What are the usable results from the ACF and PACF functions?

The autocorrelation and partial autocorrelation functions helped determine model parameter combinations to compare (see 3.2 Time series analyses: Seasonal ARIMA models - Model Choice for a greater discussion of methods). Each set of parameter combinations was modeled for intra-station comparison.

What are the usable results from the STL analysis (including model choice)?

The best model for each station was determined by which had the lowest Akaike’s Information Criterion (AIC) value (see table 2). (Note that AIC values cannot be used to compare between models computed from different datasets.) Station COH726 is best modeled by p, d, q parameters of 3, 0, and 3 respectively (AIC= 1387.880). Station LYN614 is best modeled by p, d, q parameters of 3, 0, and 3 respectively (AIC= 1557.391). Station WEY38.4 is best modeled by p ,d, q parameters of 7, 0, and 7 respectively (AIC= 1253.584).

What are the usable results from the best models for each station?

What do these results tell us about the models and the systems?

The forecasted mean for all three models mostly overlays the real data, although there is significantly greater variation in the real data than in data forecasted for the same time period. There are several months of real data for each station that exceed the upper 95% confidence limit for forecasted data. In contrast, there are almost no months of real data that are less than the lower 95% confidence limit. (Note the lower 95% confidence limit for all three models drops below 0” of precipitation, indicating not rainfall is predicted.) This could indicate that the models are not handling the extent of the variation in the data well and may be biased towards better predicting low and average rainfall months. Alternatively, this could indicate that anomalously large rainfall events are driven by non-seasonal or indirectly related factors.

The model forecasts for future dates for all three stations show such little variation compared to the real data that is concerning. Plotting actual versus predicted monthly precipitation for each station (independent of date) shows that the models exhibit a strong pull towards the mean, where the spread of predicted values is narrowly concentrated around a mean compared to the spread of actual values.

It’s likely that month is not a strong indicator for rainfall amount because the climate is highly variable year-to-year and seaside towns are prone to high-intensity rainfall events during storms like Nor’easters or large snowfalls. I anticipate that adding other data to the models (e.g., temperature, wind direction, or humidity) would greatly increase a model’s predictive power in modeling precipitation.