Analysis: Part II

Paige Miller

11/3/2017

Analysis Goals:

1. Data exploration and descriptive statistics
2. Characterize daily patterns of SPC, temperature, and conductivity at each site
3. Calculate dissimilarity between daily patterns at each site using Hierarchical clustering
4. Cross-correlation between stage height and conductivity (what the lag is and how related the time series are).
5. Help write analysis portion of methods section.

4. Cross-correlation between stage height and conductivity

We first excluded known probe errors. Joined with conductivity data by date. For now, I only have water level for:

- Brooklyn
- Carr
- Tanyard
- Trail
- Shoal
- Turkey

Missing for:

- Brick
- McNutt
- Tallassee
- Big

Table 1: Instances of negative water level observations removed prior to analysis.

name	DateTime	WaterLevel
Bear	2016-01-16 08:05:00	-2.692
Bear	2016-01-16 08:10:00	-2.427
Bear	2016-01-16 08:15:00	-2.323
Shoal	2016-04-14 09:05:00	-0.025
Turkey	NA	-29.503
Turkey	NA	-30.459
Turkey	NA	-29.560

Water level descriptives

We have observations ranging from 2015-06-18 21:15:00 to 2017-07-07 09:20:15. But not all sites were observed at all point in that range. Only Carr has measurements in 2017.

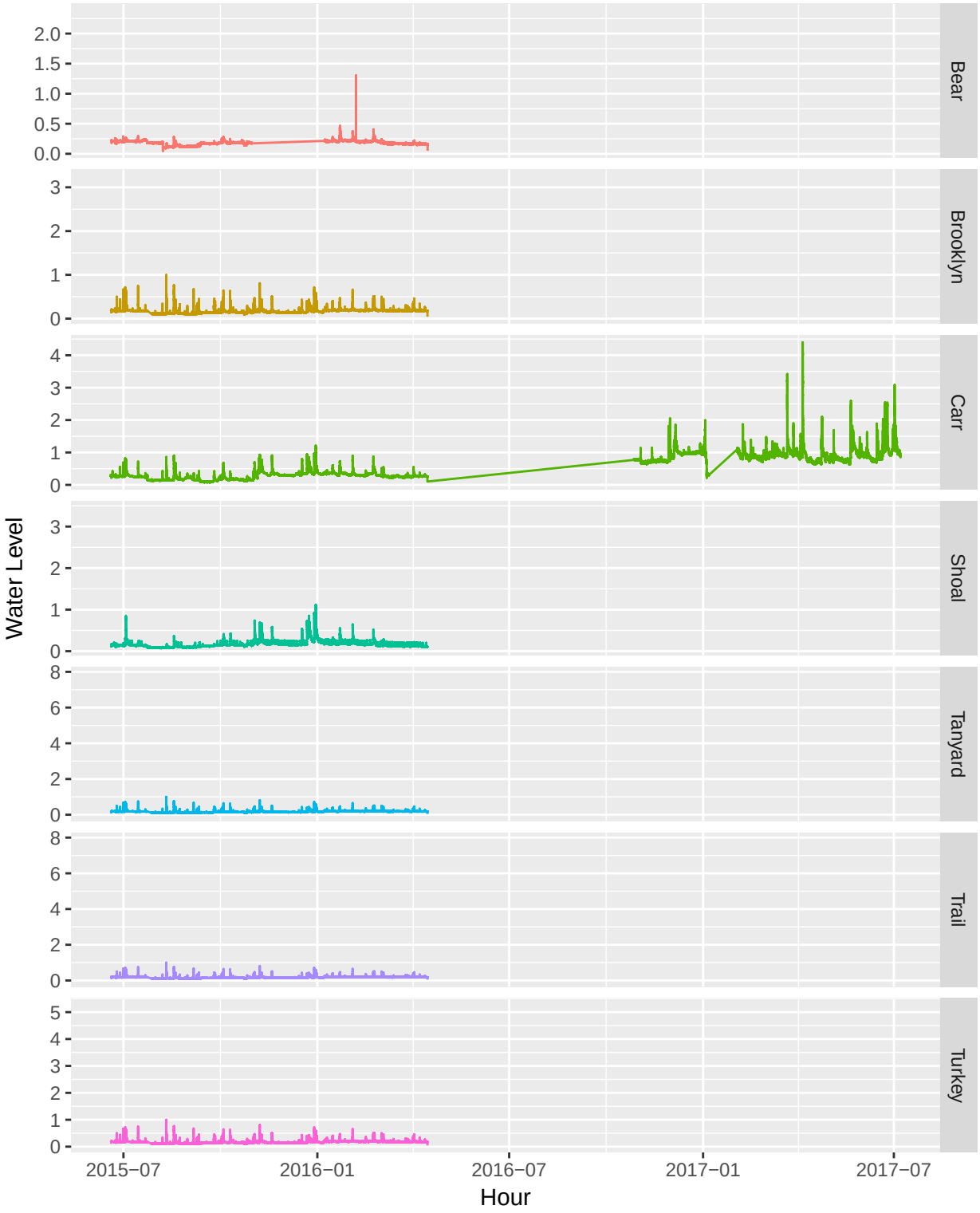

Figure 1: Water level by site.

Table 2: Number of water level observations per site (taken at 5 minute intervals).

Site_Name	Observations
Brooklyn	224805
Shoal	211382
Turkey	205750
Carr	204313
Tanyard	200615
Trail	192581
Bear	182353
McNutt	86647
Brickyard	86568
Tallassee	85231

Table 3: Mean and SD for water level and barom by by site.

name	mean_water	sd_water
Bear	0.31	0.16
Brooklyn	0.36	0.24
Carr	0.55	0.36
Shoal	0.37	0.24
Tanyard	0.30	0.23
Trail	0.30	0.23
Turkey	0.50	0.39

The water level is highest at Carr and lowest at Trail. The standard deviation in water level was highest at Turkey and lowest at Bear.

One way to standardize water level data might be to use Box Cox transformation. Box Cox test indicated that a negative power transformation would be appropriate. I can do this, but I think it may not be necessary (can have a chat about this later).

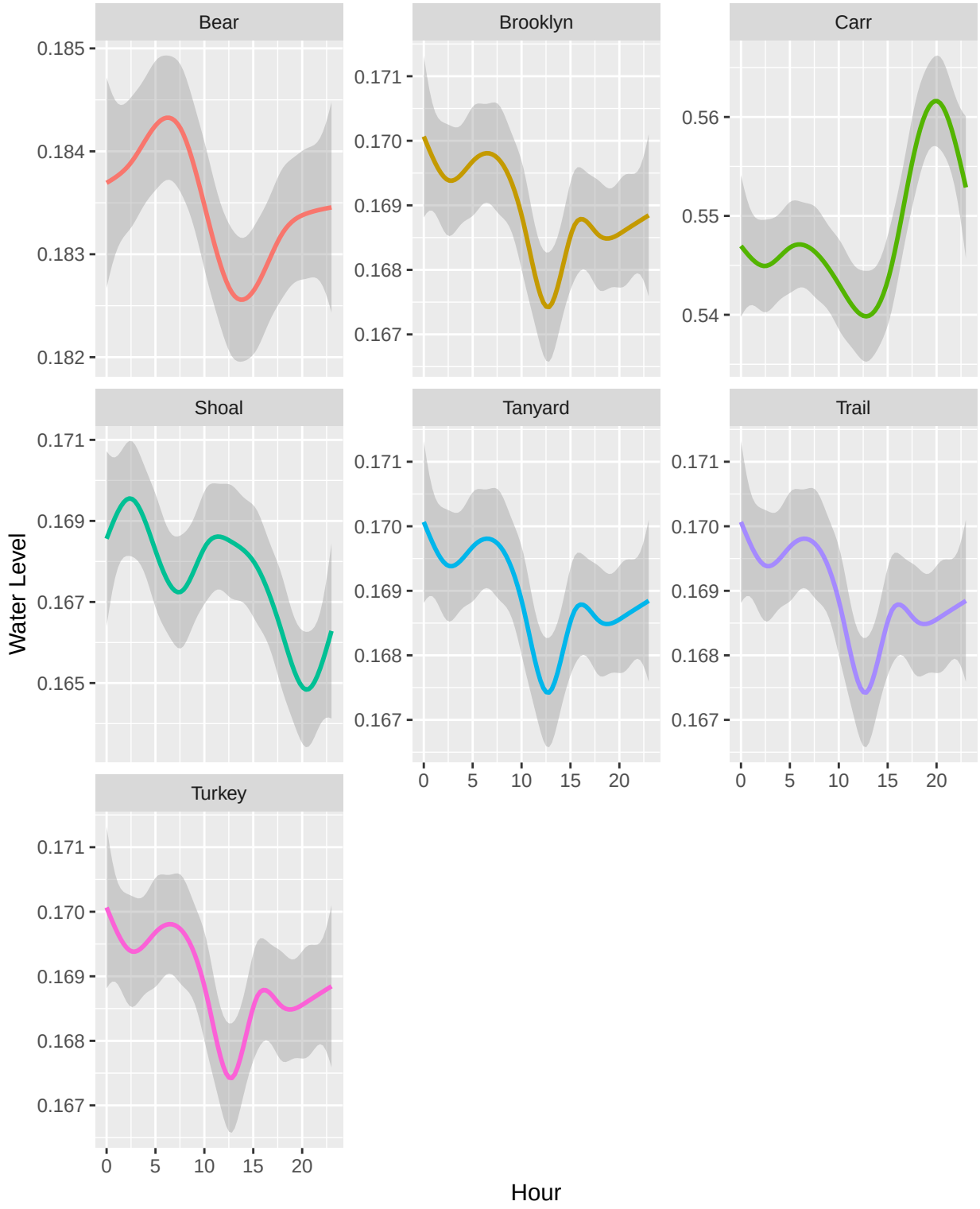

Figure 2: Water level by site aggregated by hour (and averaged each hour).

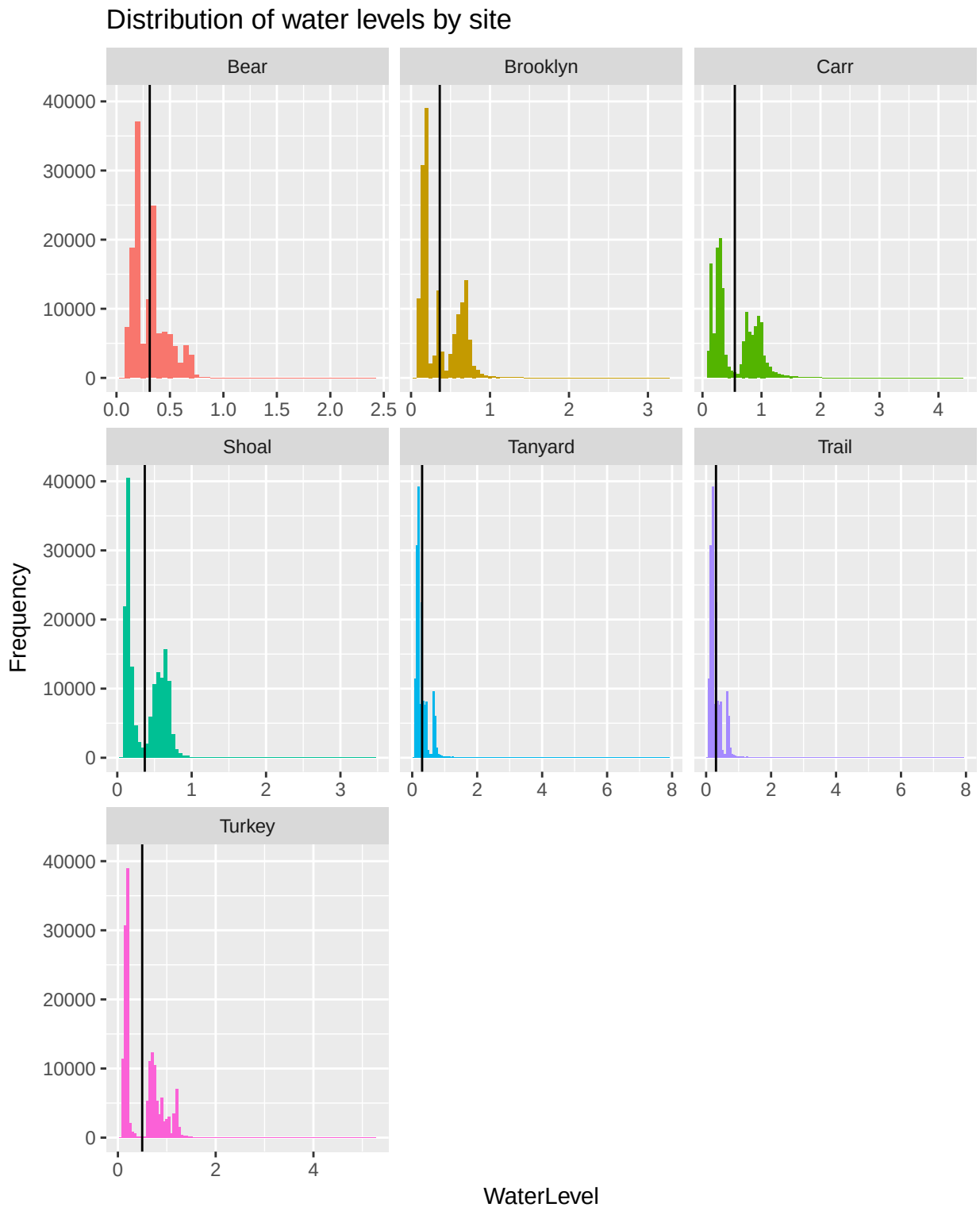

Figure 3: Since water level distributions are heavily right skewed (mean pulled to the right), normalizing the curves is causing there to be negative values. Black lines represent means.

Cross correlation with water level and SPC: Carr

The basic problem we're considering is the description and modeling of the relationship between stage height (water level) and SPC.

In the relationship between two time series (yt and xt), the series yt may be related to past lags of the x-series. The sample cross correlation function (CCF) is helpful for identifying lags of the x-variable that might be useful predictors of yt.

A negative value for h is a correlation between the x-variable at a time before t and the y-variable at time t. For instance, consider h of neg. 2. The CCF value would give the correlation between $x_t - 2$ and y_t.

First let's just plot a chunk of the Carr time series from 2016-01-01 to 2016-02-01. We know we have both conductivity and water level measurements during this time interval.

Clearly our data are not stationary (i.e., variables at time t are related to variables at time t-1, t-2, ...). We can first difference to transform a non-stationary series to a (at least a weakly) stationary series apt for assessing cross-correlation.

First differenced water level plot of Carr from 2016-01-01 to 2016-02-01 shows obvious correlation. Spikes in water level seem to be followed by dips in SPC then rises in SPC.

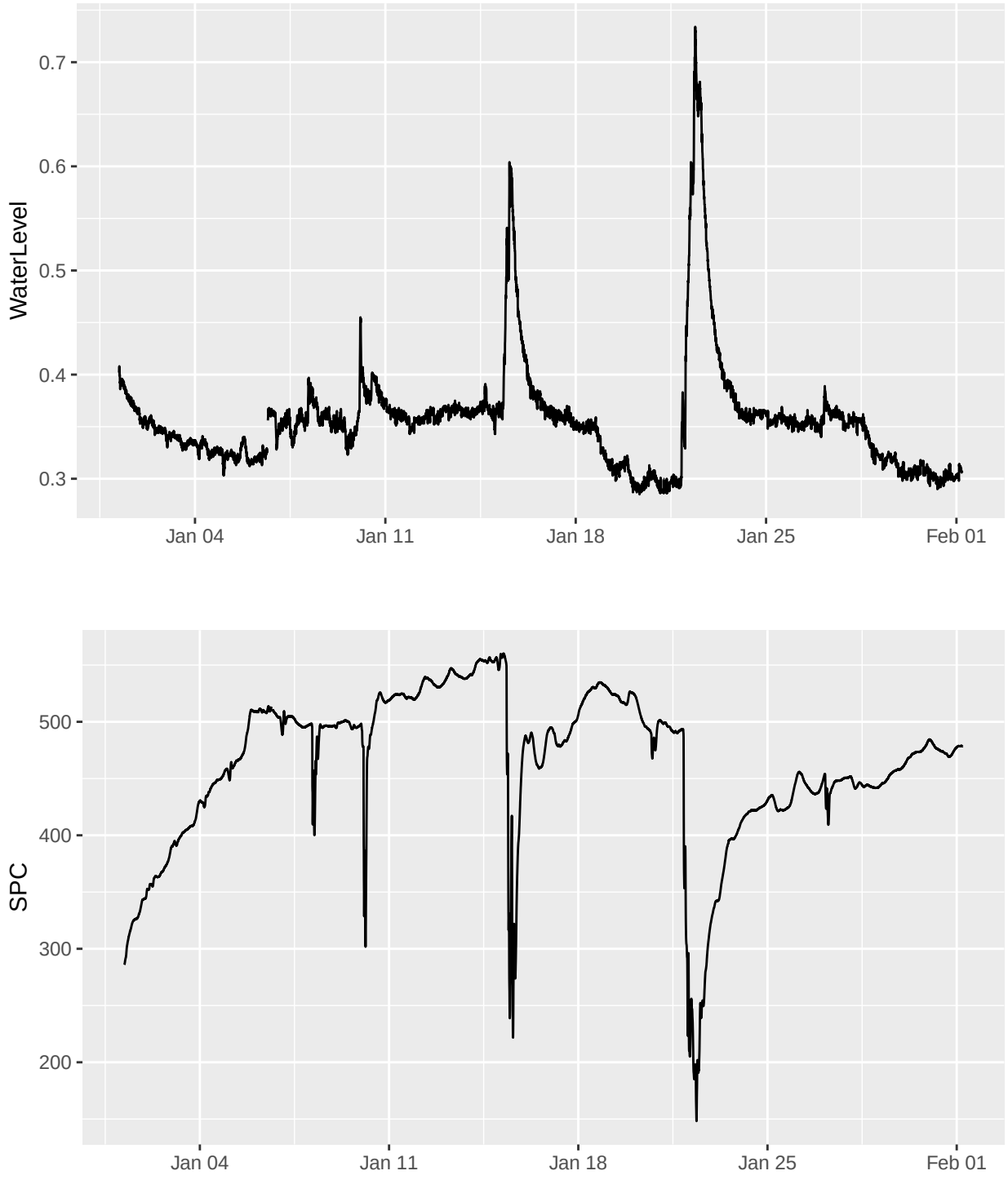

Figure 4: Water level and SPC at Carr from 2016-01-01 to 2016-02-01. Data are not standardized or transformed.

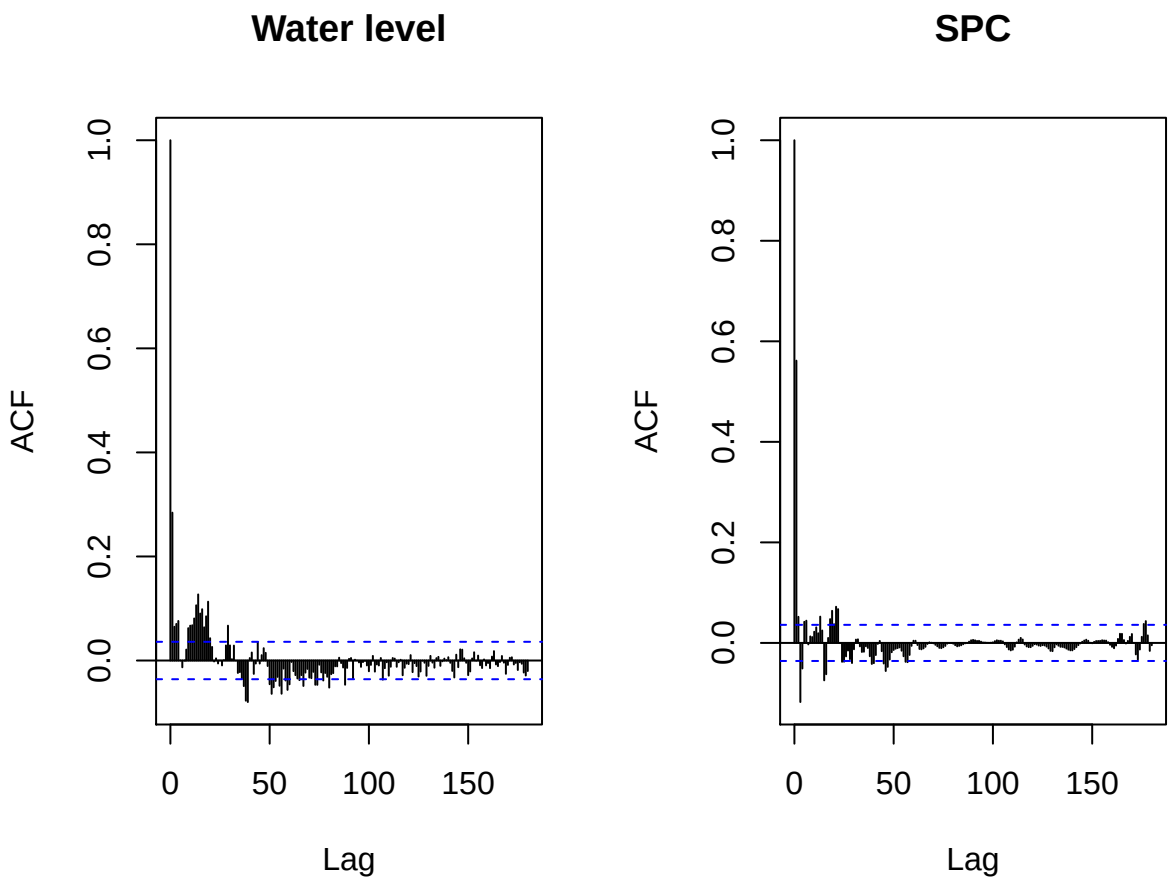

Figure 5: ACF plots of first-differenced water level and SPC at Carr. One lag represents a 15 minute interval.

Data for Carr during 2016-01-01 to 2016-02-01 were grouped by 15 minute chunks and then the averaged for each chunk.

ACF plots of first-differenced water level and SPC at Carr from 2016-01-01 to 2016-02-01 show that water level remains correlated with itself for almost two hours. SPC remains correlated with itself for about an hour.

ACF of SPC: Brooklyn, Carr, Shoal, Turkey

Cross correlation of water level and SPC: all sites

Now we're interested in assessing the correlation between daily stage height signals and daily SPC for the rest of the sites in 15 minute intervals.

Apparently in the year from February 2015 to Feb 2016, only 4 sites had both water level and SPC taken consistently? Possibly not true, but will look at it later.

Next steps

- Look at days with and without precipitation separately.

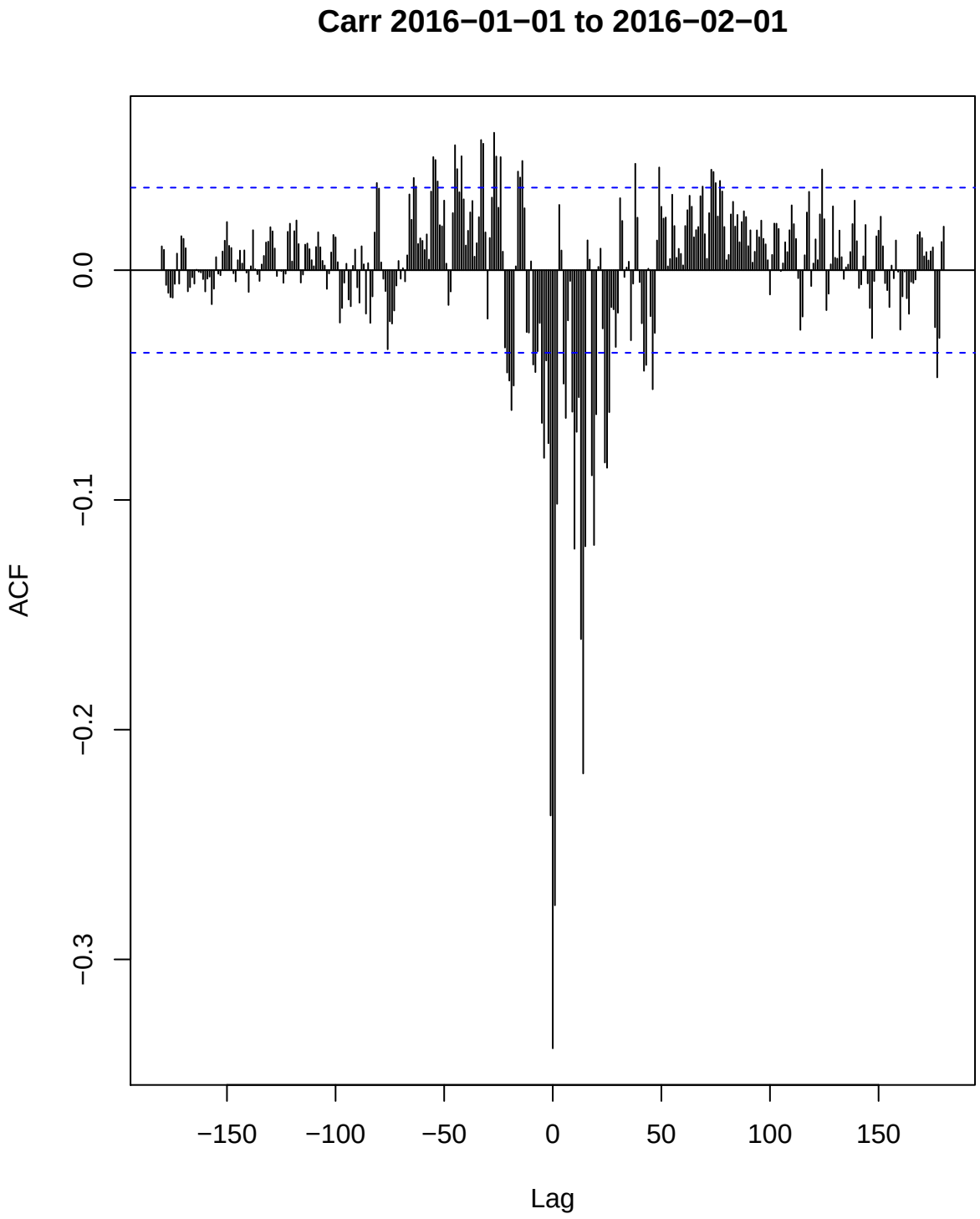

Figure 6: Cross correlation between stage height and SPC. Lags are 15 min intervals. The two series are negatively correlated (-0.3)

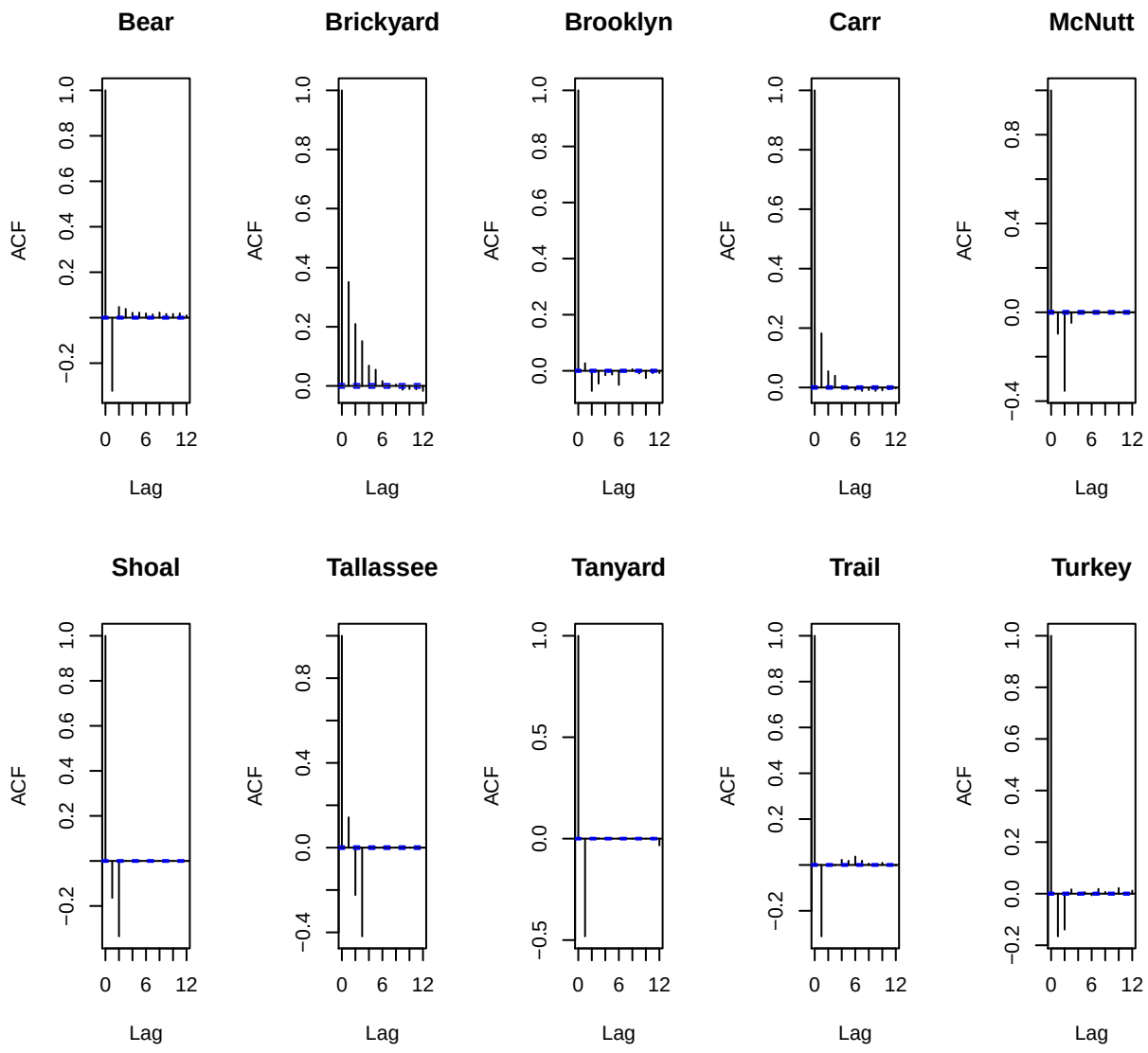

Figure 7: ACF of all sites SPC. All time series are in 5 minutes intervals (so one lag is 5 minutes).

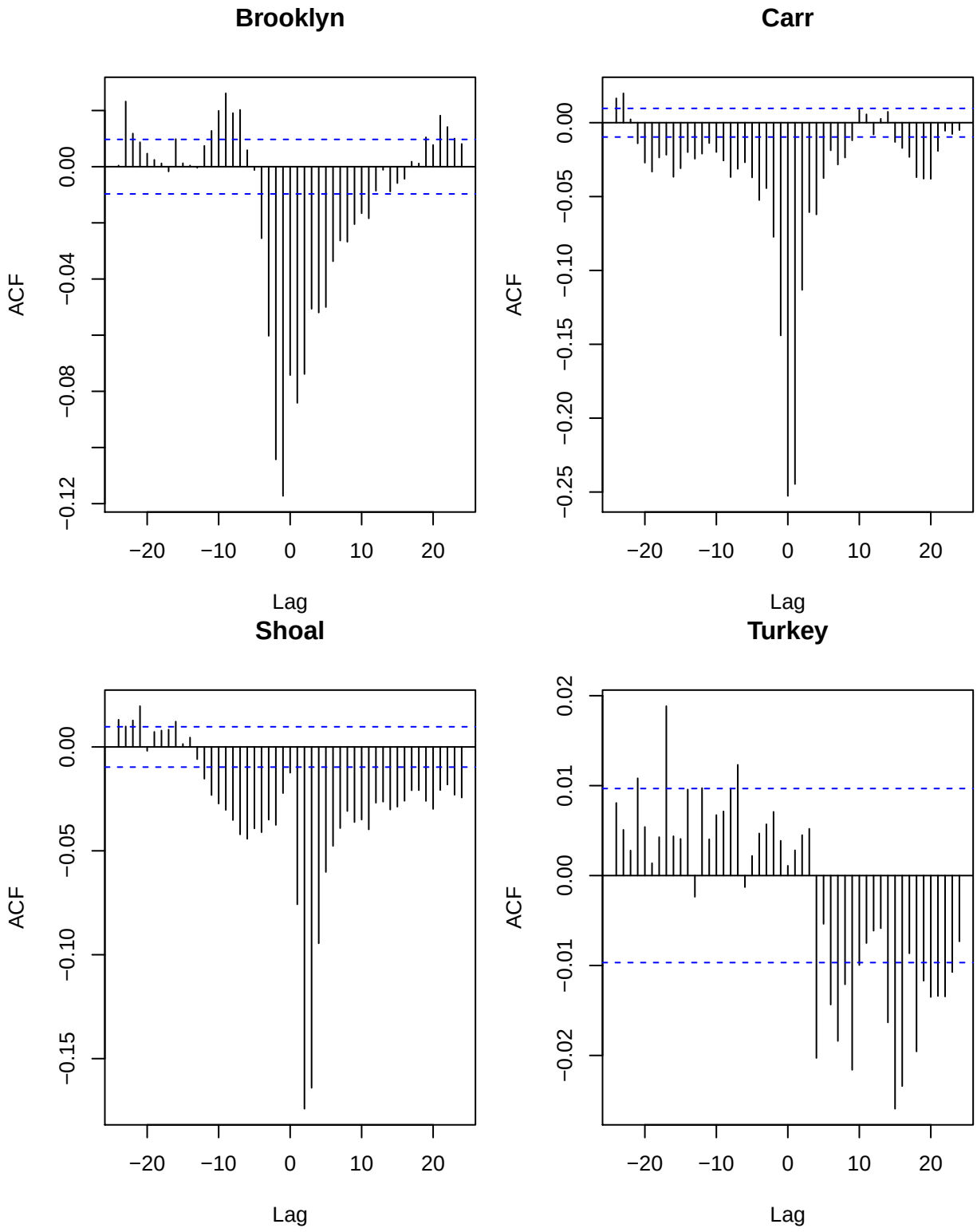

Figure 8: CCF plots of four sites with comparable time series. Lags are 5 minute intervals.