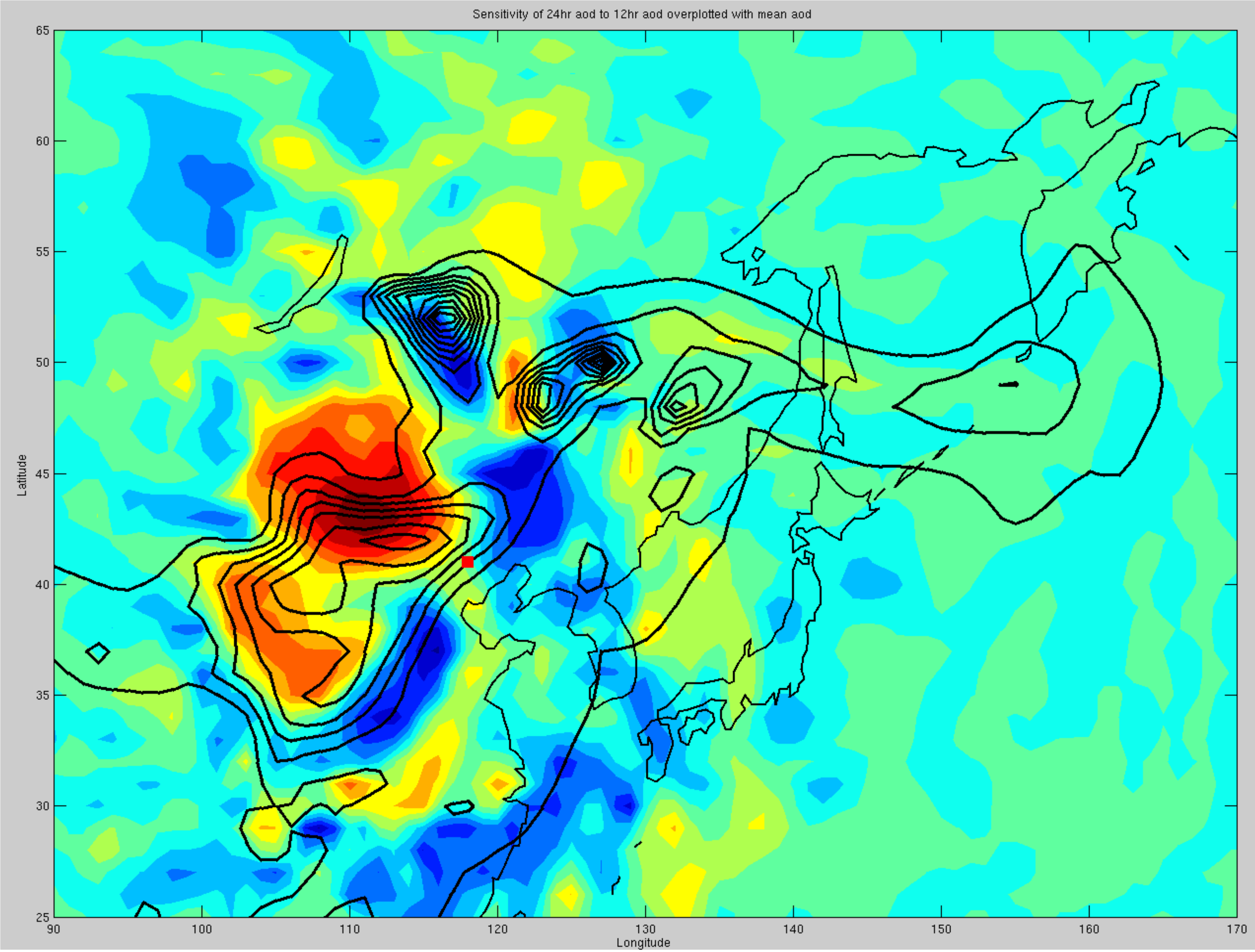

aod plume at 118E, 41N
aod
forecast lead (hrs)
0
0.5
1
1.5
2
2.5
3
3.5
4
0
12
24
36
48
60
72
84
96
108
120



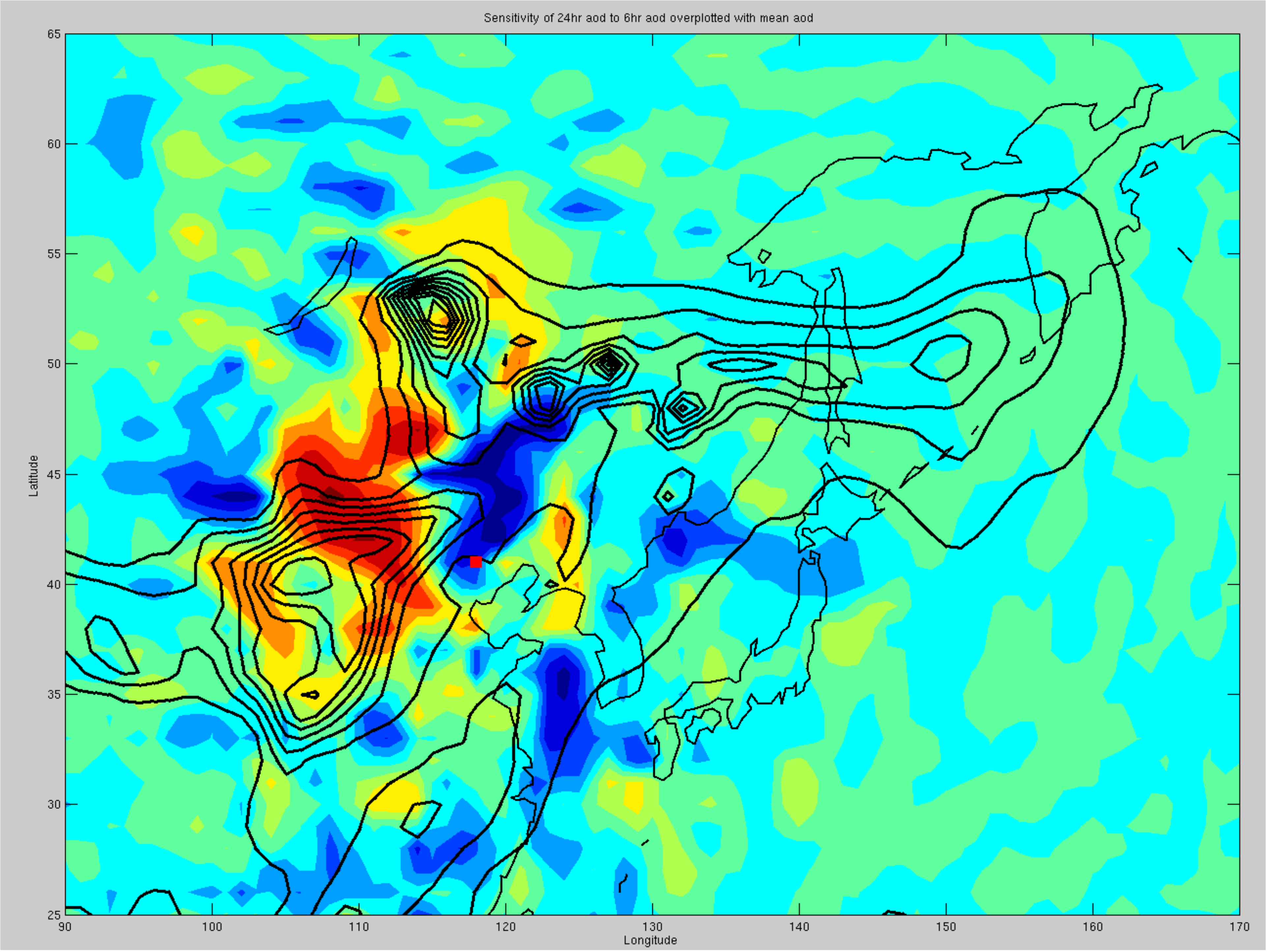
Sensitivity of 24hr aod to 24hr aod overplotted with mean aod
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



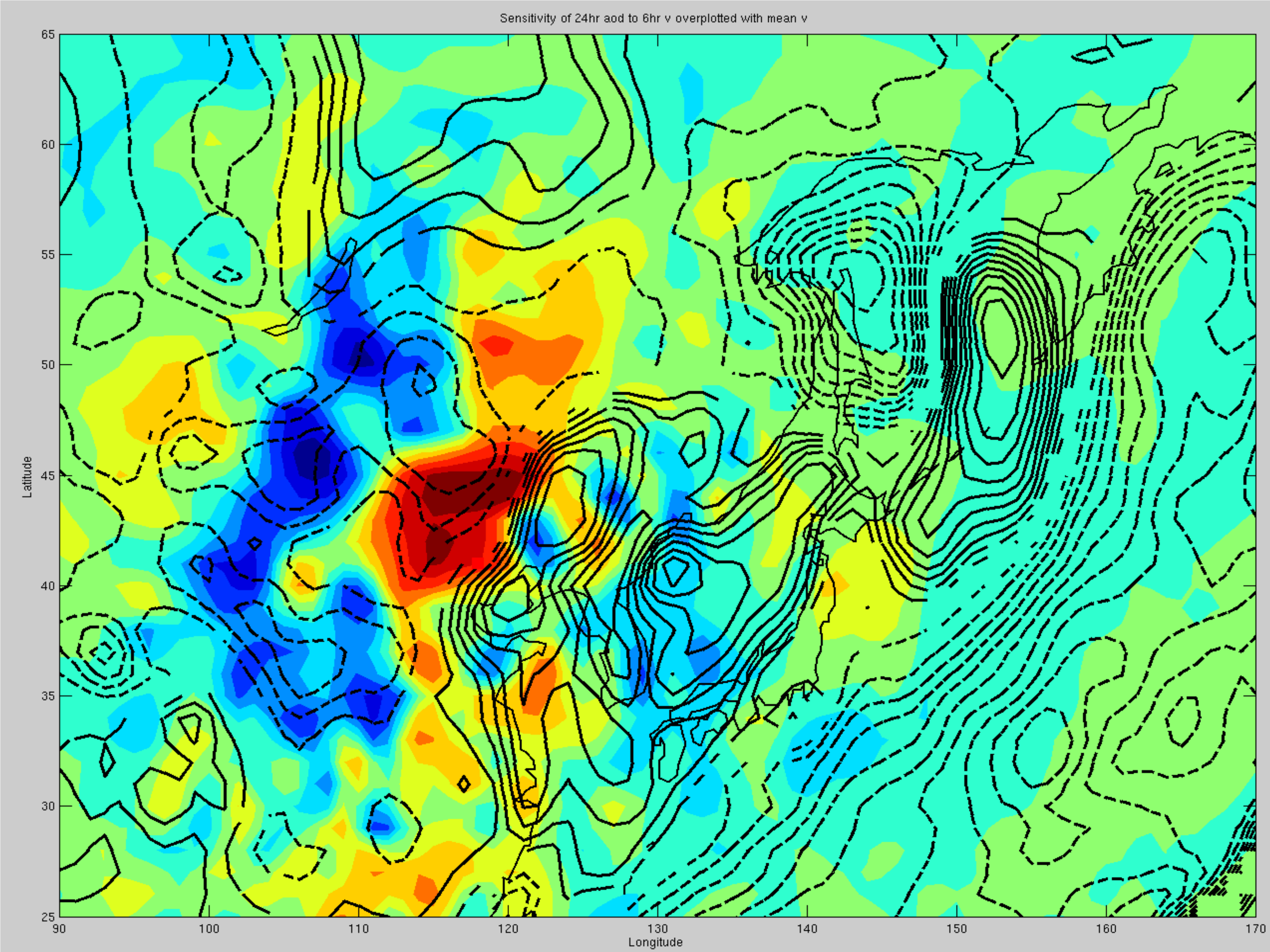
Sensitivity of 24hr aod to 18hr aod overplotted with mean aod
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



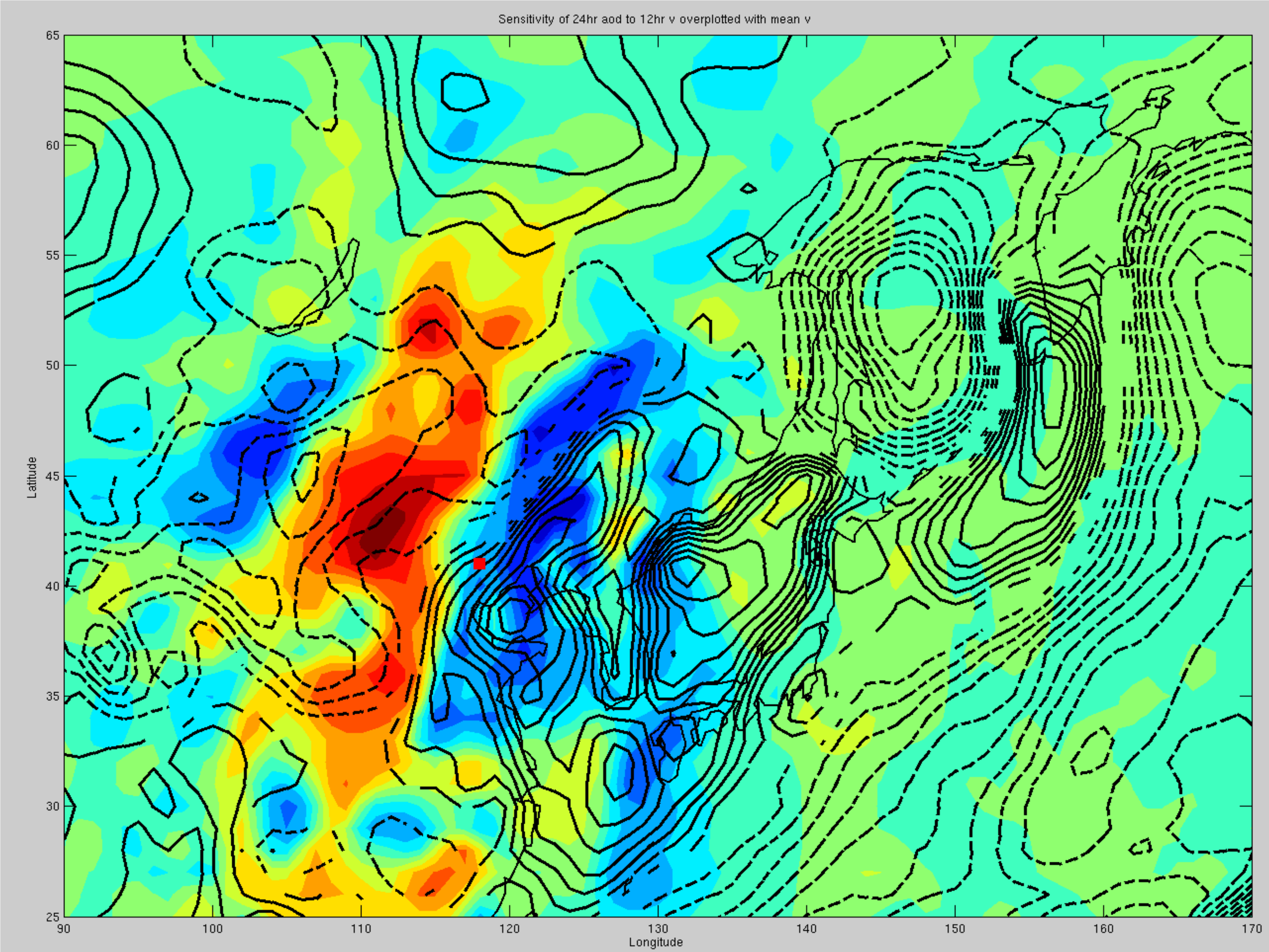
Sensitivity of 24hr aod to 12hr aod overplotted with mean aod
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



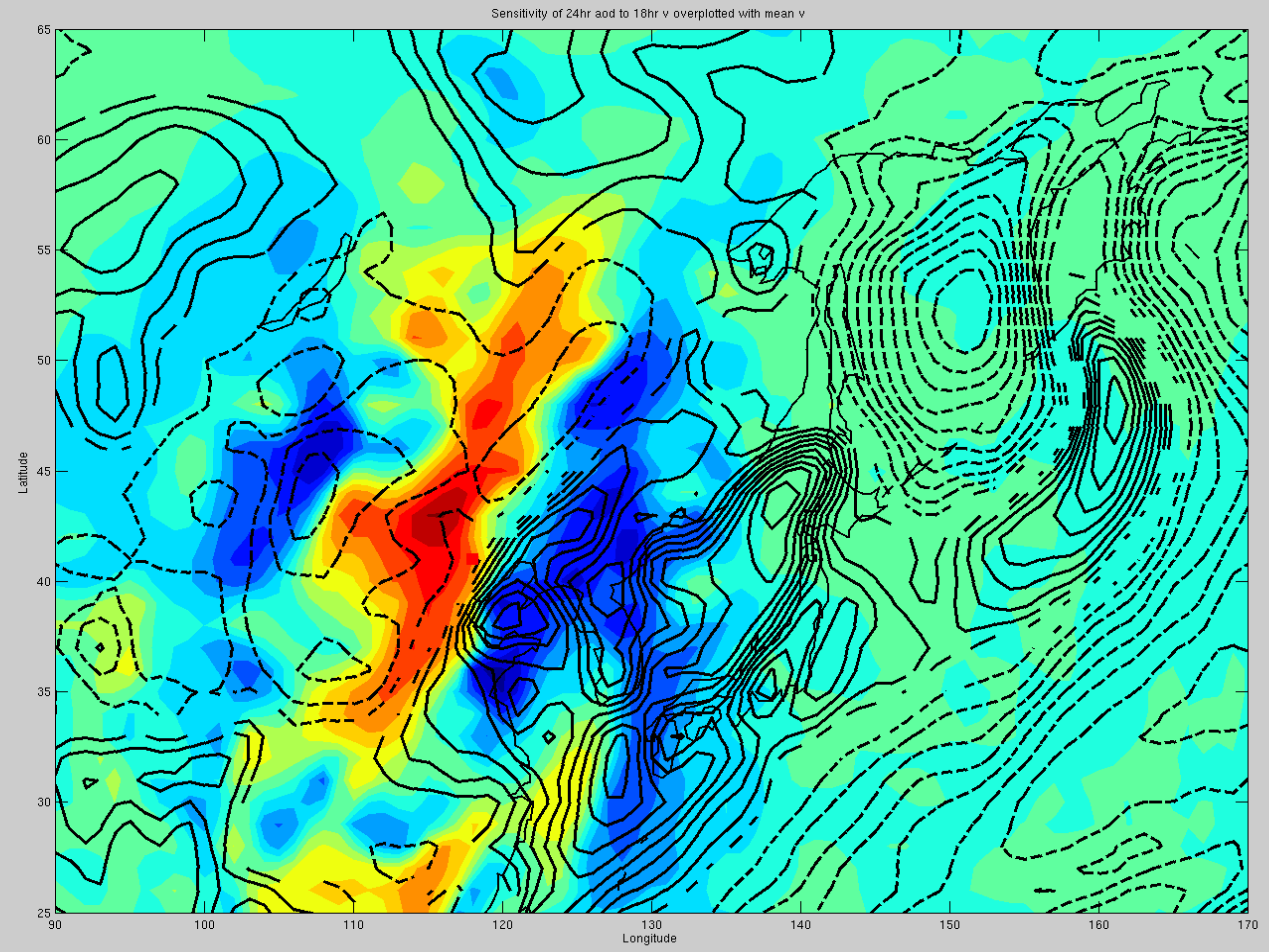
Sensitivity of 24hr aod to 6hr aod overplotted with mean aod
65
60
55
50
45
40
35
30
25
Latitude
90
100
110
120
130
140
150
160
170
Longitude



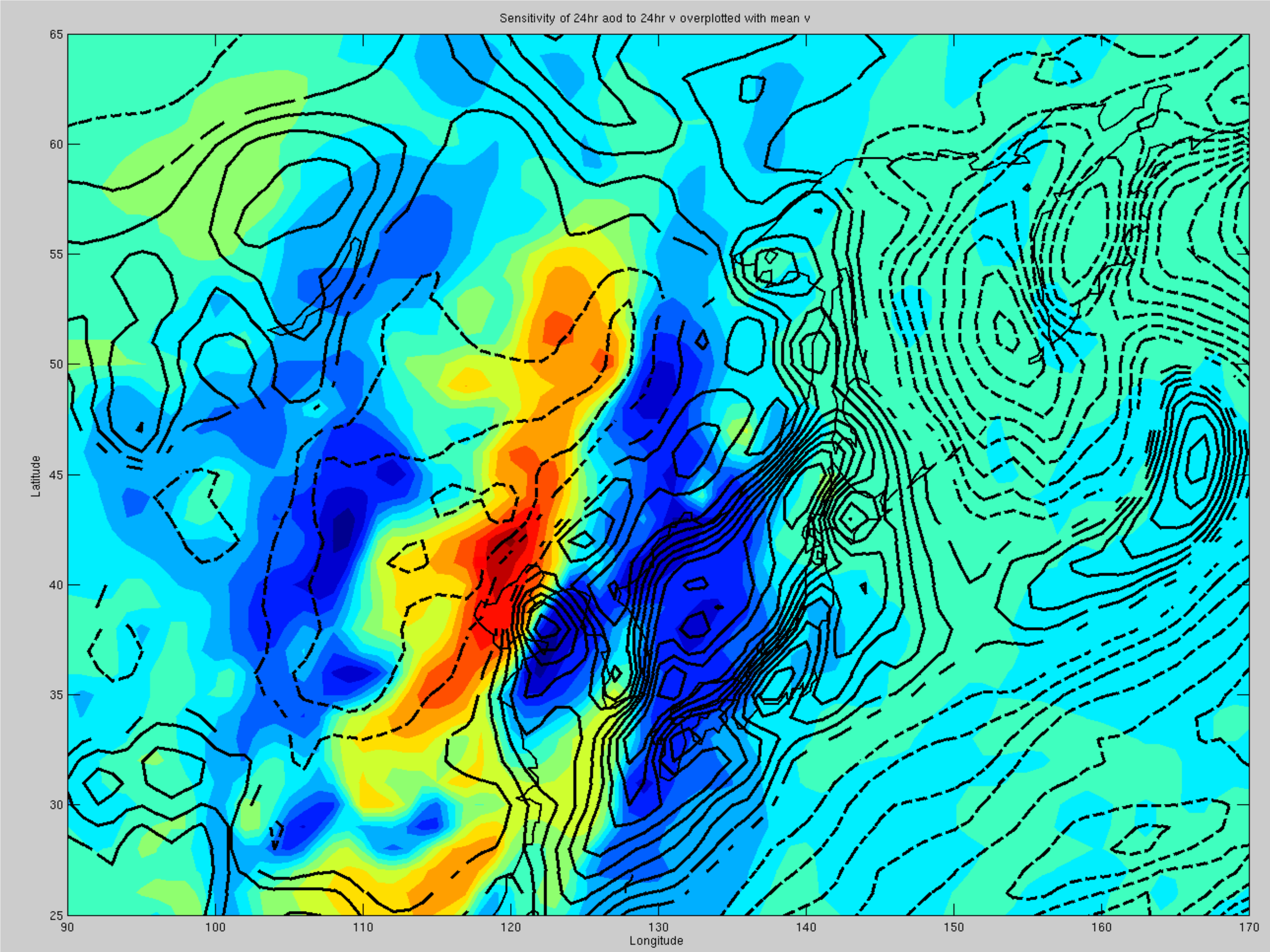
Sensitivity of 24hr aod to 6hr v overplotted with mean v
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



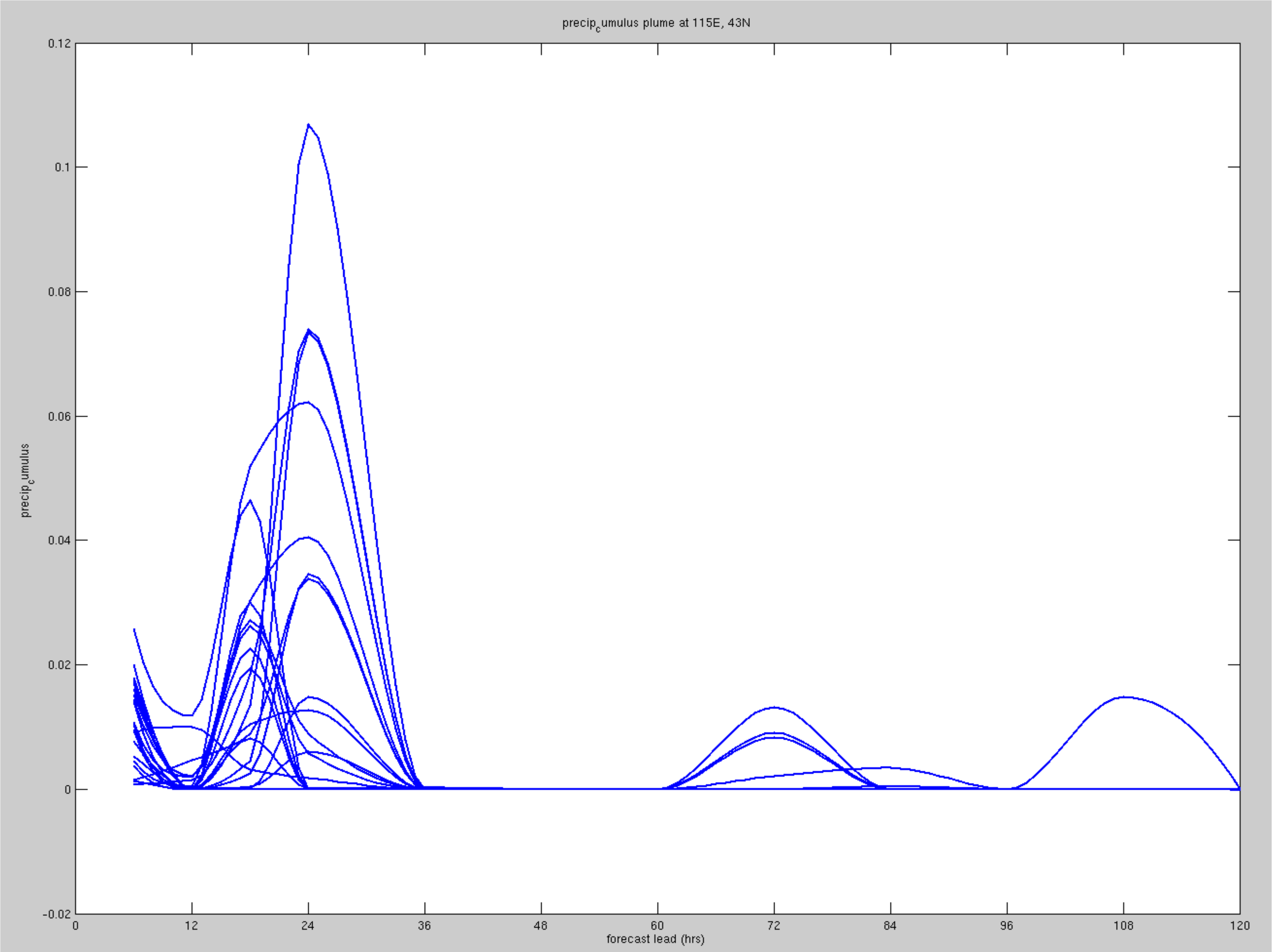
Sensitivity of 24hr aod to 12hr v overplotted with mean v
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



Sensitivity of 24hr aod to 18hr v overplotted with mean v
Latitude
Longitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170



Sensitivity of 24hr aod to 24hr v overplotted with mean v
Latitude
65
60
55
50
45
40
35
30
25
90
100
110
120
130
140
150
160
170
Longitude



precip$_c$umulus plume at 115E, 43N
precip$_c$umulus
0.12
0.1
0.08
0.06
0.04
0.02
0
-0.02
0
12
24
36
48
60
72
84
96
108
120
forecast lead (hrs)

Where does the value of ensembles lie for aerosols?

- Routine analysis of forecast products
- Partition sensitivities between meteorology and sources
- Identify sensitivity between what we can observe (e.g. AOD) and what we care about (e.g. $PM_{2.5}$)
- Target observing assets that will have the largest impact on what we care about
- Identify strong sensitivities between meteorology and aerosols in an inline model.
- Multi-model sensitivity (although can be difficult to interpret)