



# **NWS Verification Team Meeting 07/10/08**

## **Review of the EVS verification exercises**

**James Brown, Julie Demargne,  
Yuqiong Liu, Christopher Ware**



# Three exercises

## 1. Exercise 1

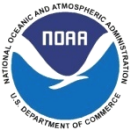
- Synthetic temperature forecasts
- Regression of obs. & ens. mean, without bias

## 2. Exercise 2

- Same synthetic temperature forecasts
- Regression with biased mean or spread

## 3. Exercise 3

- Real streamflow forecasts from MARFC



# Metrics considered

## Box plots (detailed picture of errors)

- Useful for data exploration (outliers etc.)
- Can construct in several ways....
- ....we pooled errors by forecast lead time

## Mean CRPS (summary picture of errors)

- Score for each forecast, then averaged.
- Good overview of 'conditional' biases....
- ...e.g. bias with increasing lead time or obs.



# Metrics considered

## Reliability & discrimination (CT & ROC)

- When Y was forecast, what was observed?

*"Our model predicts a 90% chance of flooding."*

**RELIABLE if observed 9/10 times issued (CT).**

- When X was observed, what was forecast?

*"When we observe Action Stage only, our model predicts a 100% chance of Flood Stage."*

**Cannot DISCRIMINATE AS from FS (ROC).**



# **Exercise 1 (demo., questions, results)**



# Questions

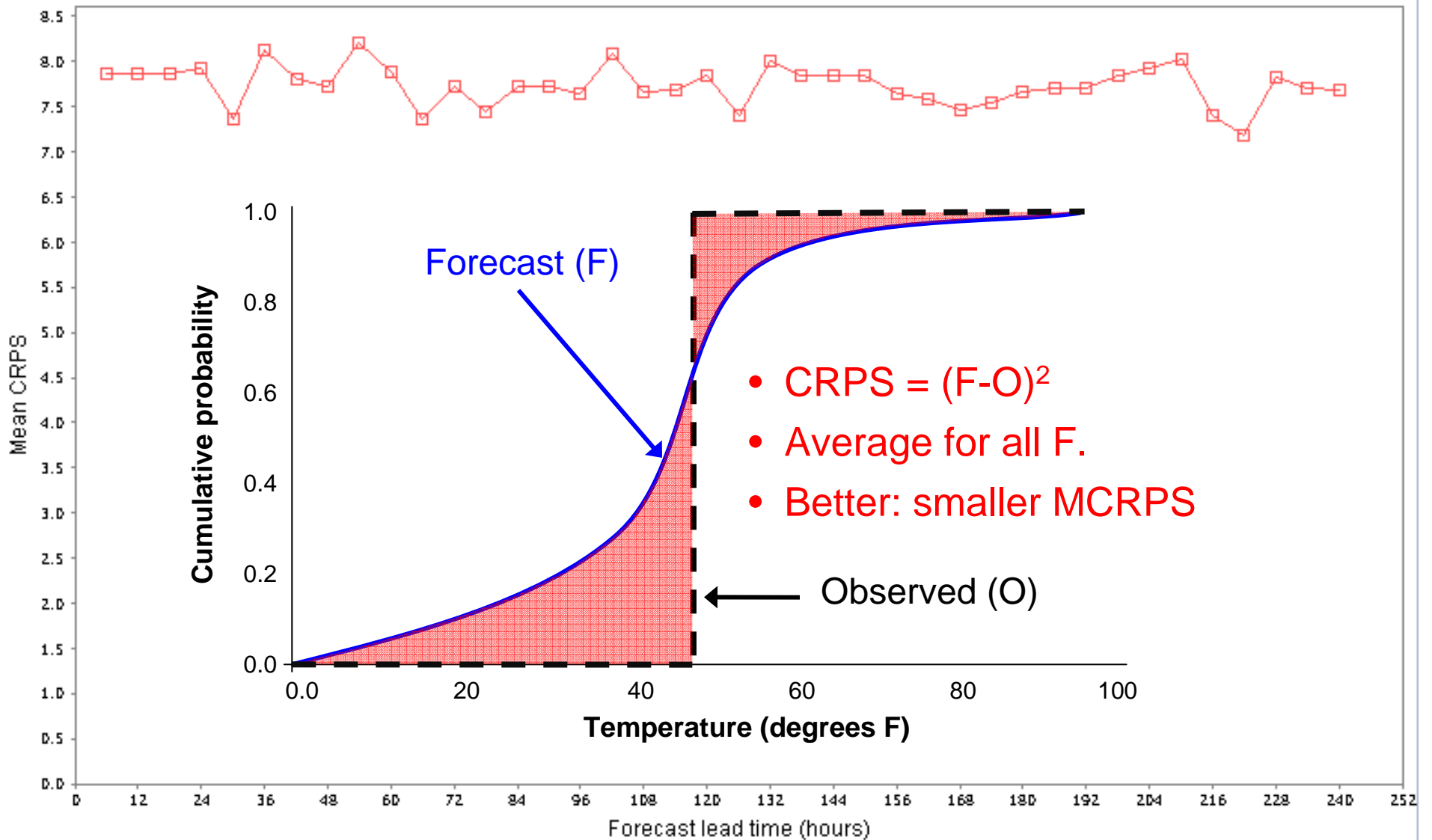
## A) “Do the results look as expected?”

- “Do the errors consistently increase with lead time?”
- “What can you tell from the box plots with regard to biases in the ensemble mean and spread (if any)?”

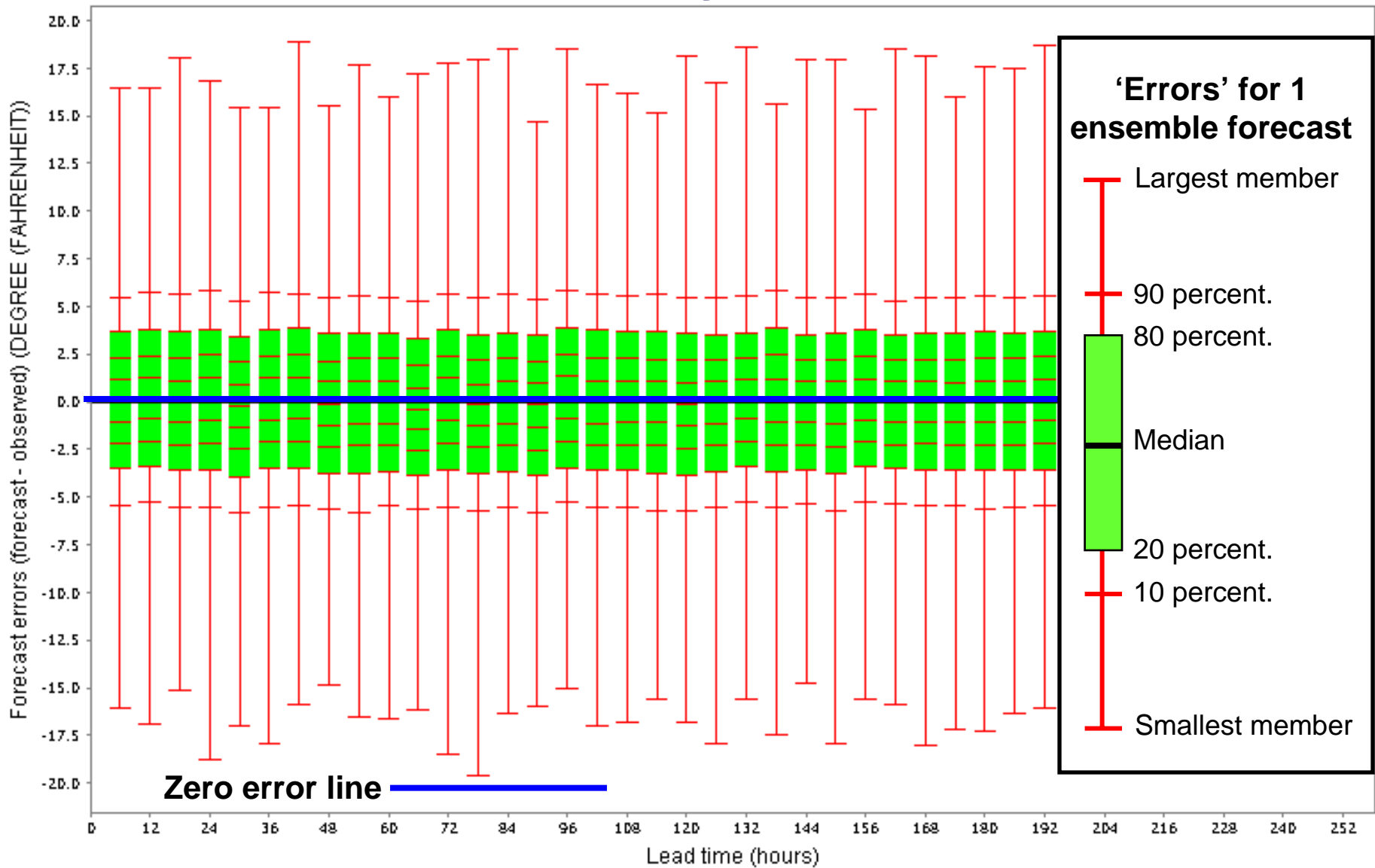
## B) “Why aren’t the results even better?”

- “For example, why is the line in the Talagrand plots not perfectly diagonal?”

# Mean CRPS

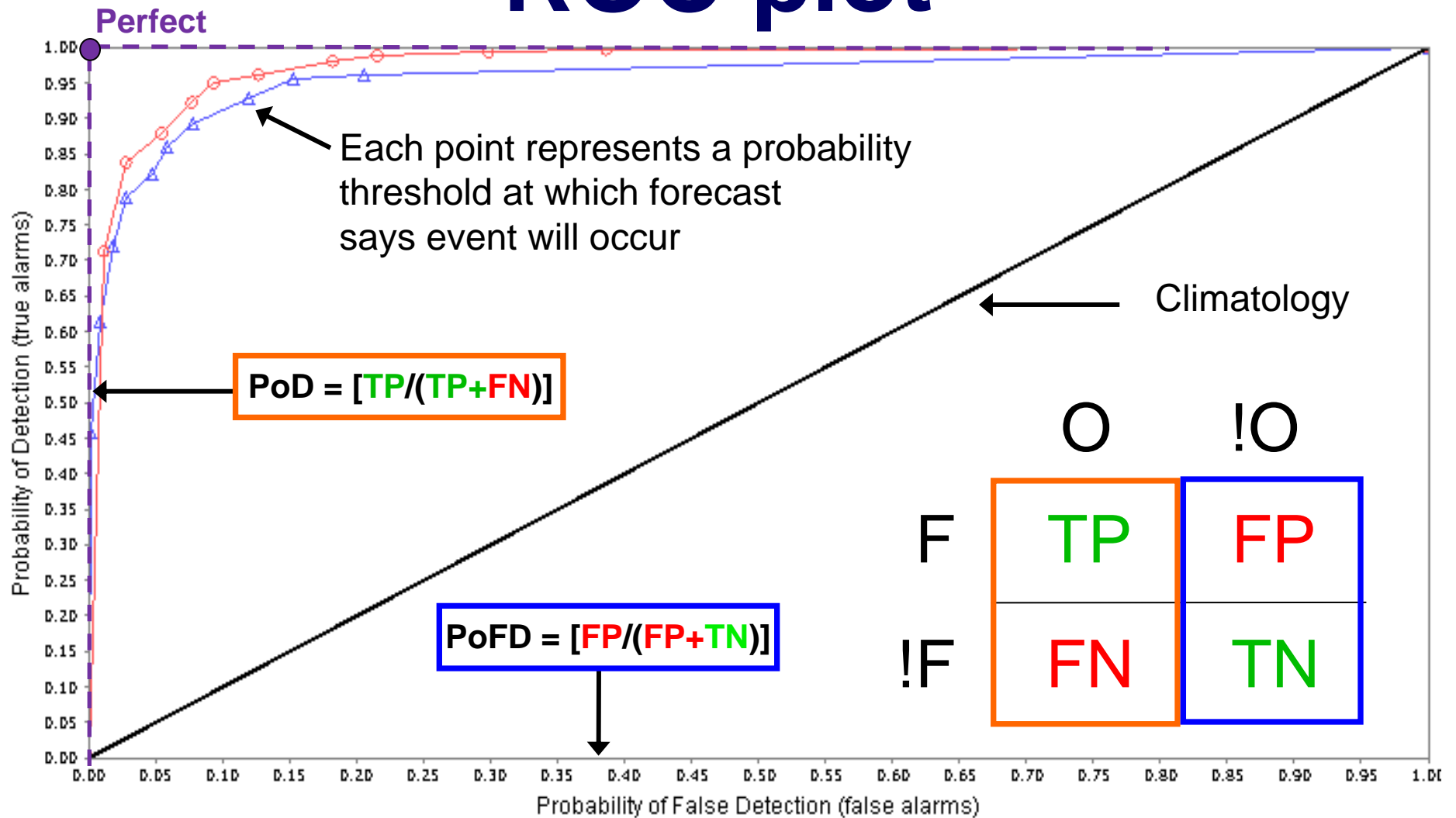


# Box plot by lead time





# ROC plot

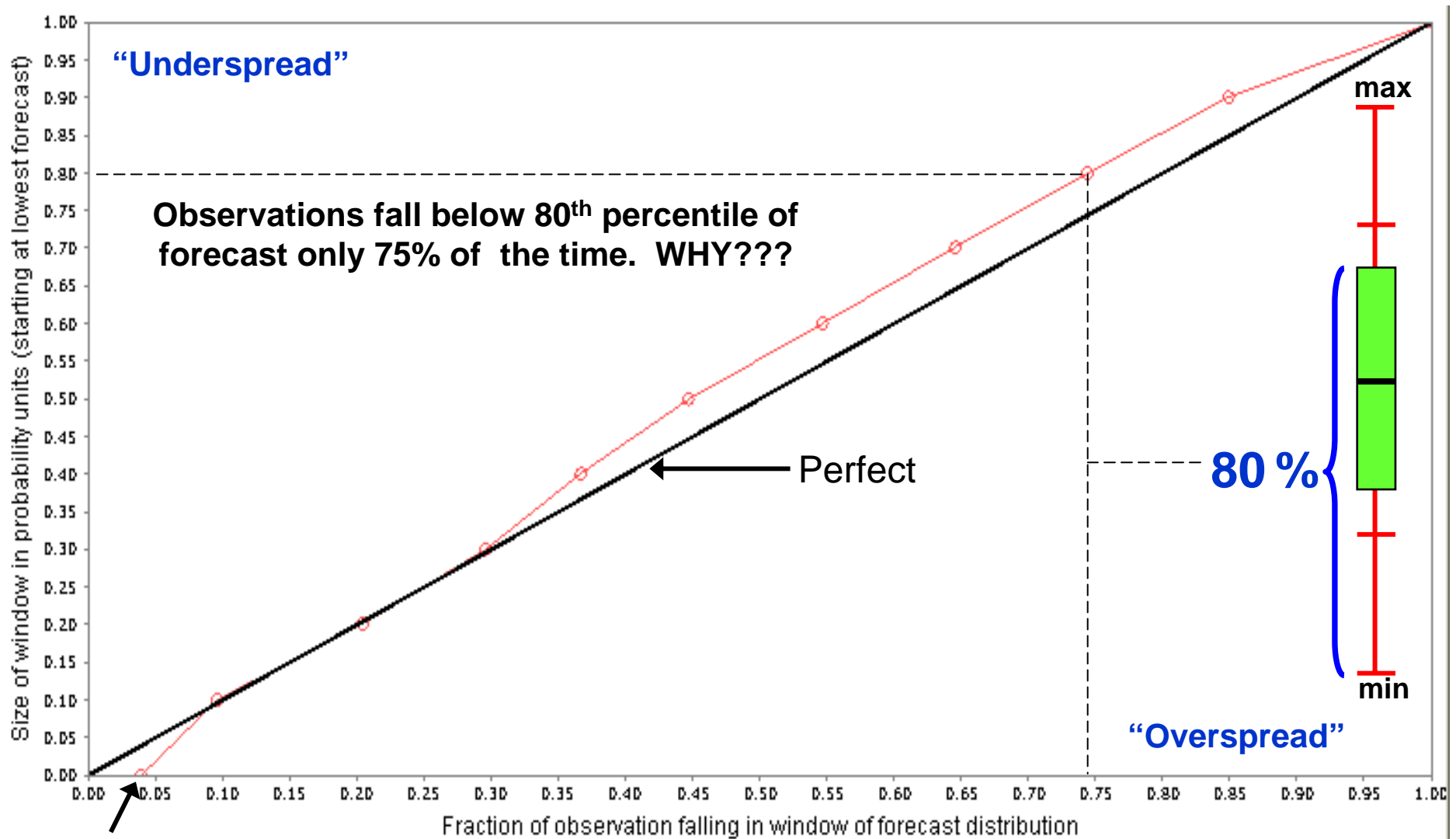


— Random guess (no skill)    -o- P[ob] > 0.3 (74.061)    -△- P[ob] > 0.7 (85.401).

1<sup>st</sup> event: T > 74 DegF

2<sup>nd</sup> event: T > 85 DegF

# Talagrand (ex. from 48hrs)



Observations fall below all forecasts 4% of the time



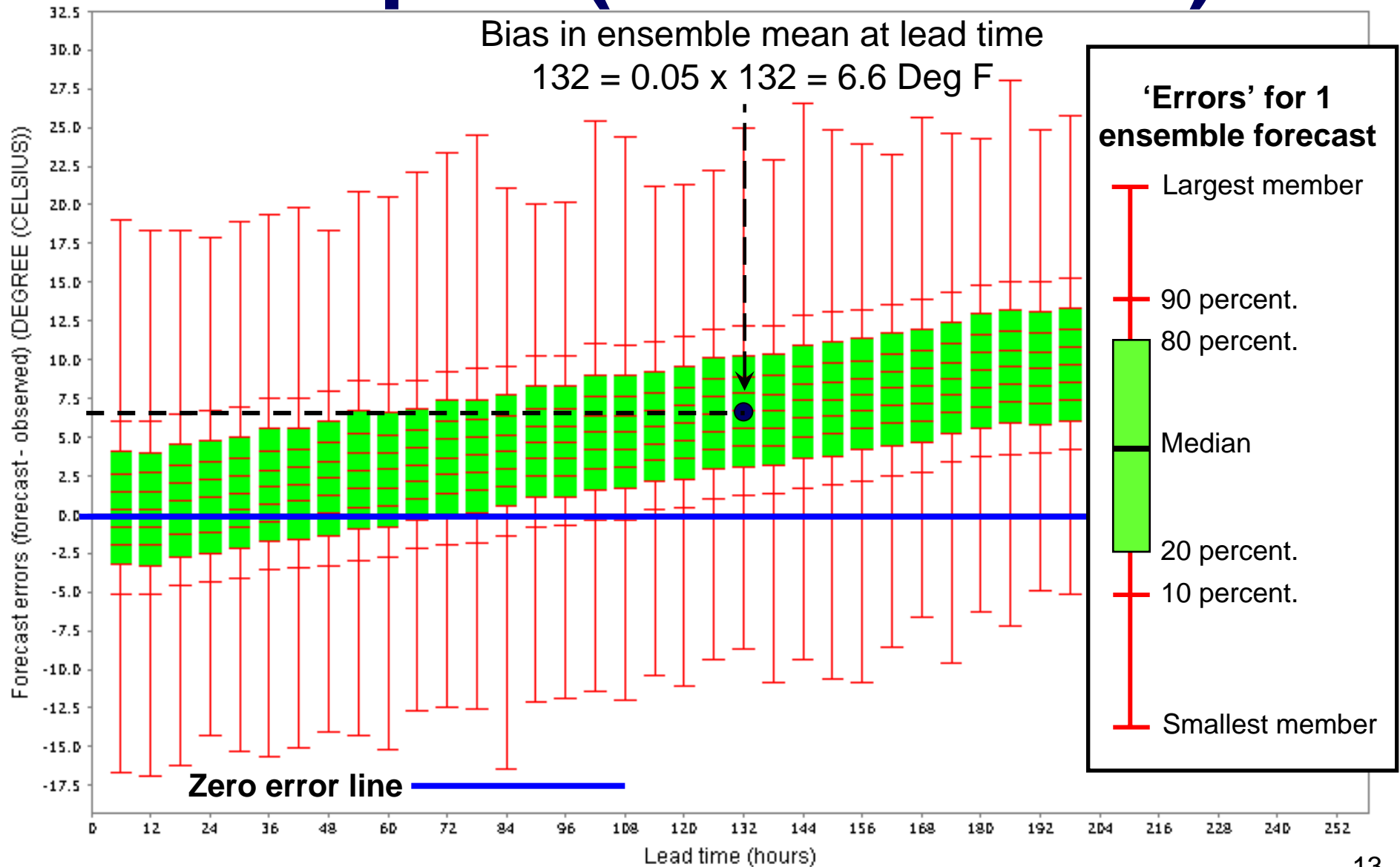
# **Exercise 2 (demo., questions, results)**



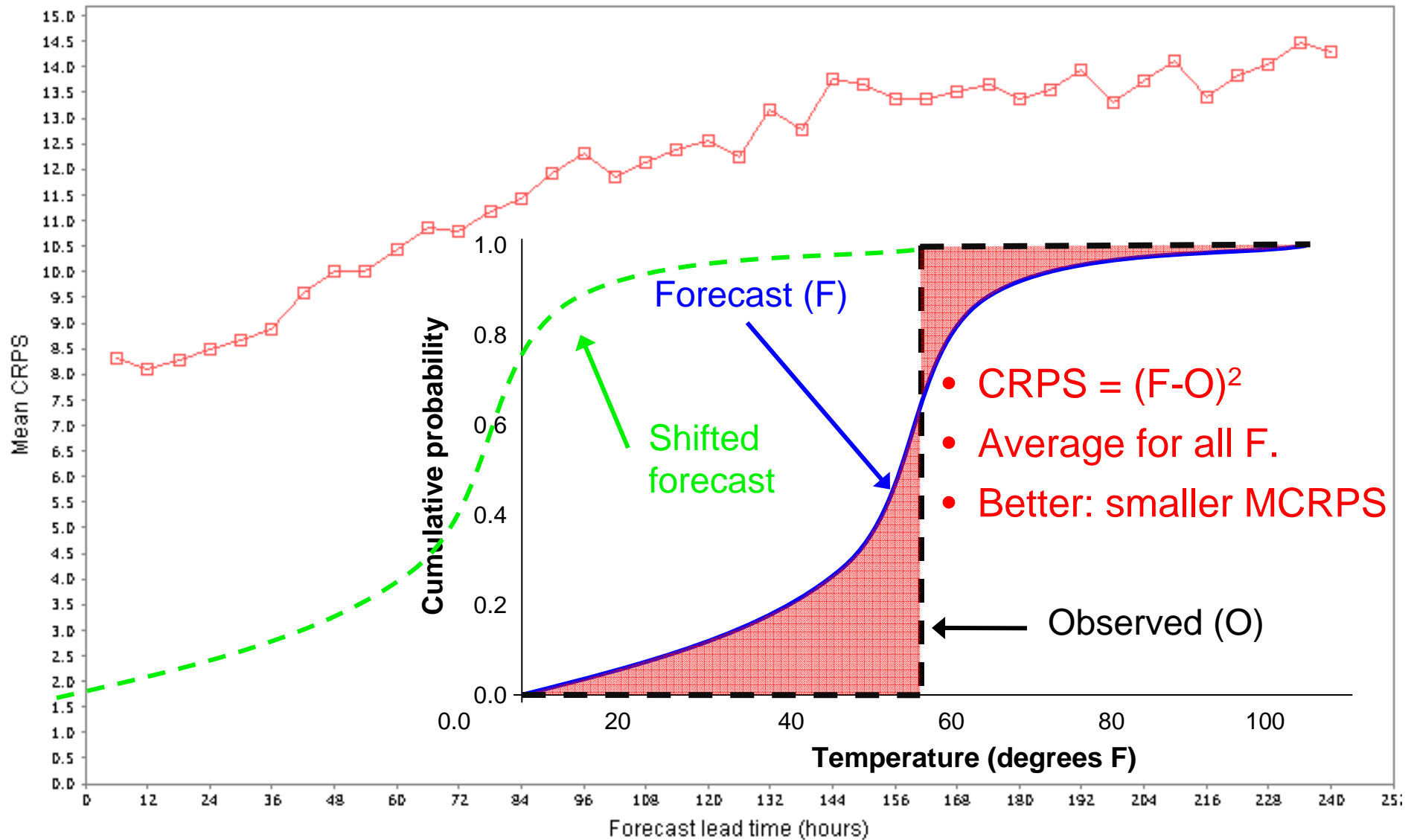
# Questions

- A) “Can you tell that the forecasts with biased mean deteriorate in quality with lead time?”**
- B) “Why are the values for the MCRPS much larger for the scenario with biased mean?”**
- C) “For the biased mean scenario, examine the Cumulative Talagrand diagrams with increasing lead time. Are there any changes with lead time and, if so, what do they indicate?”**

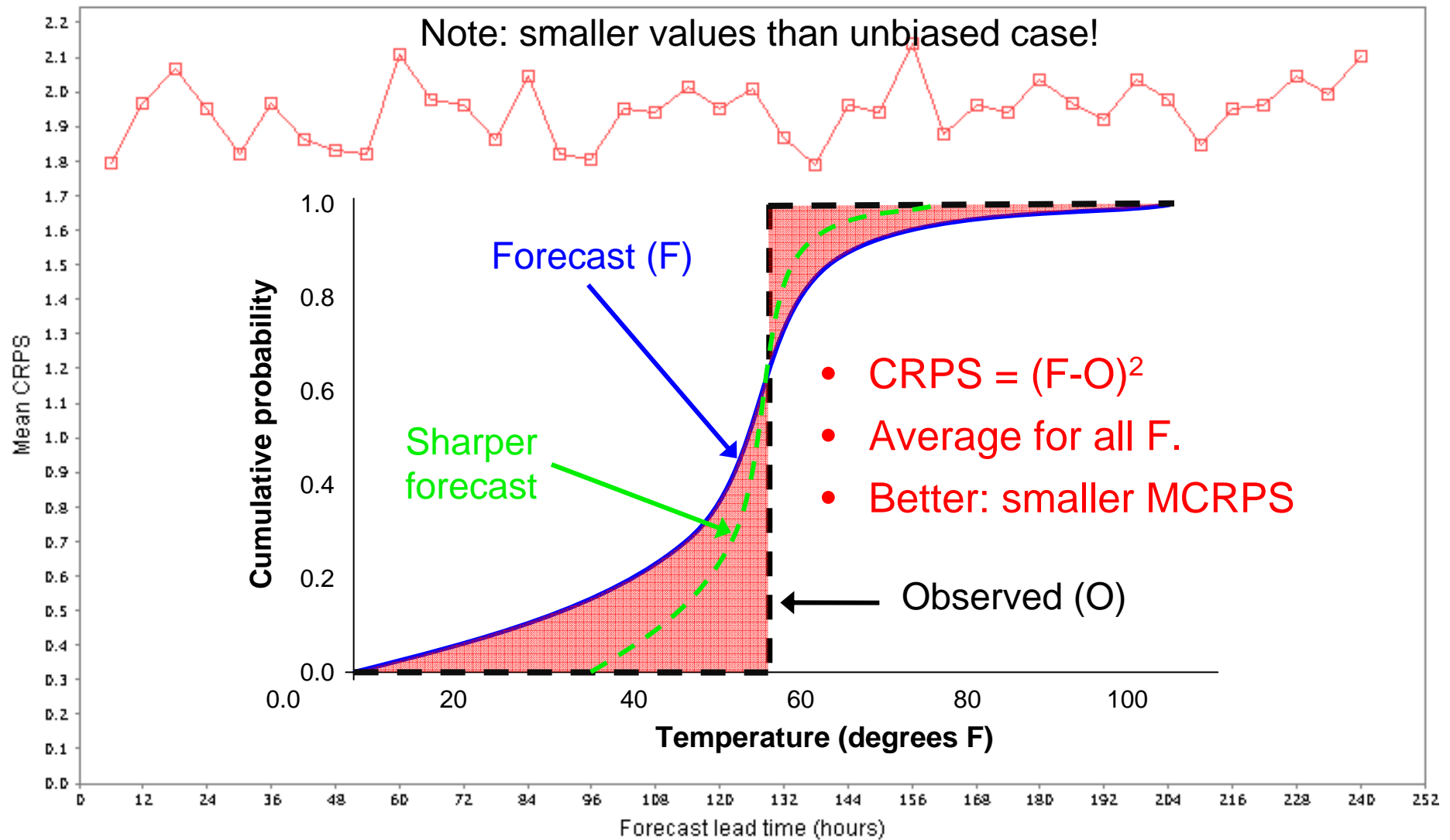
# Box plot (biased mean)



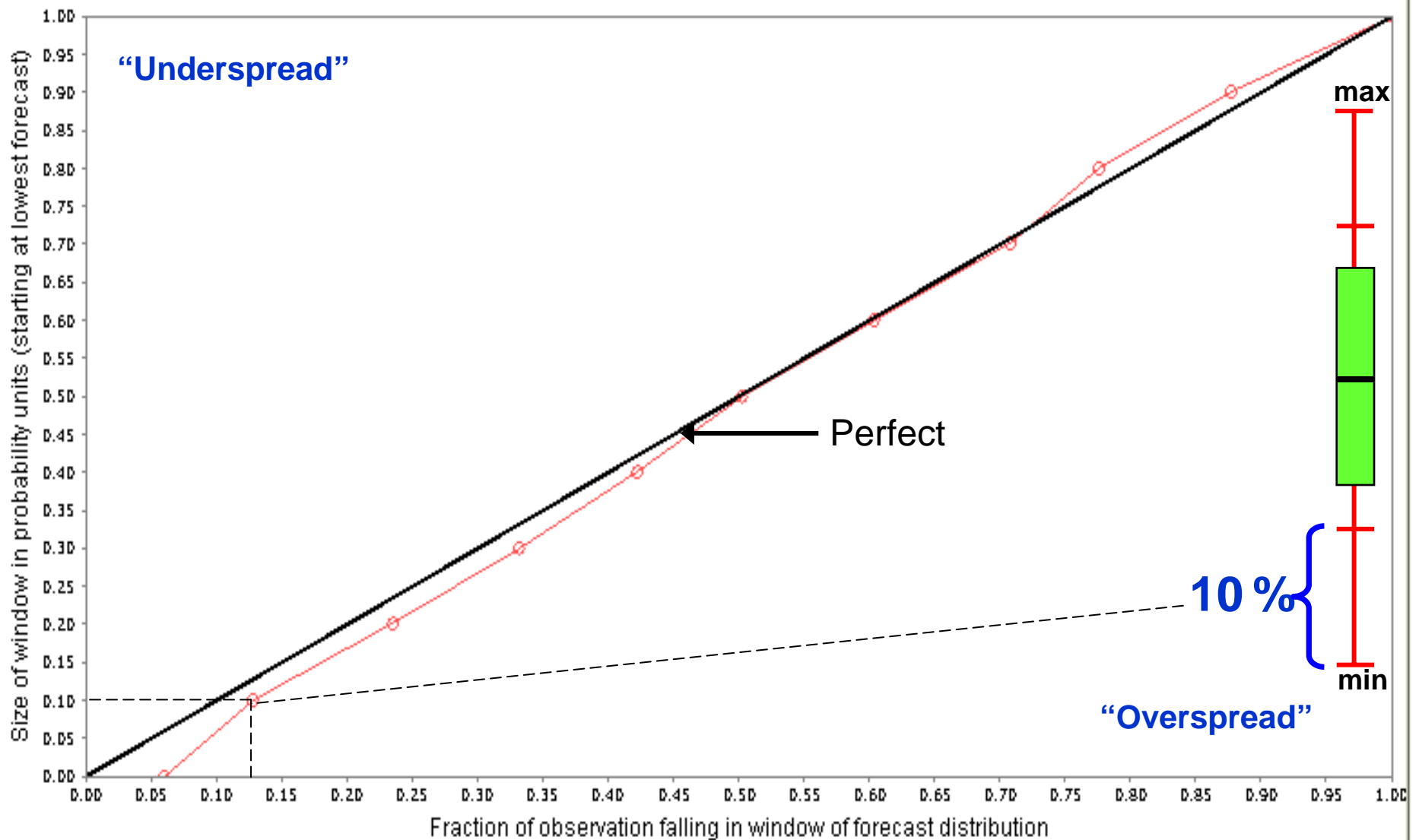
# Mean CRPS (biased mean)



# Mean CRPS (biased spread)

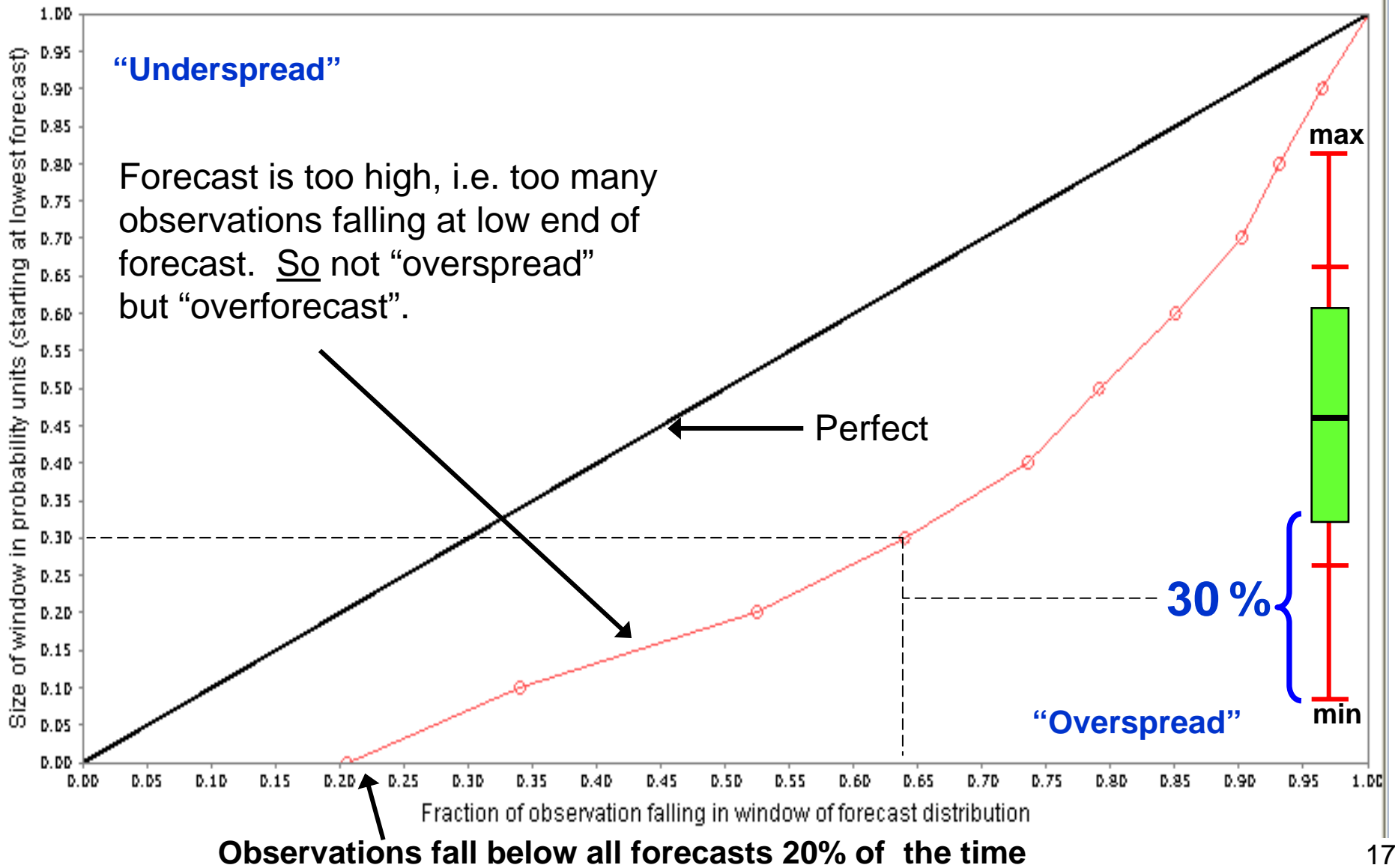


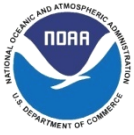
# Talagrand (biased mean, 6)





# Talagrand (biased mean, 60)





# **Exercise 3 (demo., questions, results)**



# Questions

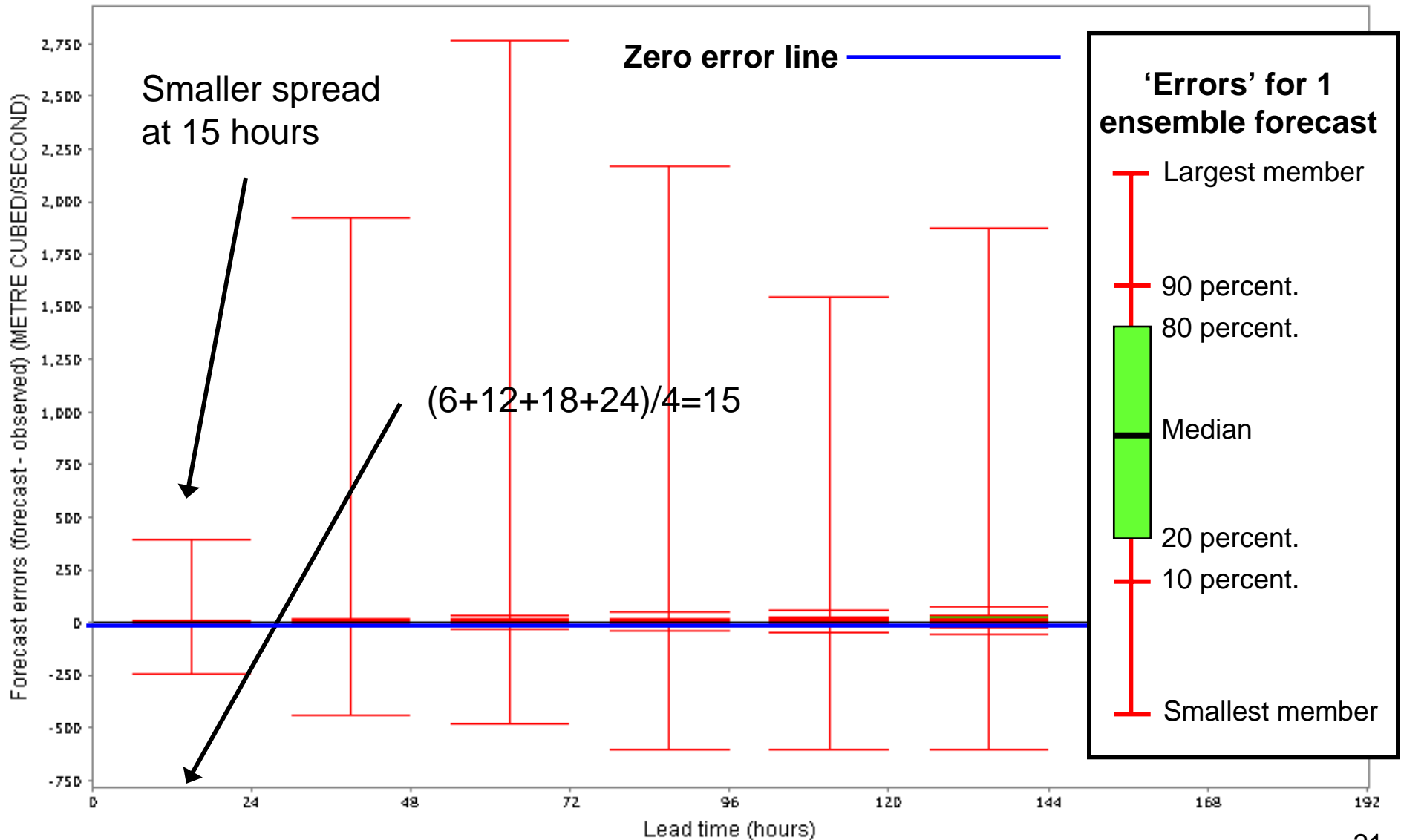
- A) “Why is there now 15 hours between each lead time instead of 6?”**
- B) “Examine the box plots pooled by lead time. What type of forecasting bias is present?”**
- C) “Notice from the box plots that the ensemble spread is low at small lead times. How does this impact the reliability at short lead times (e.g. compare diagrams at day 1 and 7)?”**
- D) “Given box plots of errors by observed value, how do errors vary with observed value?”**



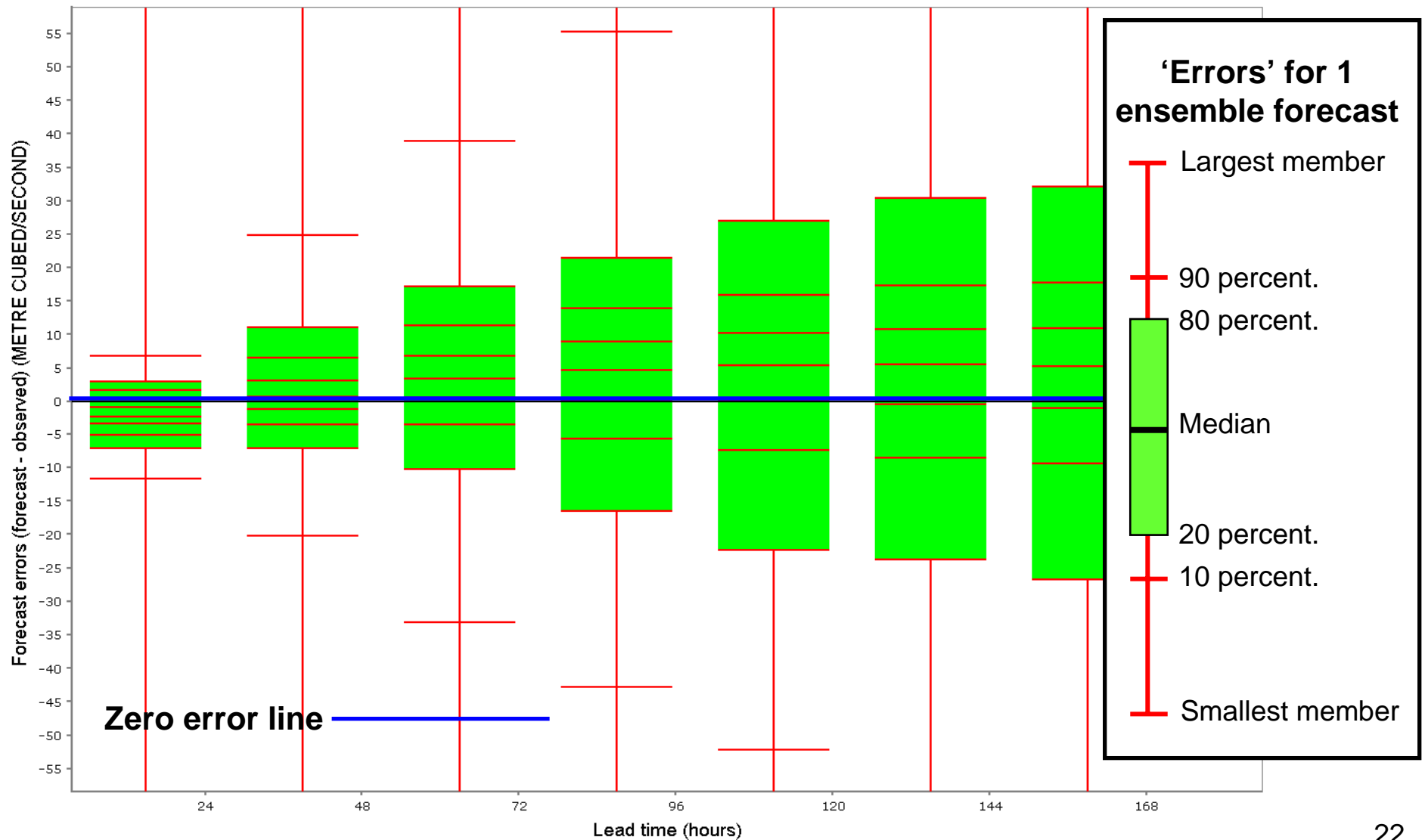
# Questions

- E) “Examine the deterministic verification metrics for the ensemble means. What do they indicate about the forecasts?”**
- F) “Do the ROC curves vary as expected?”**

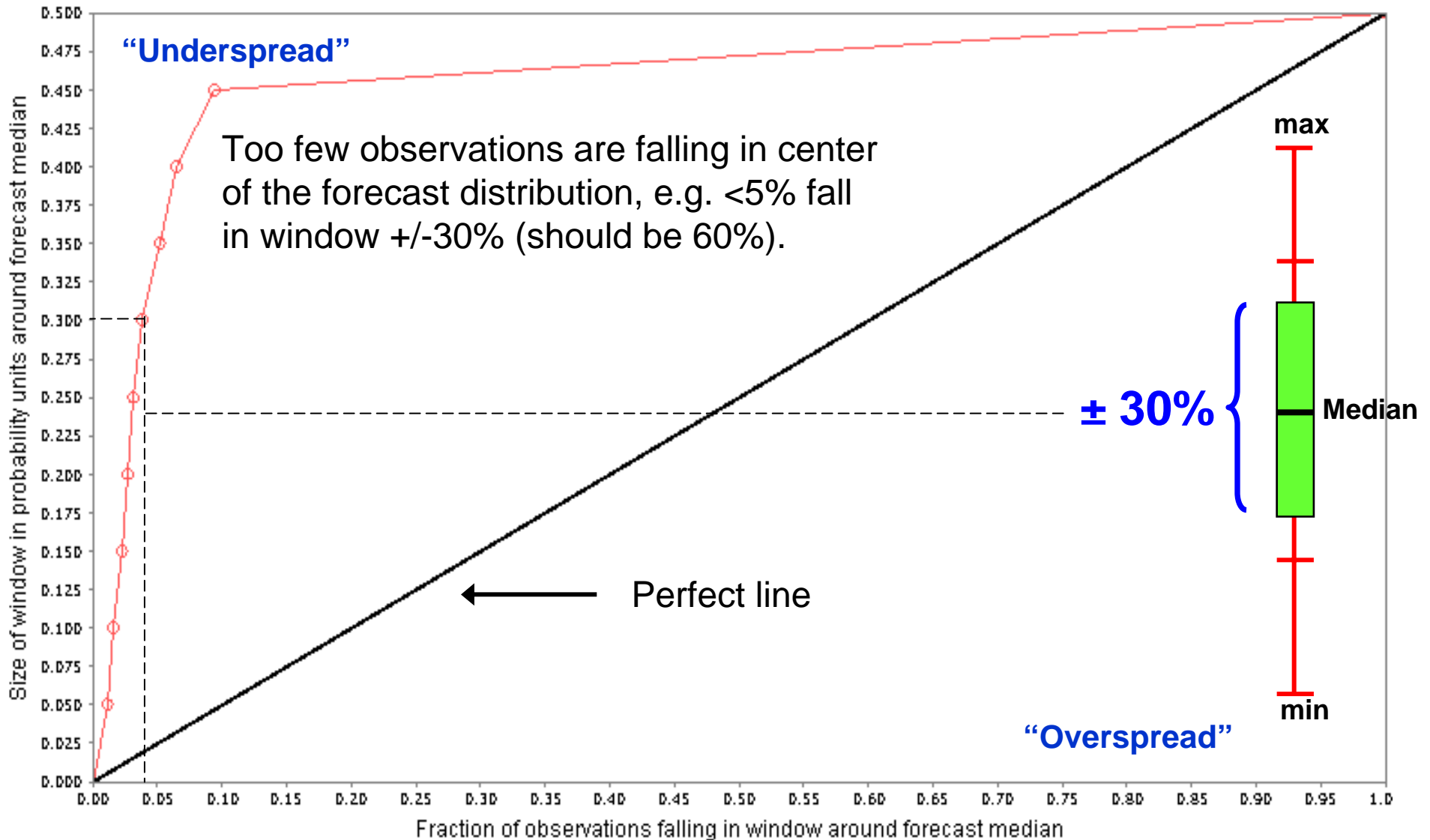
# Box plot by lead time



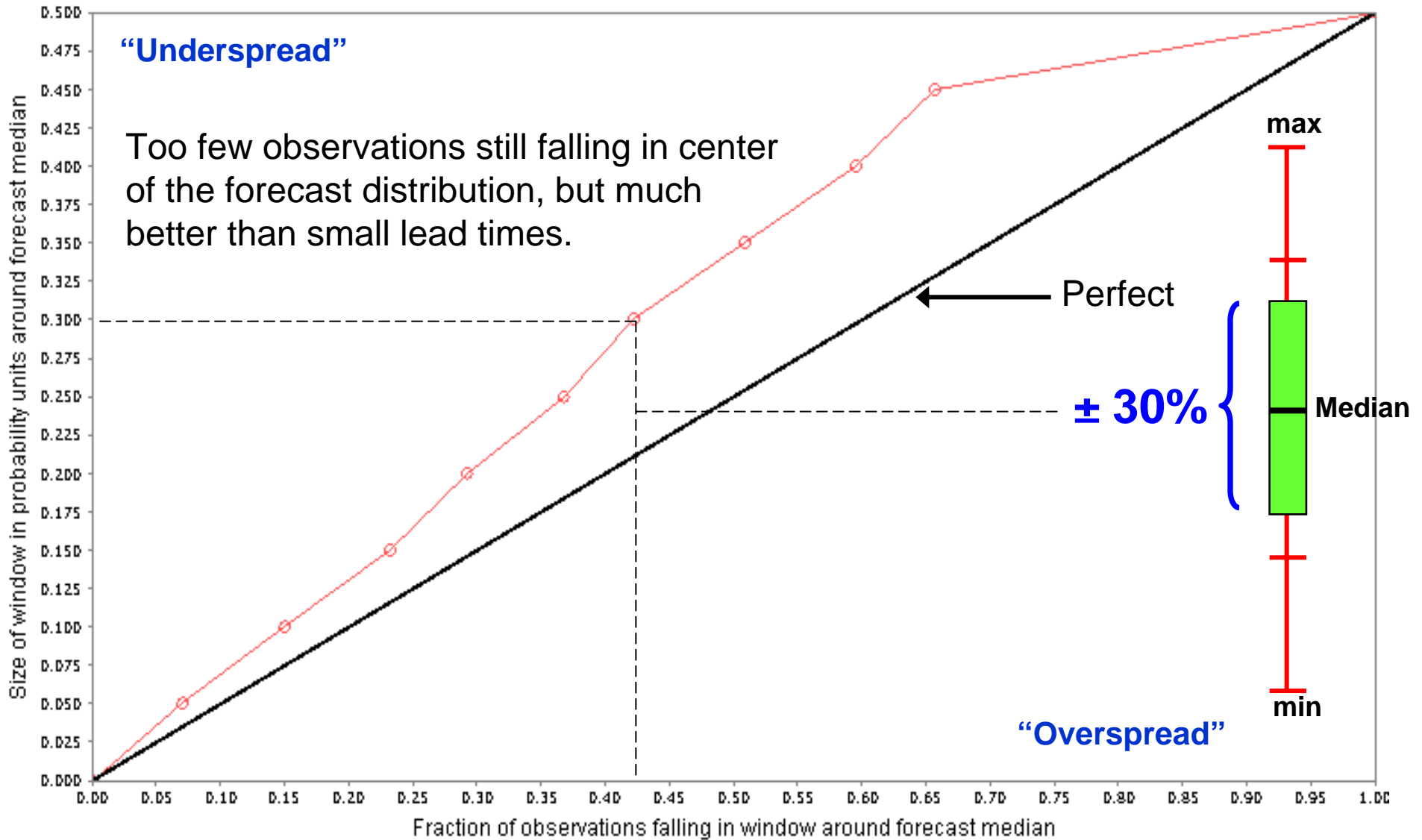
# Box plot by lead time (zoom)



# Talagrand (day 1)

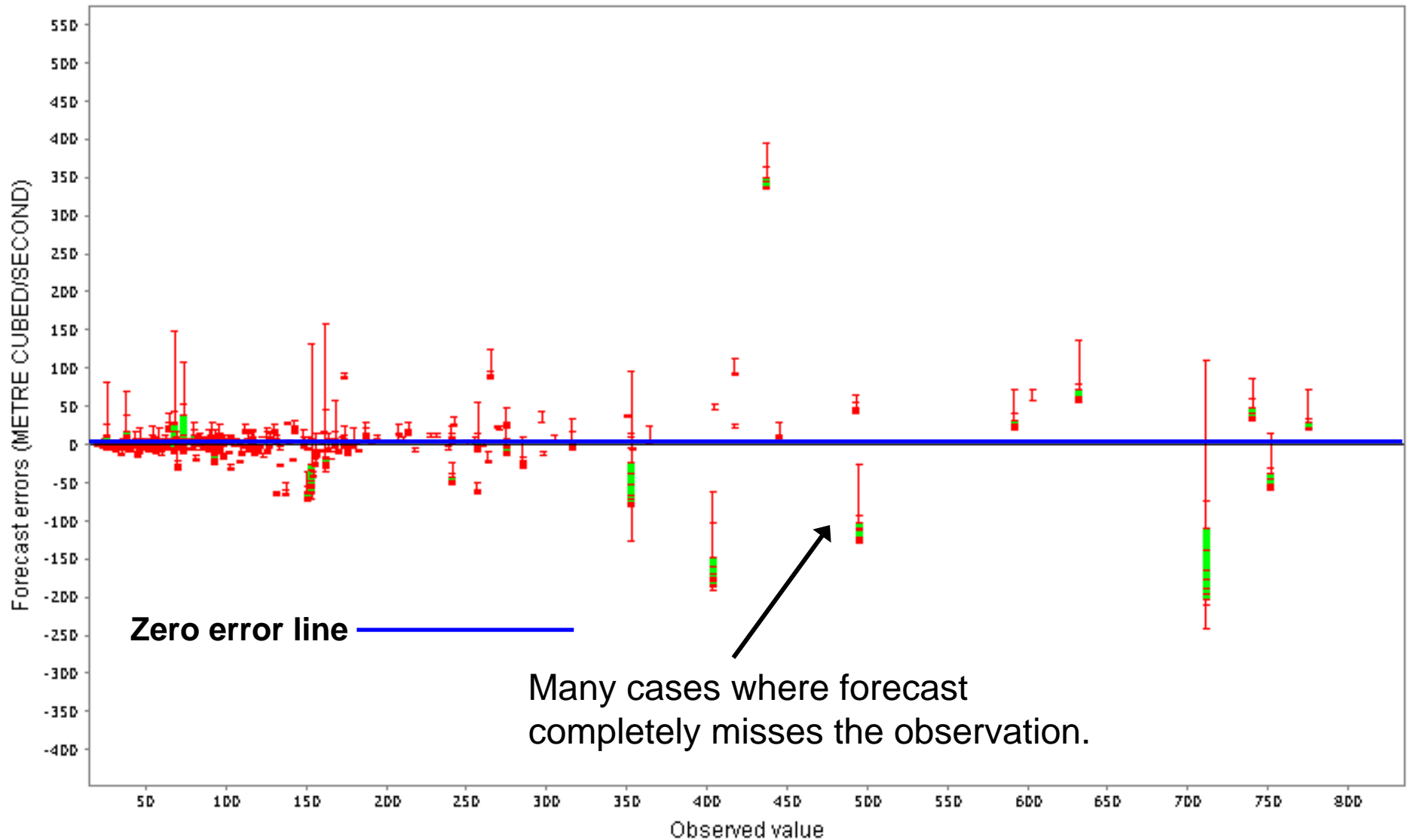


# Talagrand (day 7)

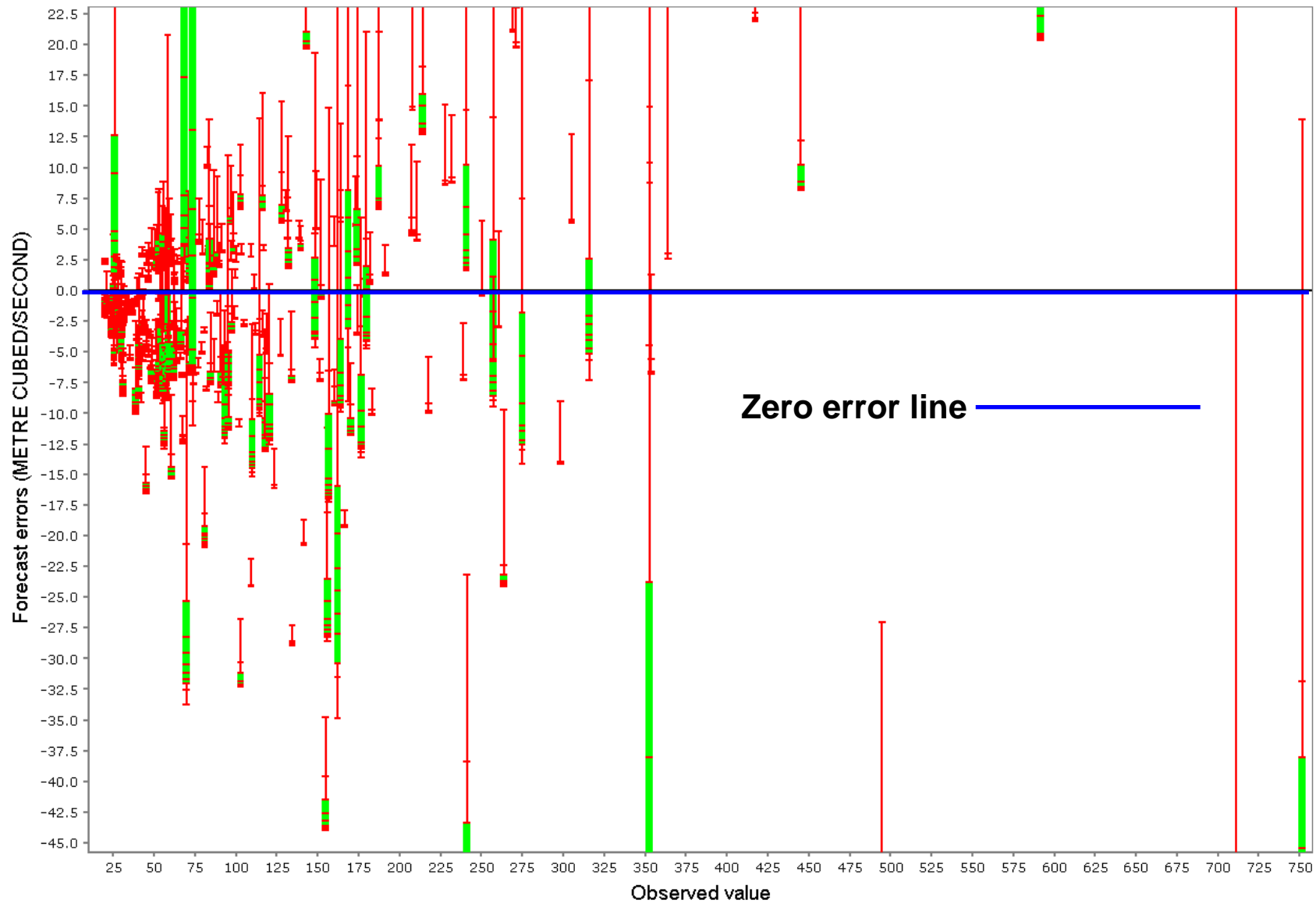




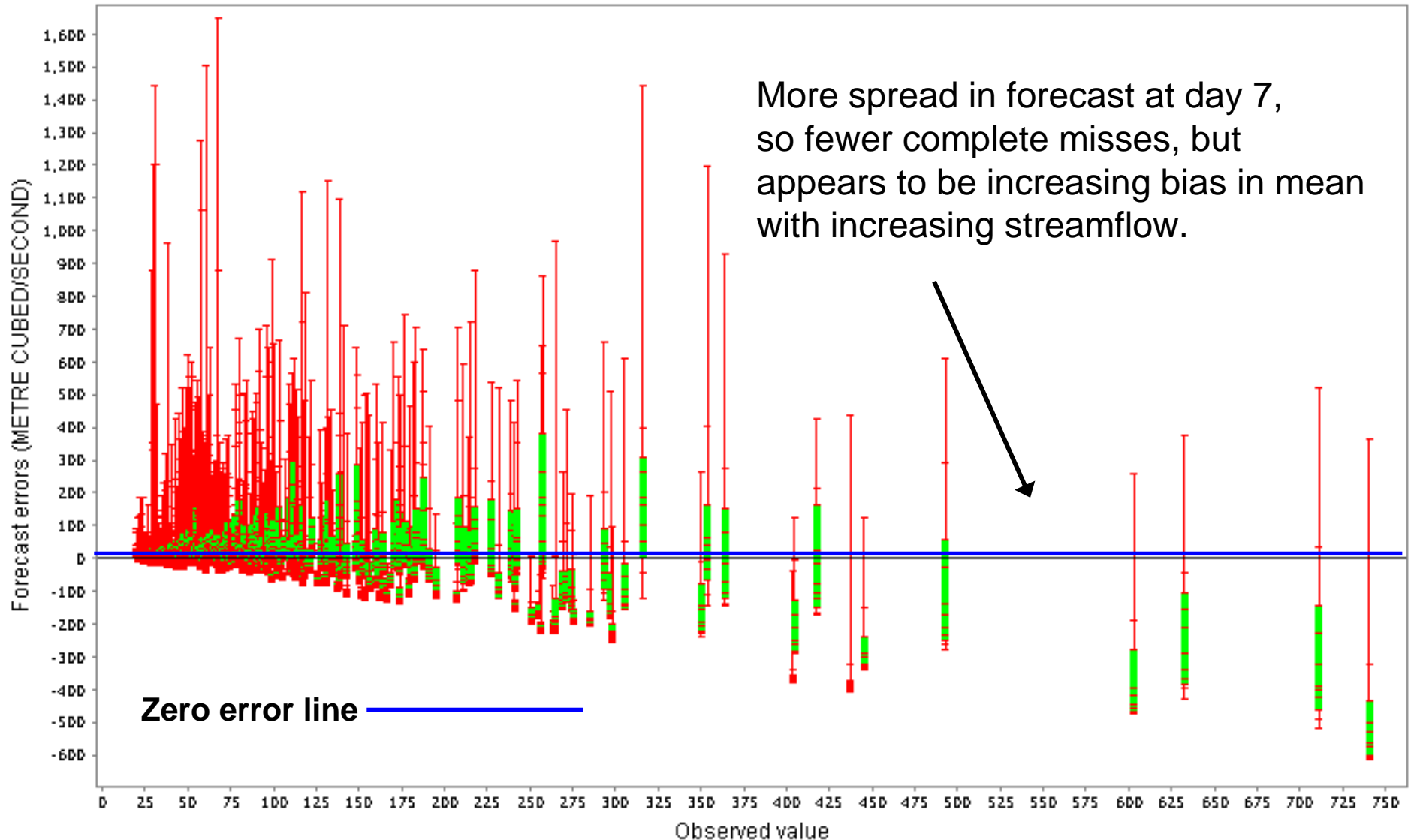
# Box plot by obs. (day 1)



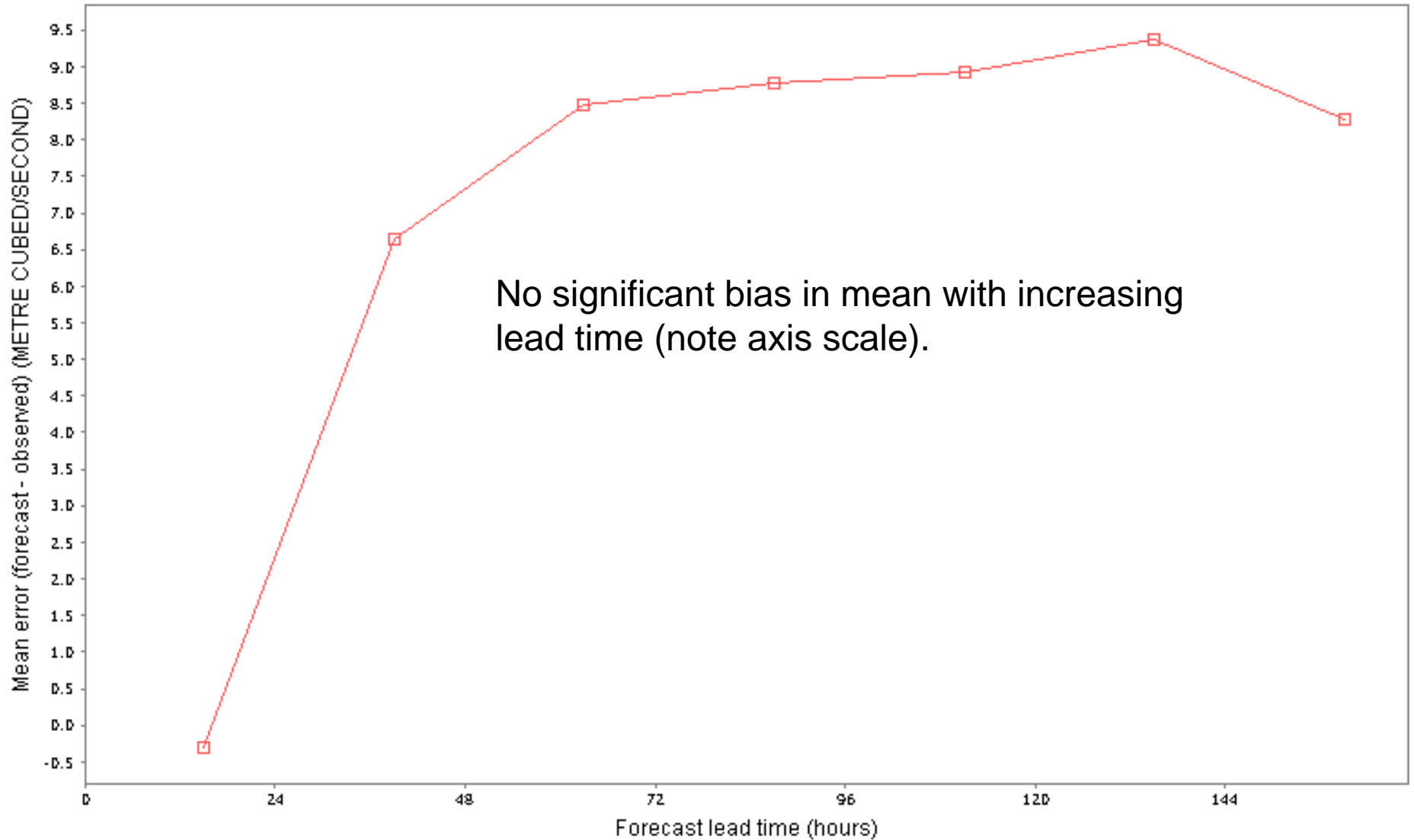
# Box plot by obs. (day 1) zoom



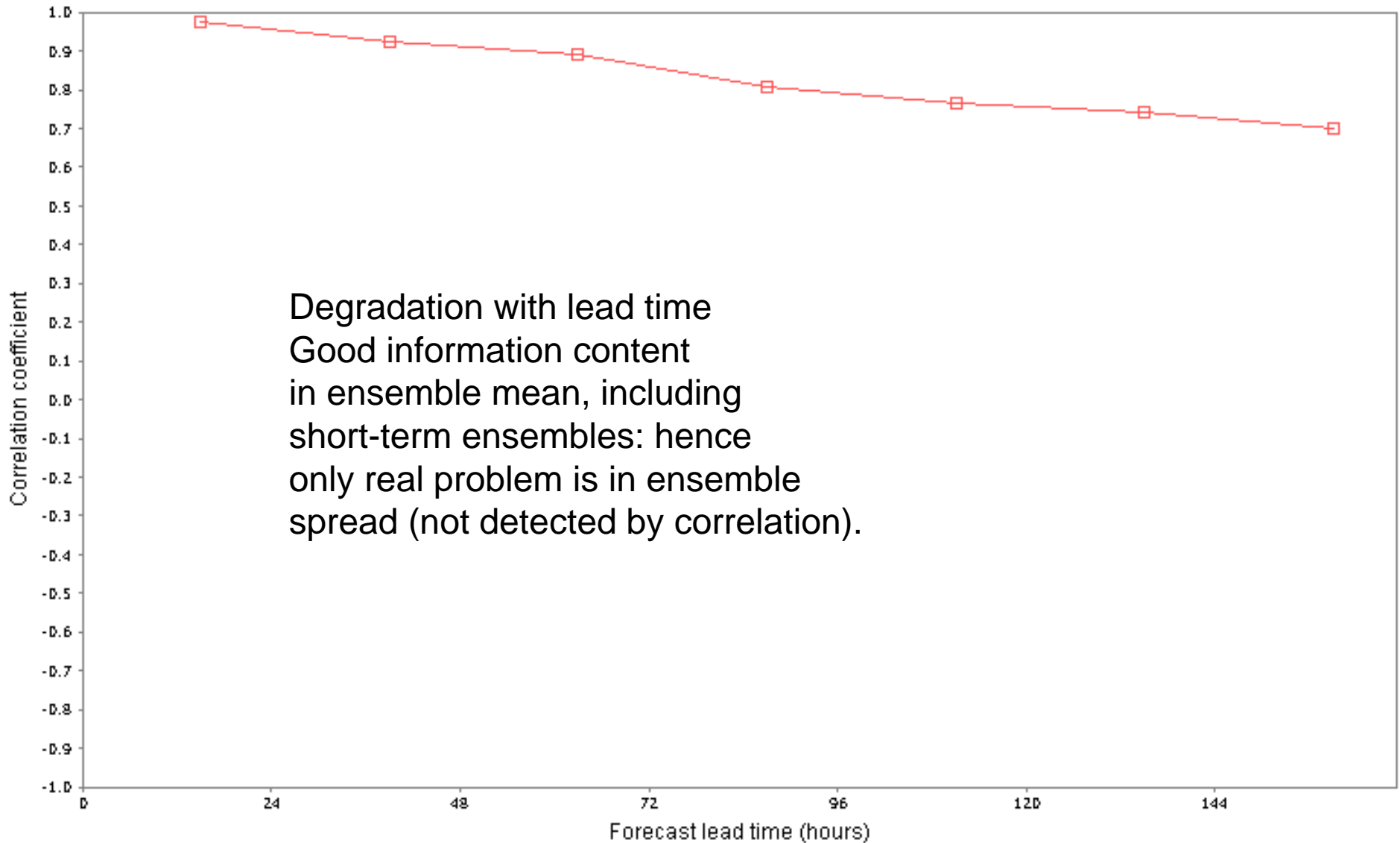
# Box plot by obs. (day 7)



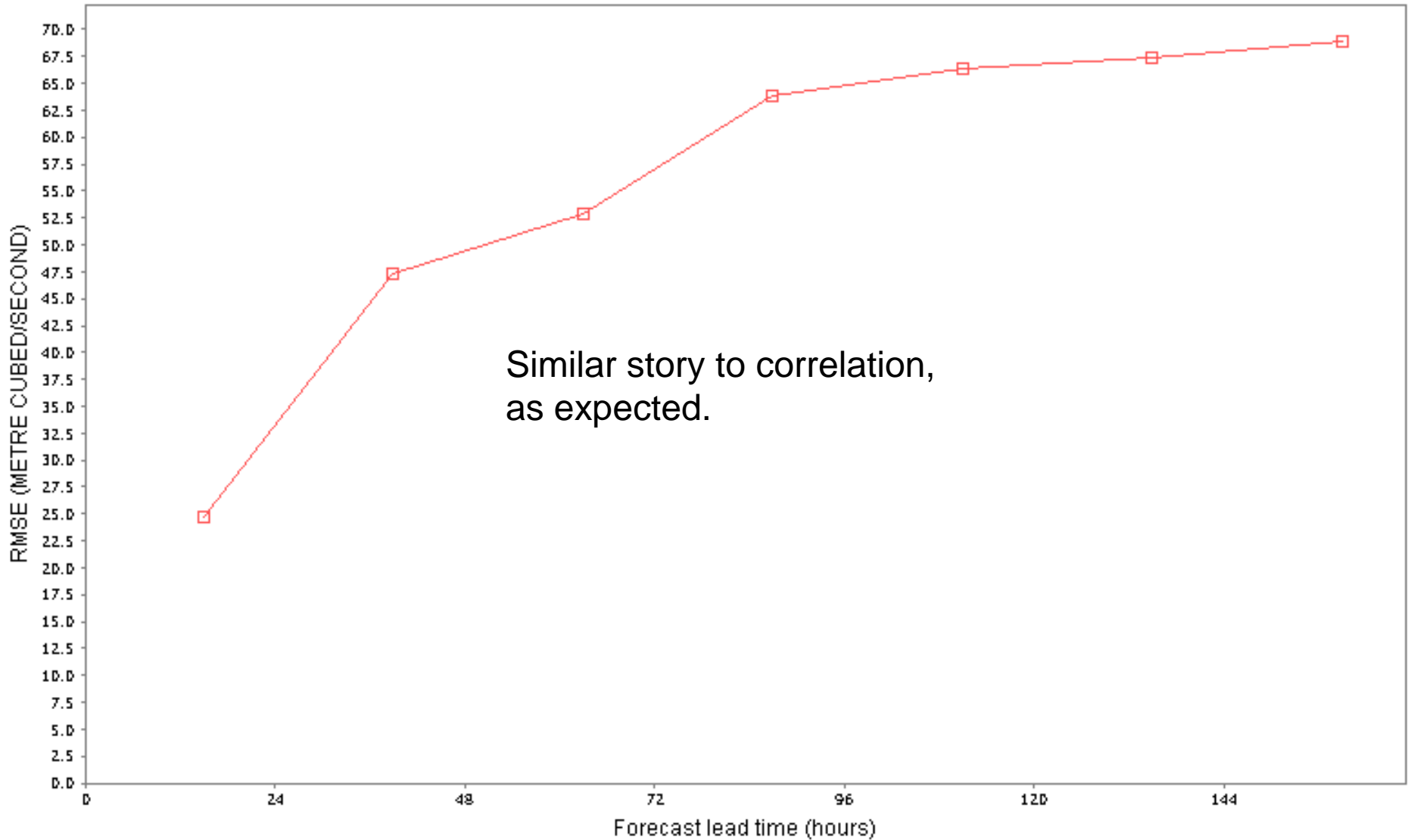
# Mean error of ens. mean



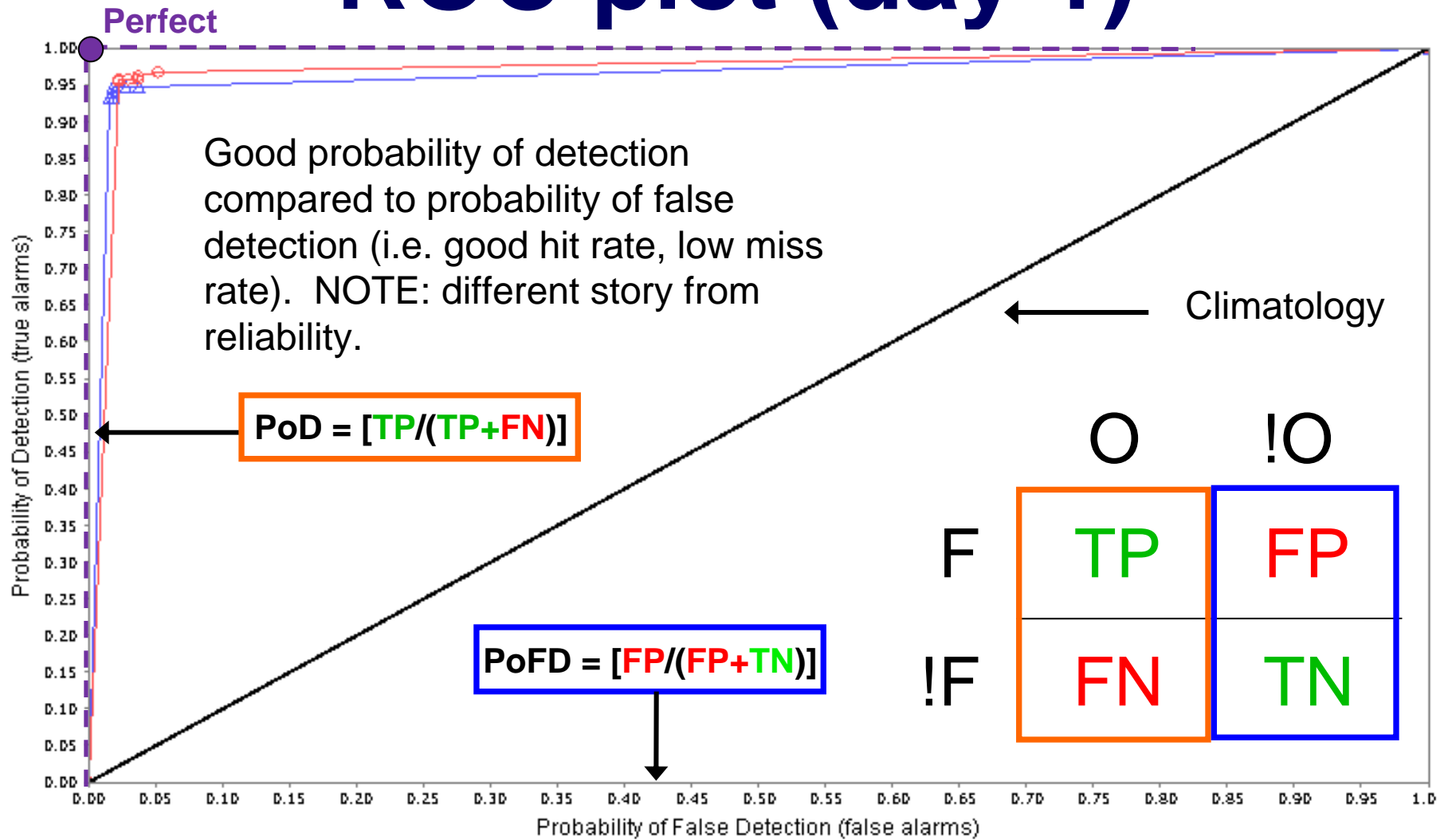
# Correlation of ens. mean



# RMSE of ens. mean



# ROC plot (day 1)

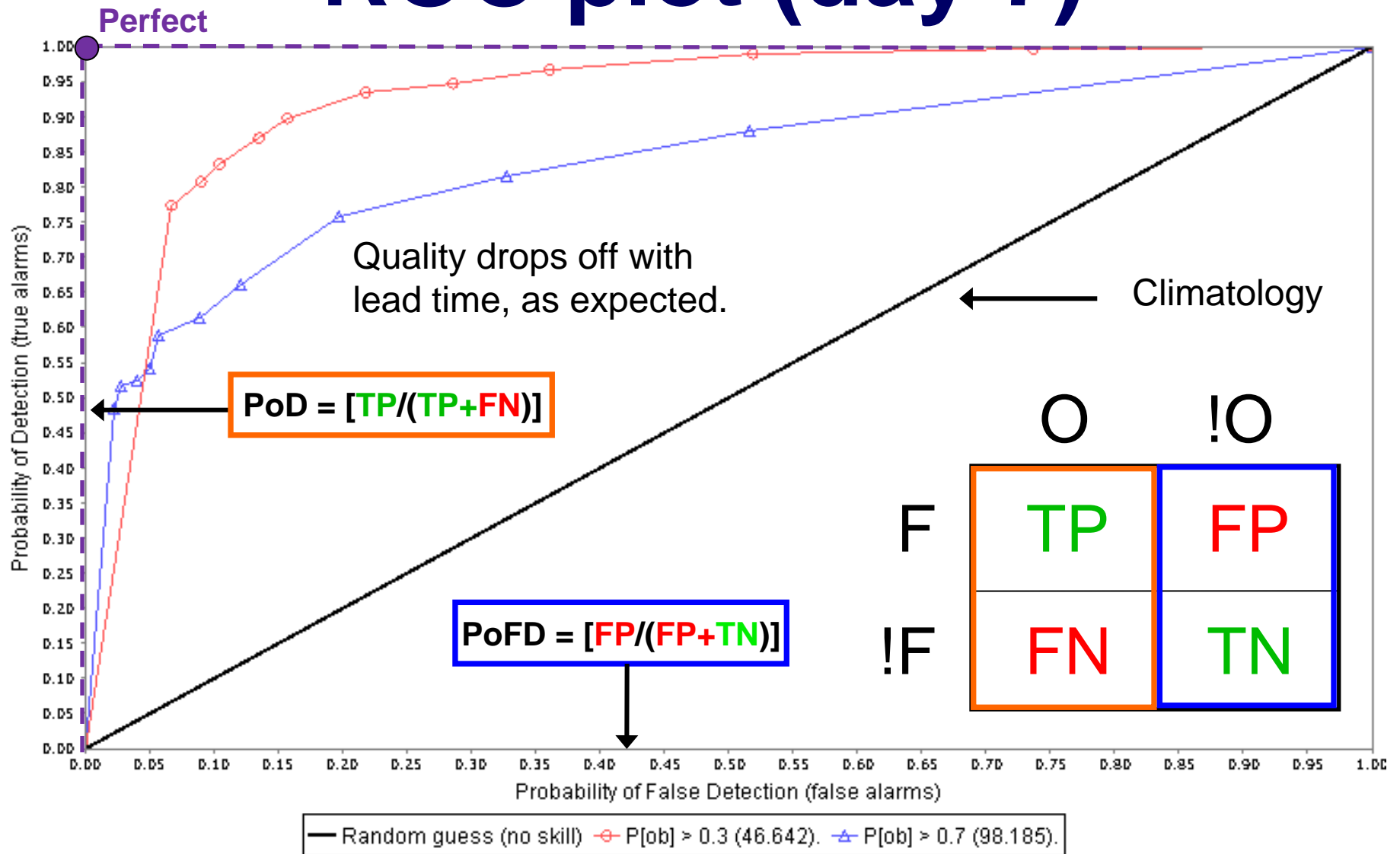


— Random guess (no skill)    ◊ P[ob] > 0.3 (46.642)    ▲ P[ob] > 0.7 (98.185)

1<sup>st</sup> event: Q > 46 cms

2<sup>nd</sup> event: Q > 98 cms

# ROC plot (day 7)







# Questions and discussion