

Hydrologic Ensemble Forecasting Service (HEFS)

Seminar C Exercise 2: Answers

Hank Herr

HEFS Training Workshop (4), Silver Spring, MD
September 19, 2013



Question 1

❑ Question:

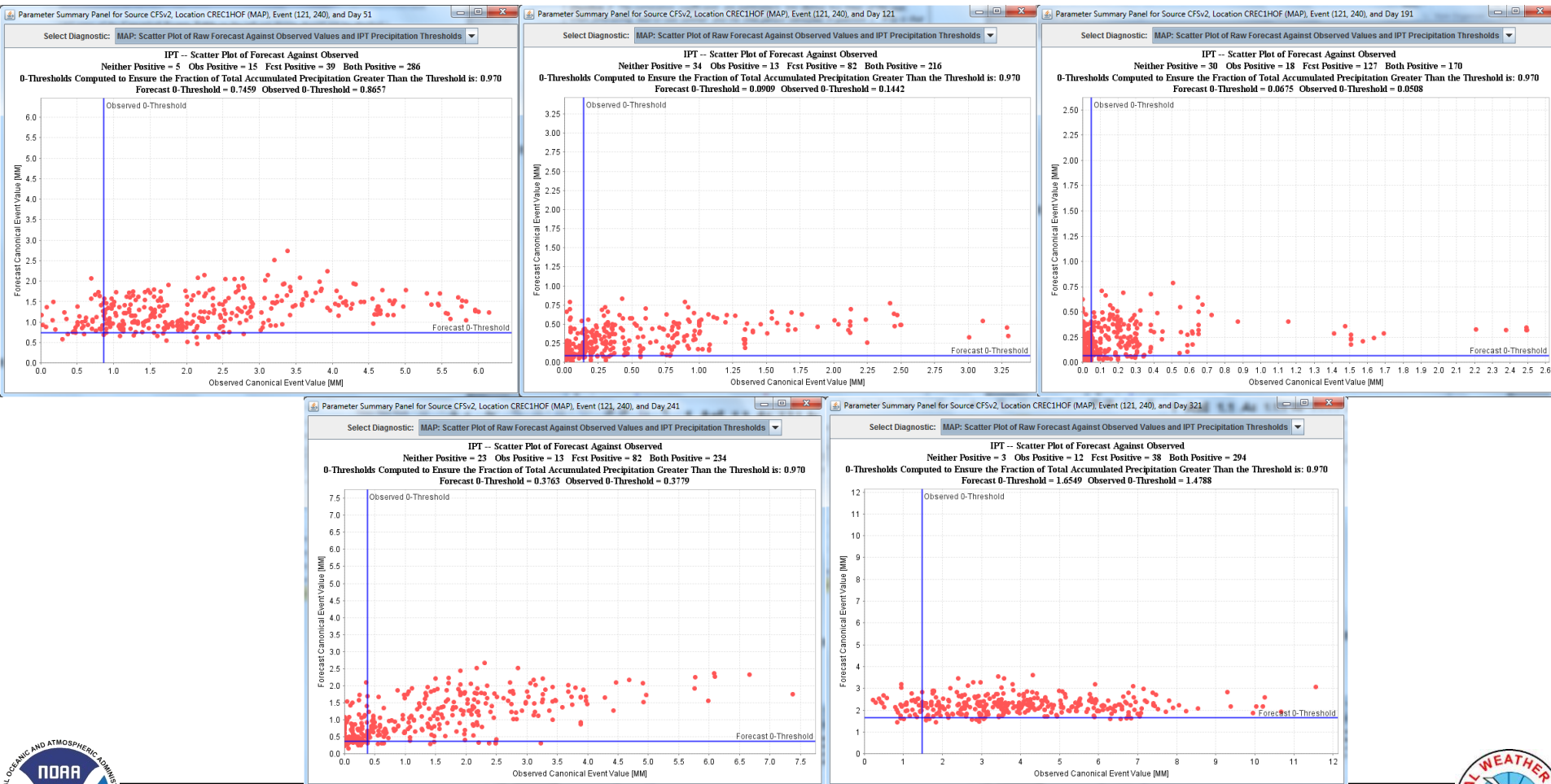
- o CFSv2 displays particularly low correlation (noted by thick stripe of blue and gray blocks) when the 30-day events cover approximately what time period of the year?

❑ Answer

- o Late fall and winter

Question 2

- The following are the resulting scatter plots in order. The lowest correlations occur when the observed values are largest (last plot; day 321) and when the values are lowest and most prone to outliers (day 191). The highest correlation values are when both the observed and forecast values are close to zero; i.e., when the situation is least “interesting”. In general, CFSv2 appears to be unskilled for this location.



Question 3

Question:

- o Do the correlation values appear to satisfy the expected pattern relative to the lead time; that is, does the correlation decrease as lead time increases?

Answer

- o Yes

Question 4

❑ Question:

- o During which times of the year is the lead time pattern most easily identifiable? Why is that?

❑ Answer:

- o Cool Season
- o More positive precipitation values are available, yielding larger sample sizes and more reliable computed correlations

Question 5

□ Question

- When does increasing the maximum width stop having an impact on the parameters? Why is that?

□ Answer

- 365
 - For 181, there are no longer any values blocks for which more pairs are needed.

Question 6

□ Question:

- For which maximum window width do the correlation coefficients appear to be computed with “enough” data? If you determined, subjectively, that four months (121 days) is the largest allowable season, are there any times of the year for which the correlation coefficient does not appear to be estimated with enough data?

□ Answer

- 151 days seems reasonable
- No
 - For 121, there are issues in the summer (check the dark blue correlation for period (4,4))

Question 7

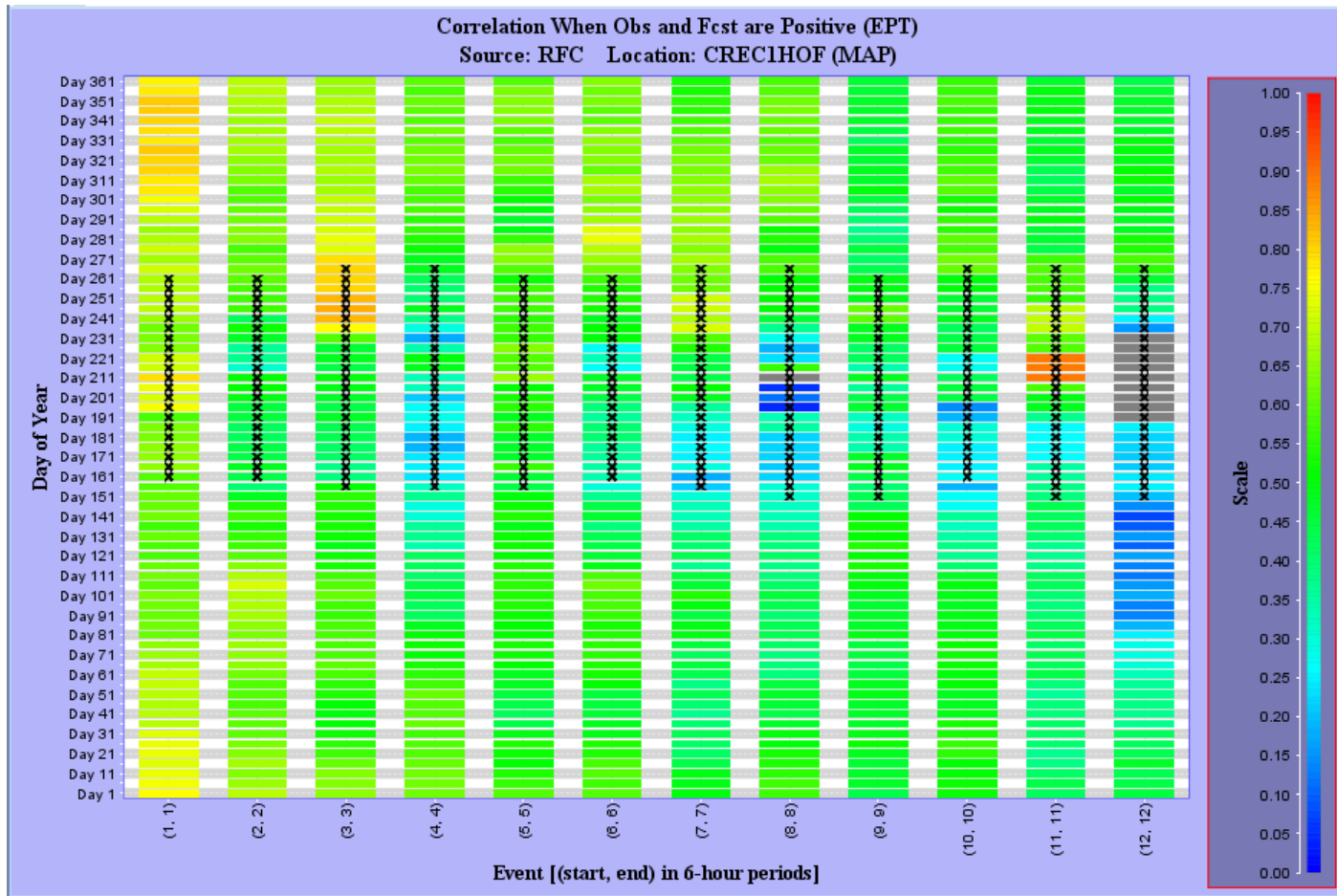
□ Question:

- o With the increased maximum window widths, are there any identifiable patterns to the correlation coefficient values by season or lead time? Can they be explained?

□ Answer:

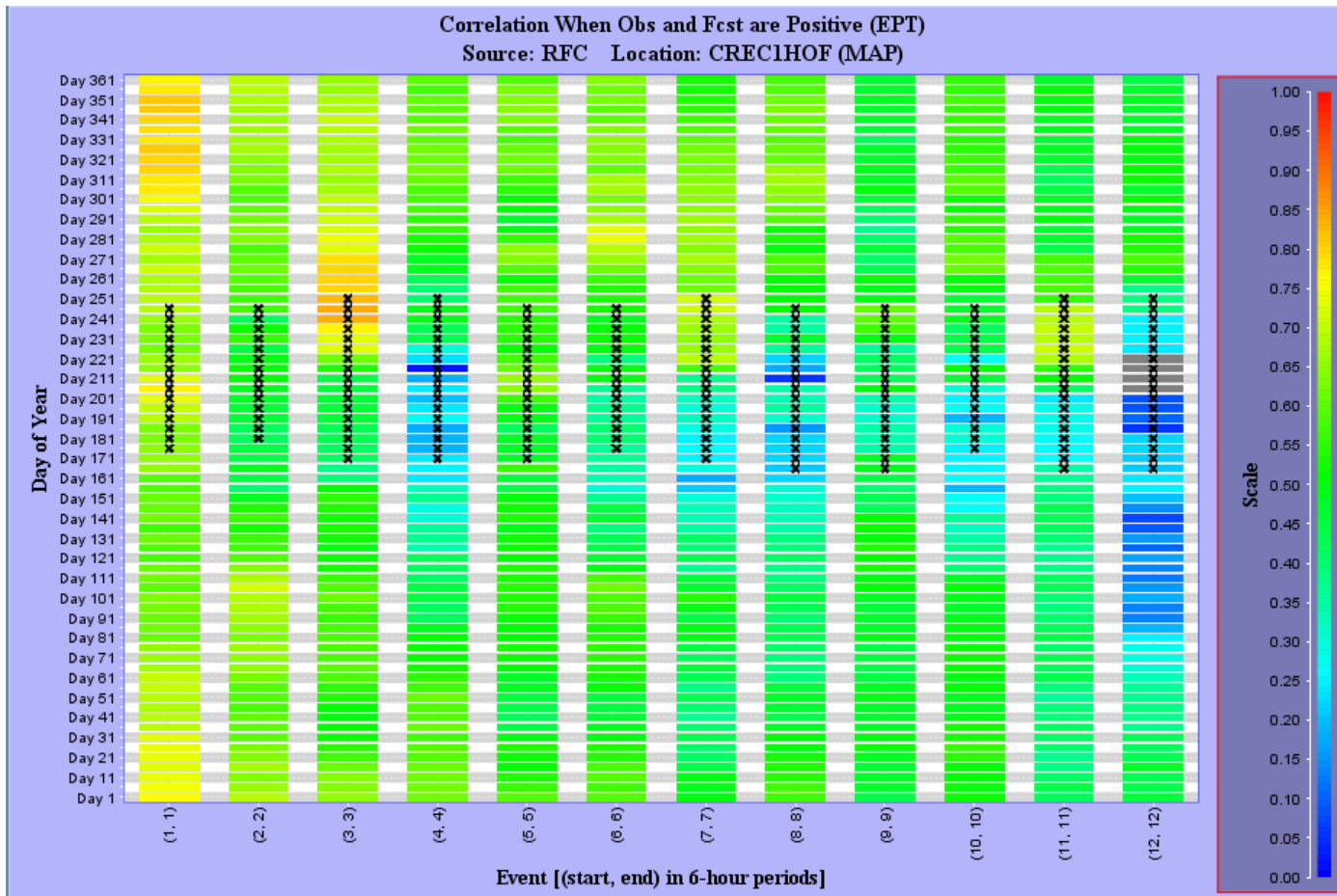
- o Season
 - Correlation is higher in cool months
- o Lead Time
 - The correlation decreases as expected. Unexpectedly, events 4, 8, and 12 are noticeably worse than the other events during the dry season. I think it might be due to an outlier that occurs on 10/19/2007 12Z (88 MM of rain; more than twice the next closest observed value on the scatter plots).

Questions 5 - 7 Plots



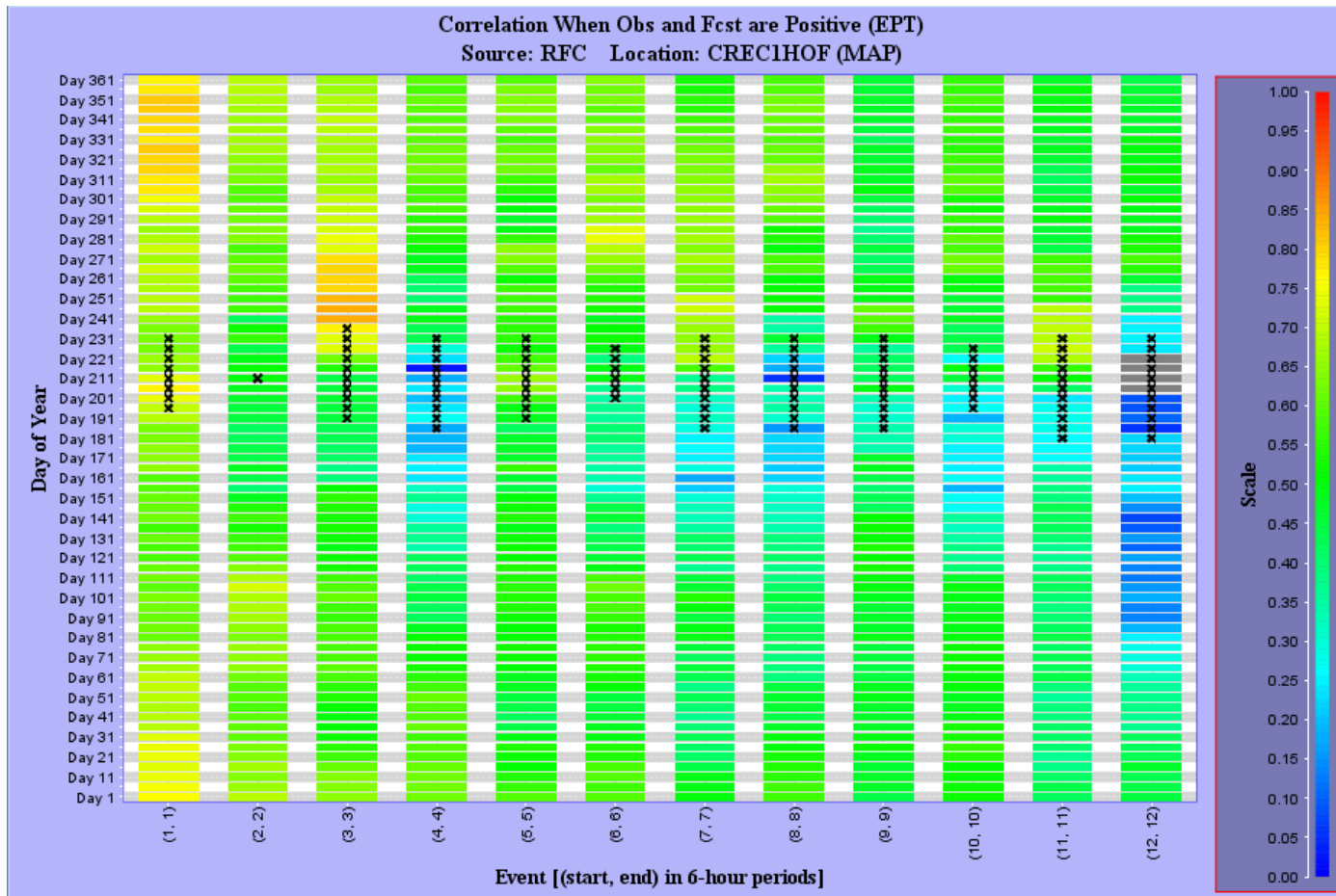
91 days

Questions 5 - 7 Plots



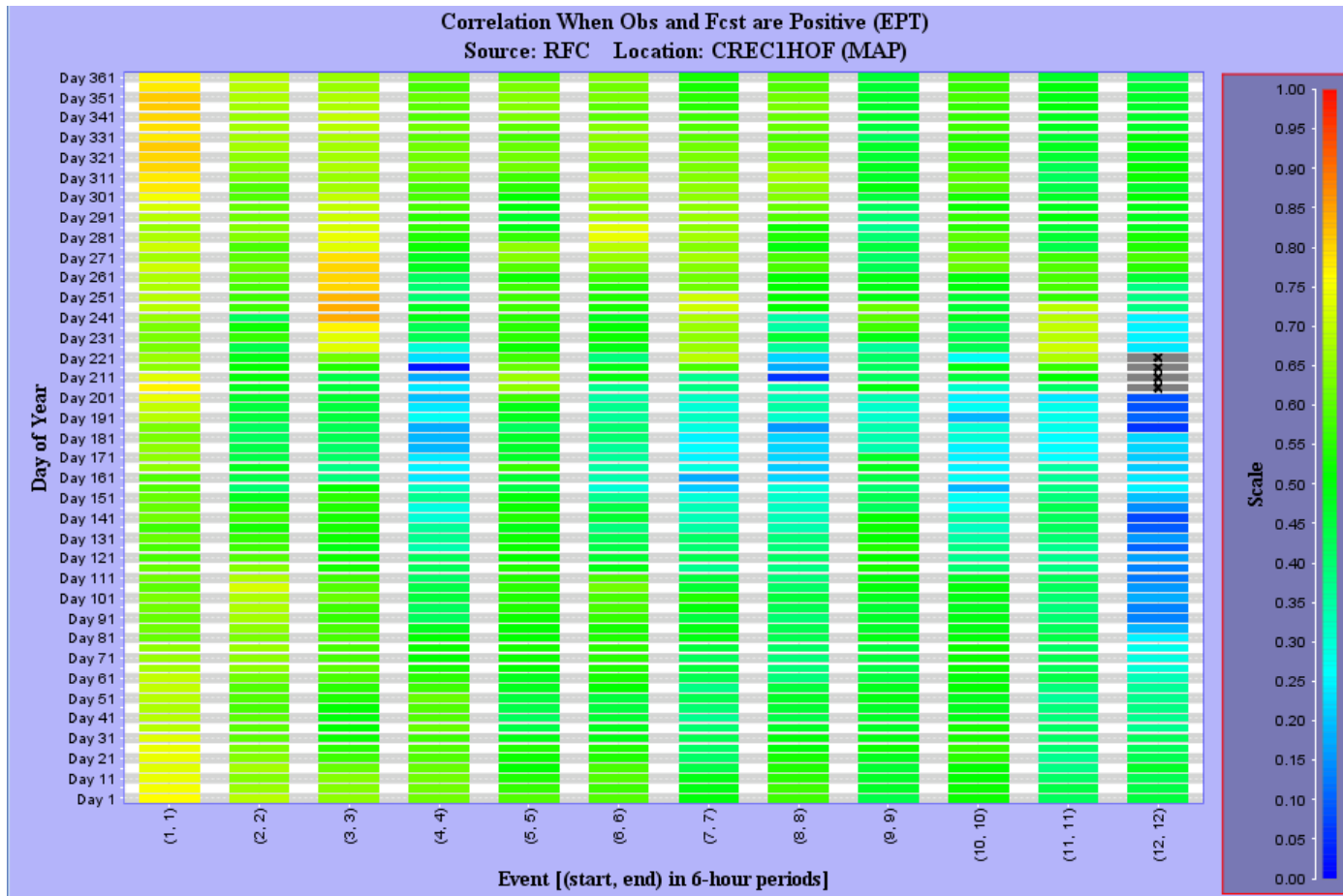
121 days

Questions 5 - 7 Plots



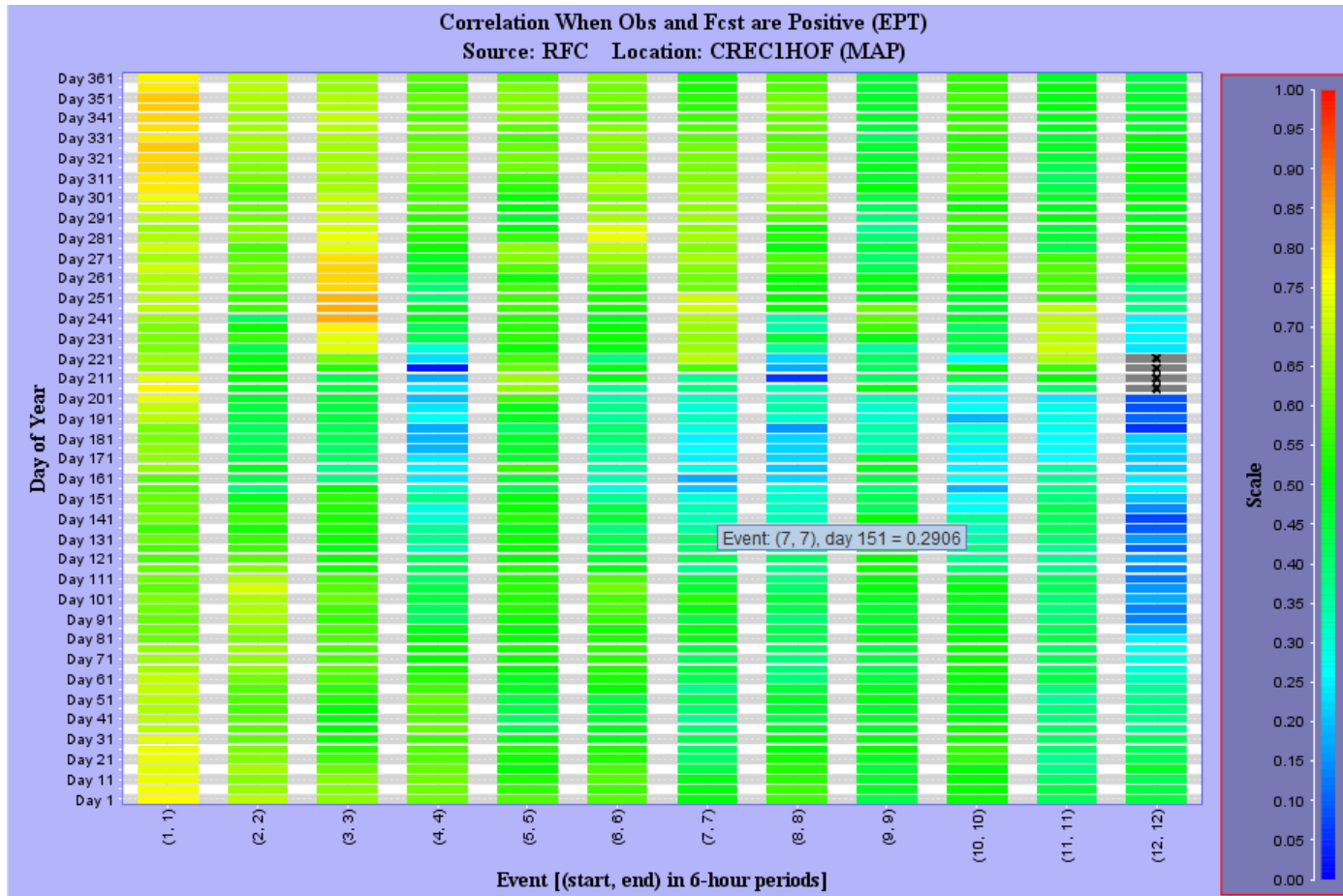
151 days

Questions 5 - 7 Plots



181 days

Questions 5 - 7 Plots



365 days

Question 8

❑ Question:

- o For which sources and which times of the year do the correlation parameters appear unreliable, likely due to small sample size? As the duration of canonical events increases, is there any change in the apparent reliability of the parameters?

❑ Answer

- o Source and times of year with unreliable correlations:
 - RFC – Summer
 - GEFS – Slightly suspect in July with sample sizes around 25-35
 - CFSv2 – focusing on events beyond 15 days (start period > 60), there appear to be no reliability issues
- o Impact from duration increasing
 - As duration increases, more positive (forecast, observed) pairs are found and parameters appear more reliable

Question 9

□ Question:

- *Suppose that the maximum window width cannot be increased any further. For each forecast source, do you believe that the source should be used in operational forecasting? Does that apply throughout the year? When used, what lead times and canonical events, if any, do you believe the forecast source will impact most significantly? How many forecast days should be used operationally?*

□ Answer:

- Volunteers?

Question 10

❑ Question:

- o Why is there no discernible seasonal pattern (or, at best, a very weak pattern)? Does it appear to be related to sample size?

❑ Answer

- o Why: CFSv2 has very little skill for this location
- o Sample size does not appear to be an issue at a 30-day duration

Question 11

❑ Question:

- o Do the correlation values appear to satisfy the expected pattern relative to the lead time; that is, does the correlation decrease as lead time increases? Are there any seasonal patterns?

❑ Answer:

- o The correlation parameter does decrease with lead time
 - It is not a strong relationship
- o GEFS appears to have higher correlation values at a 6-hour time step in the cool season

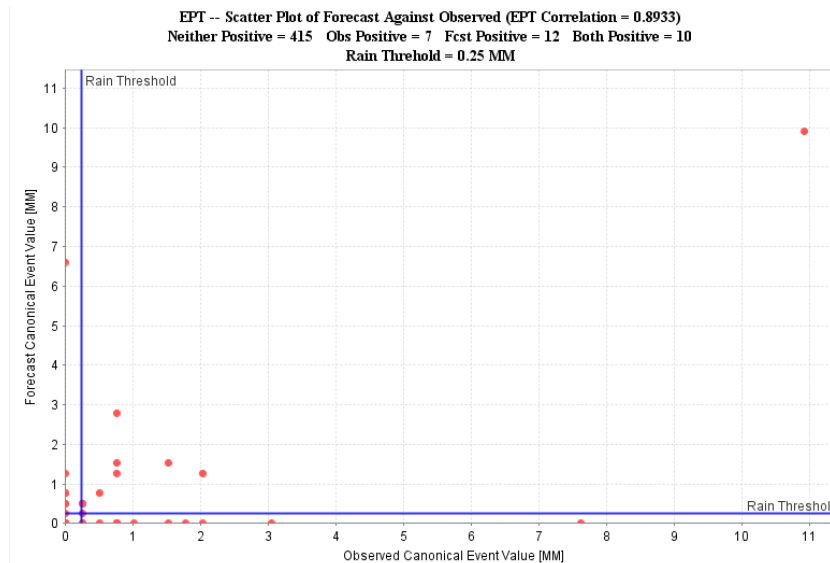
Question 12

❑ Question:

- o Does the sample size appear to be insufficient to estimate the correlation coefficient for any of the blocks shown through day 2?

❑ Answer:

- o The sample size is insufficient for the reddish blocks
 - Example:



- o Some gray blocks are also insufficient or borderline

Question 13

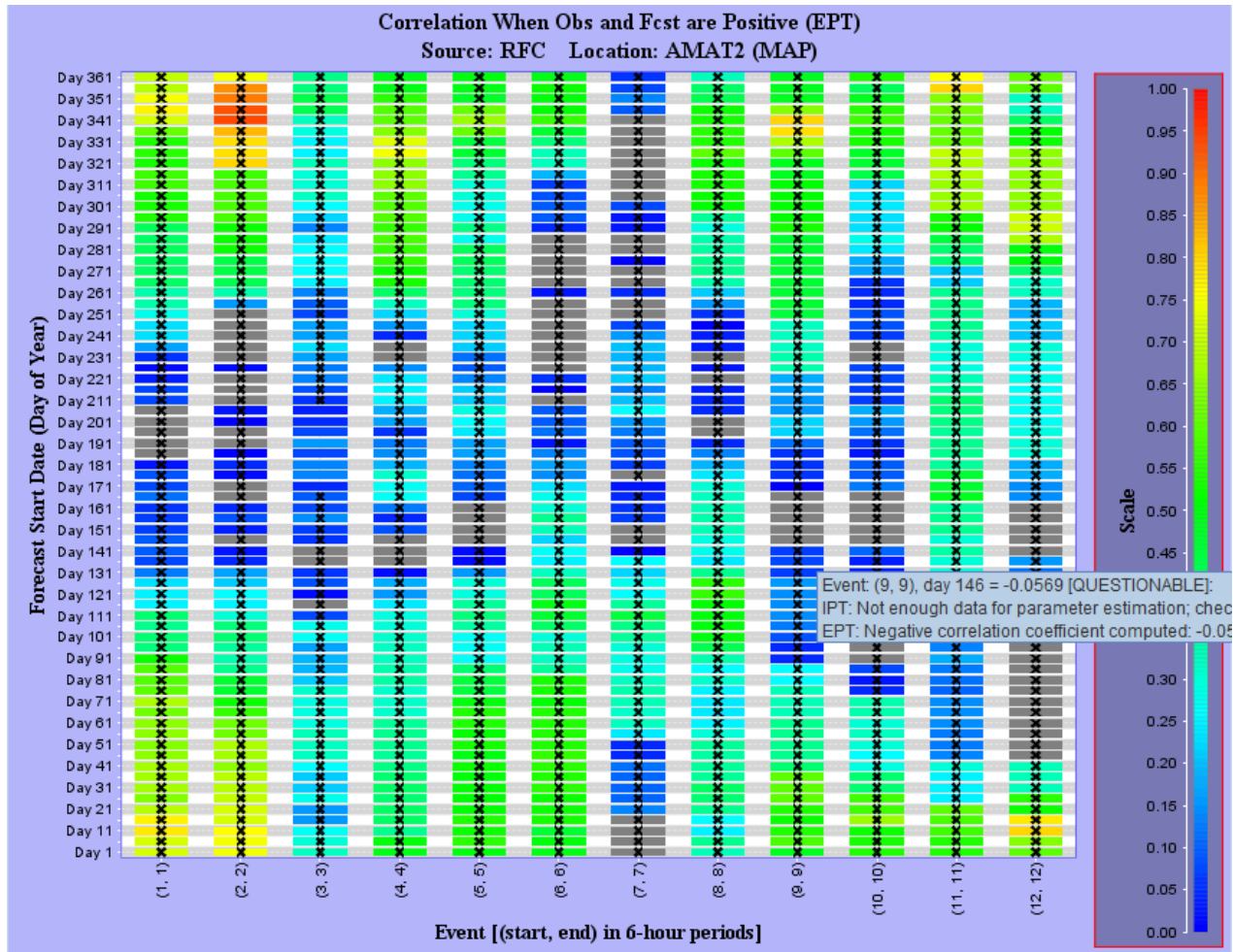
□ Question:

- For which maximum window width do the correlation coefficients through day 2 appear to be computed with “enough” data? If you determined, subjectively, that four months (121 days) is the largest allowable season, are there consistent times of the year for which the correlation coefficient does not appear to be estimated with enough data?

□ Answer:

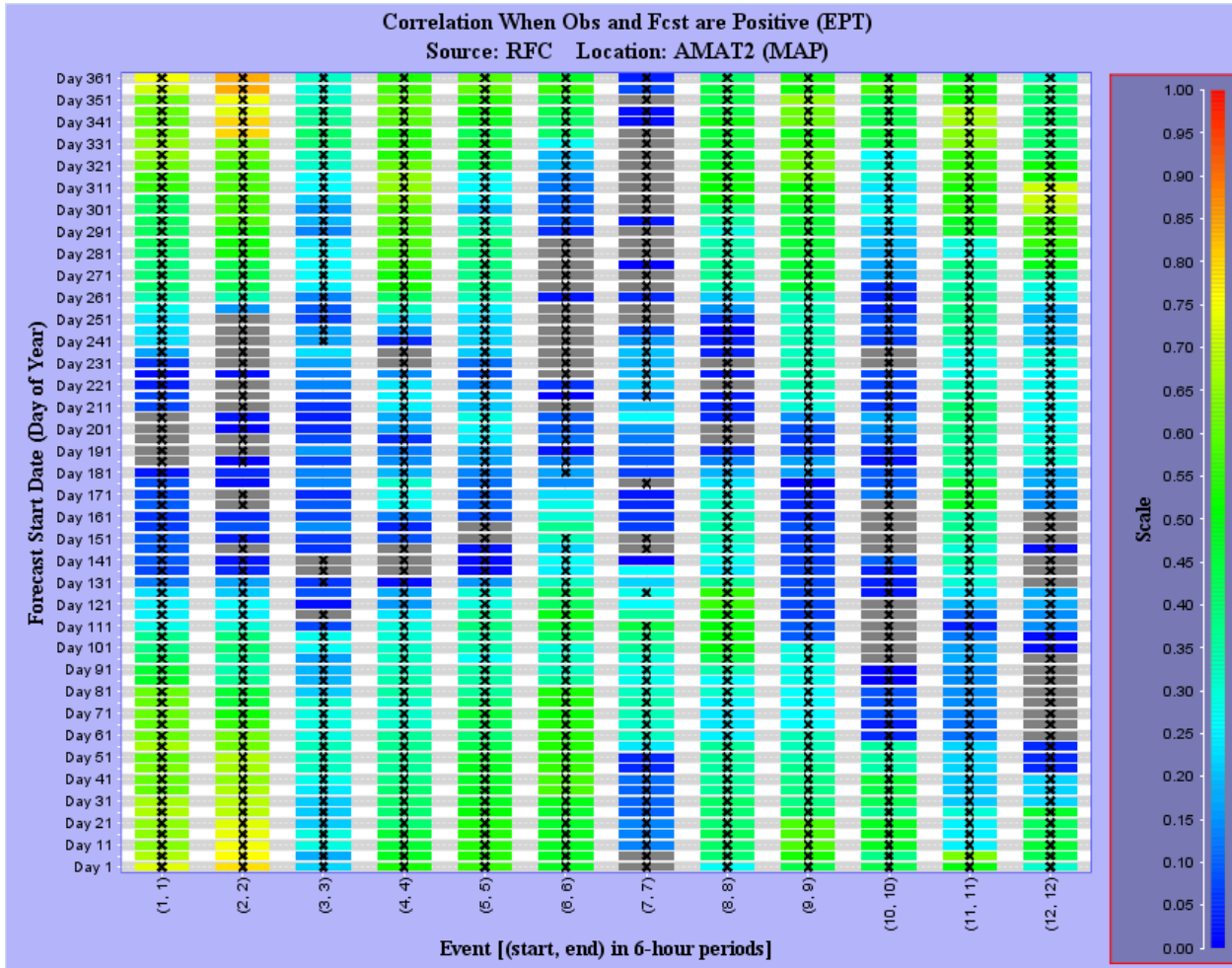
- 151 appears reasonable
 - 121 is not good based on, again, the dark orange/red blocks.
- Cool season appears to have small sample sizes.

Question 13



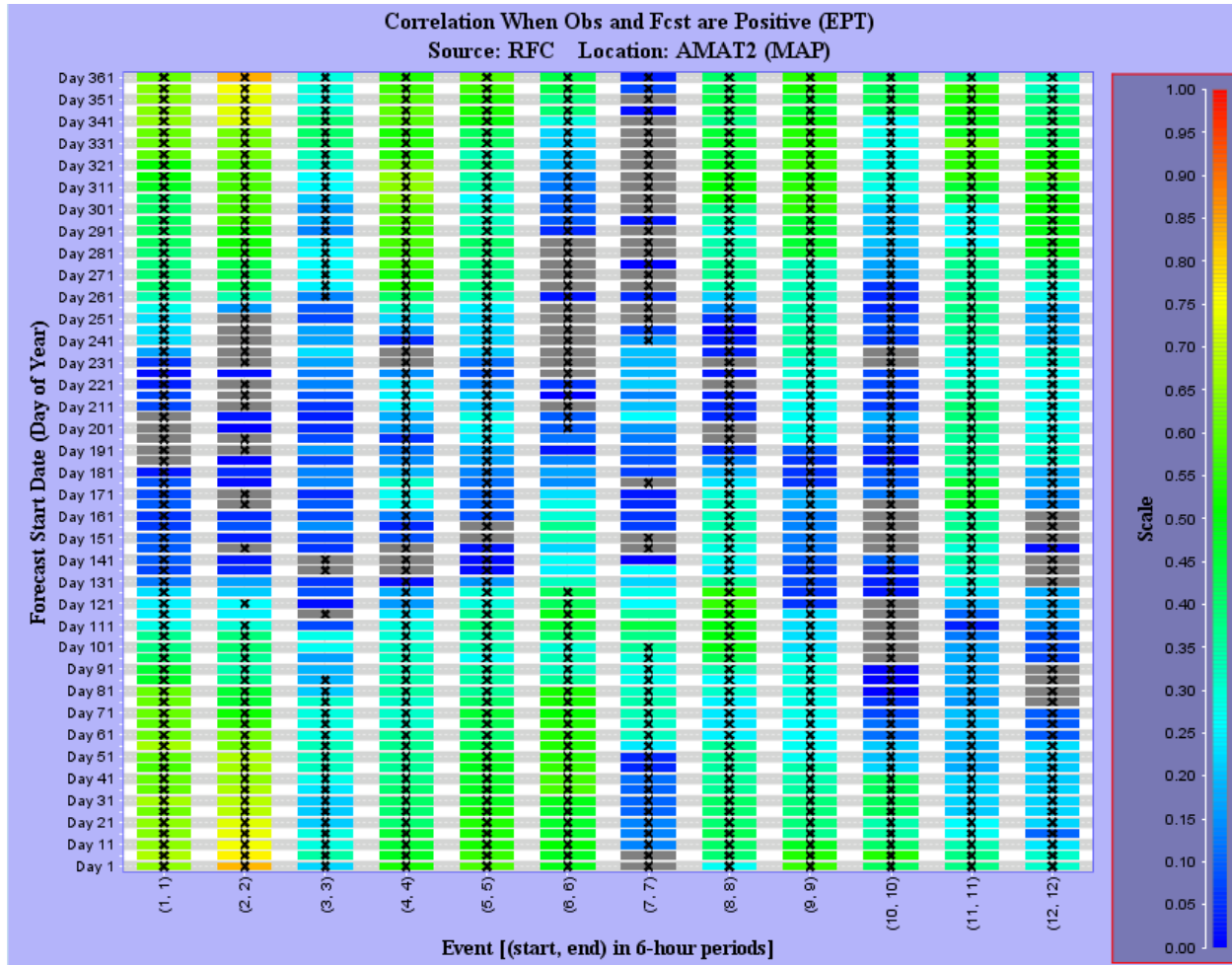
121 days

Question 13



151 days

Question 13



181 days

Question 14

❑ Question:

- During the cool season, are there any lead times that have lower correlations than the other lead times? Why might that be?

❑ Answer:

- Lead times 3 and 7
 - There is an outlier on Jan 1, 2005 (obs = 119.38 MM, fcst = 0.508 MM)

Question 15

❑ Question:

- o *For which sources and which times of the year do the correlation parameters appear unreliable, likely due to small sample size? As the duration of canonical events increases, is there any change in the apparent reliability of the parameters?*

❑ Answer:

- o Source and times of year with unreliable correlations:
 - RFC – All year, but cool season is worse.
 - GEFS – No reliability issues.
 - CFSv2 – focusing on events beyond 15 days (start period > 60), there appear to be no reliability issues
- o Impact from duration increasing
 - As duration increases, more positive (forecast, observed) pairs are found and parameters appear more reliable

Question 16

Question:

- o *Suppose that the maximum window width cannot be increased any further. For each forecast source, do you believe that the source should be used in operational forecasting? Does that apply throughout the year? When used, what lead times and canonical events, if any, do you believe the forecast source will impact most significantly? How many forecast days should be used operationally?*

Answer:

- o Volunteers?

Questions? Comments?

