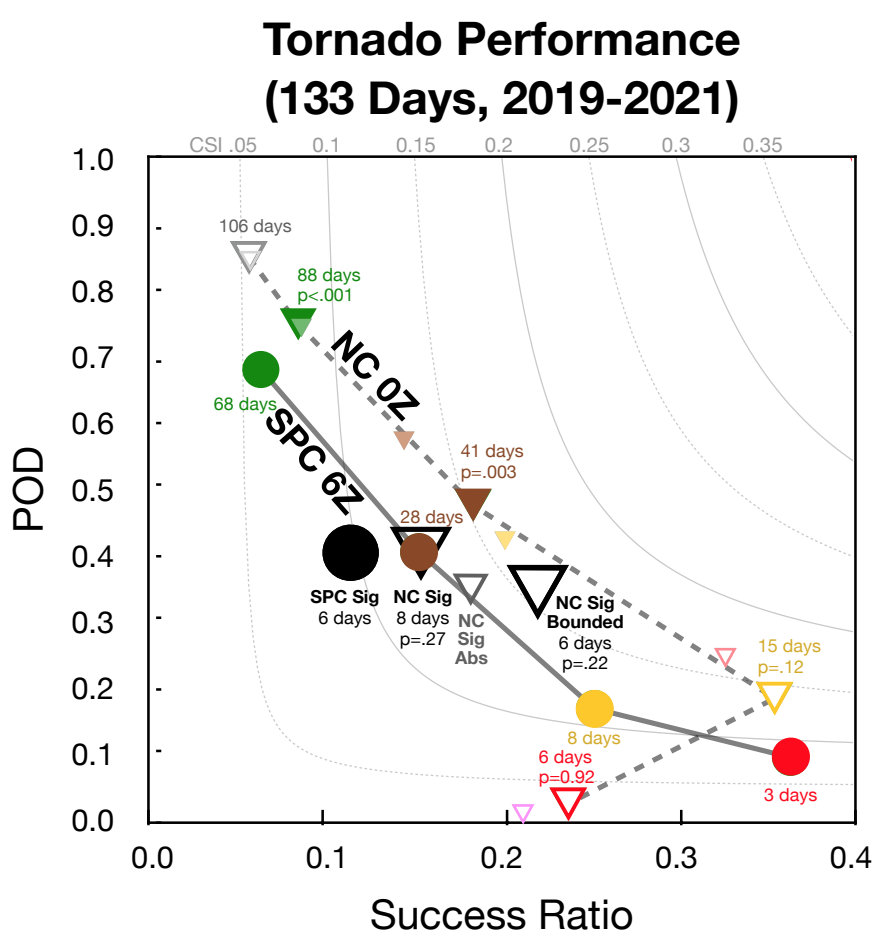


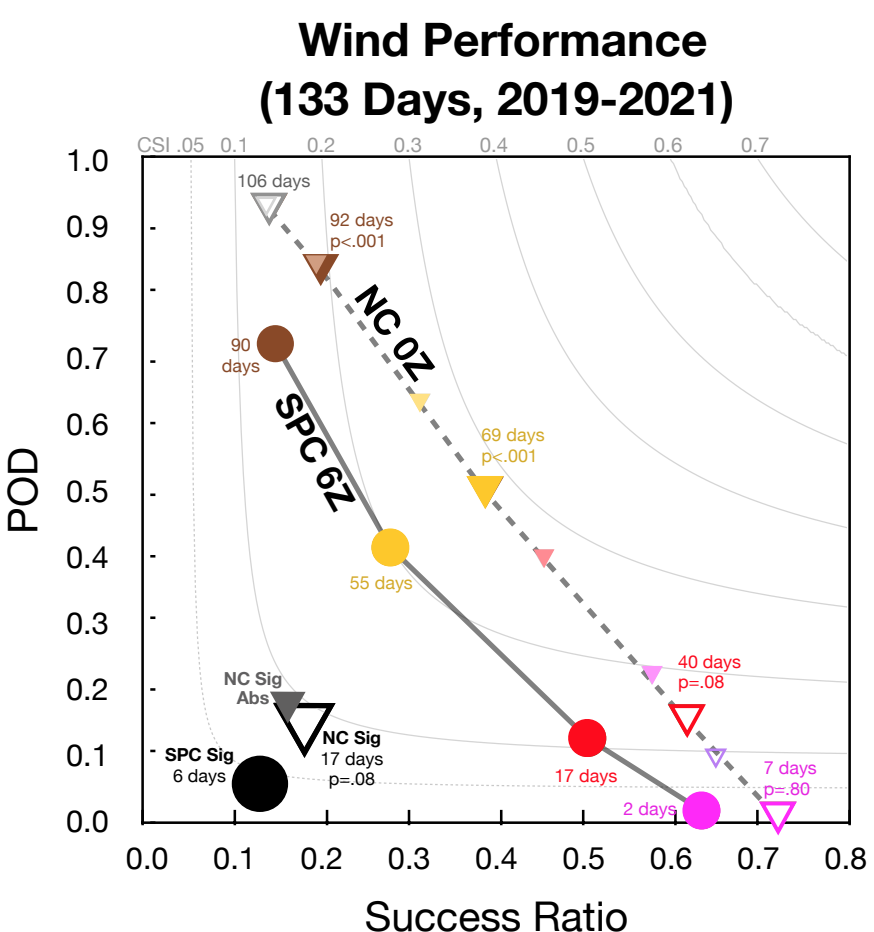
# 0Z Nadocast / 6Z SPC

## Day 1 Outlooks

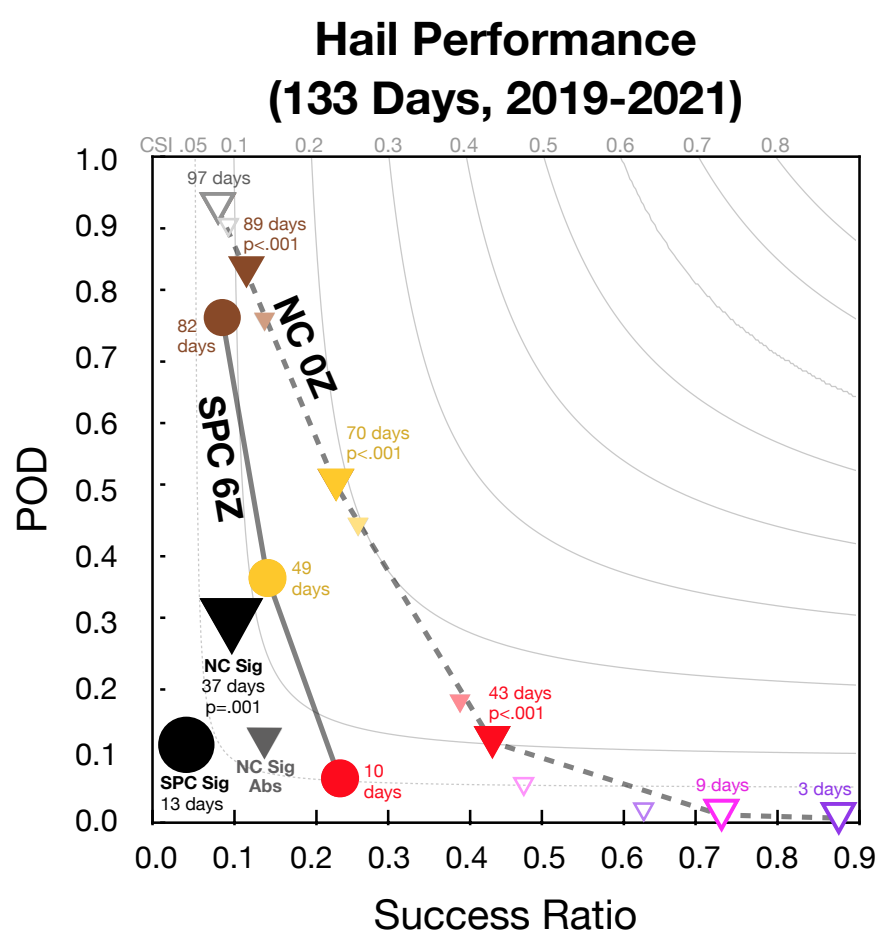
### Perf diagrams



Performance diagram for 0Z Nadocast (NC) 24hr tornado outlook (triangles) compared to 6Z SPC day 1 outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Fainter, smaller triangles are NC with absolute probability calibration instead of SPC-like calibration. Lines of constant critical success index (CSI) are drawn as light gray curves. The performance for the NC 1% threshold is included as a gray triangle. Significant tornado (EF2+) 10% thresholds are drawn in black; two symbols are shown for NC SPC-like sig predictions: the naive sig model, and the sig model bounded to never output probabilities higher than the non-sig probability. Absolutely calibrated NC naive and bounded sig predictions are similar—only one symbol is shown. Labels indicate with the number of days the threshold appears. To assess statistical significance, for each threshold the difference between the NC and SPC CSI is compared across 1 million bootstrap resamples of the 133 days. The p value for NC CSI exceeding SPC CSI is shown; the NC triangle is drawn with solid fill when  $p < .05$  (i.e. when NC is statistically more skillful).

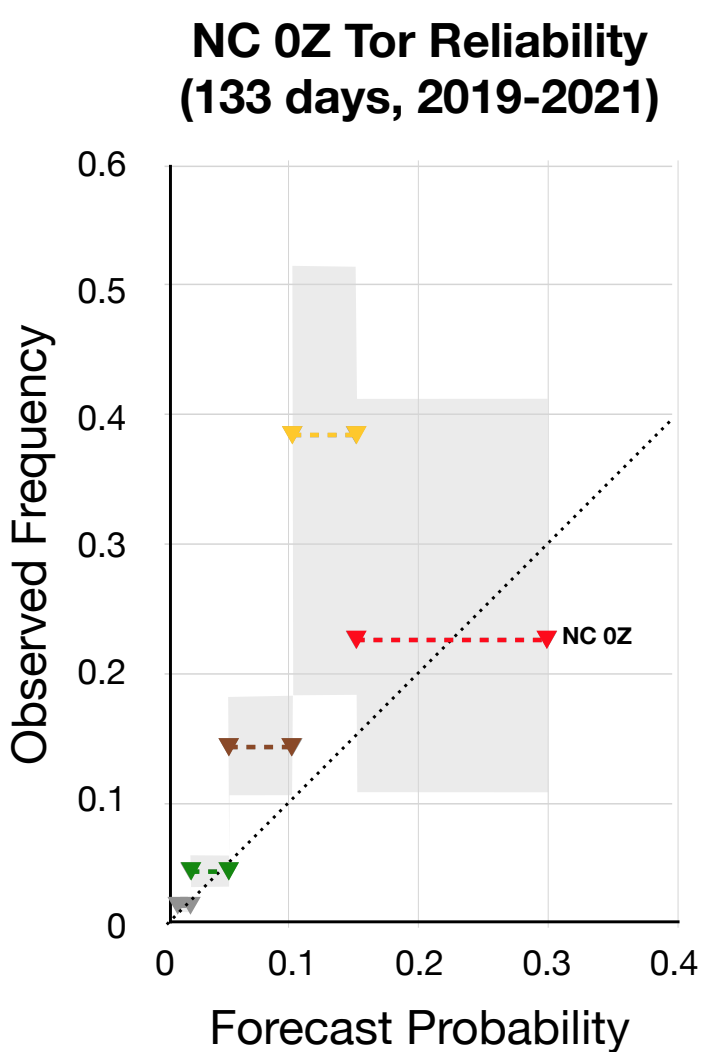


As before, but for severe wind. The gray triangle is the NC 2% threshold. The NC naive and bounded sig models produce the same results on this dataset.

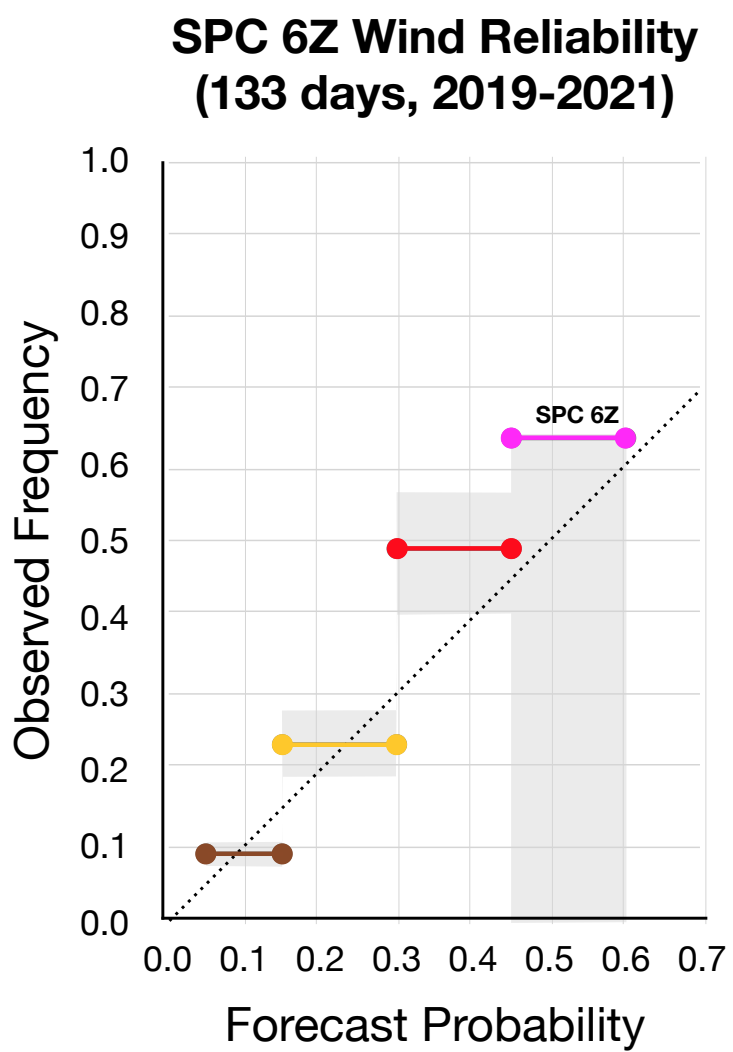
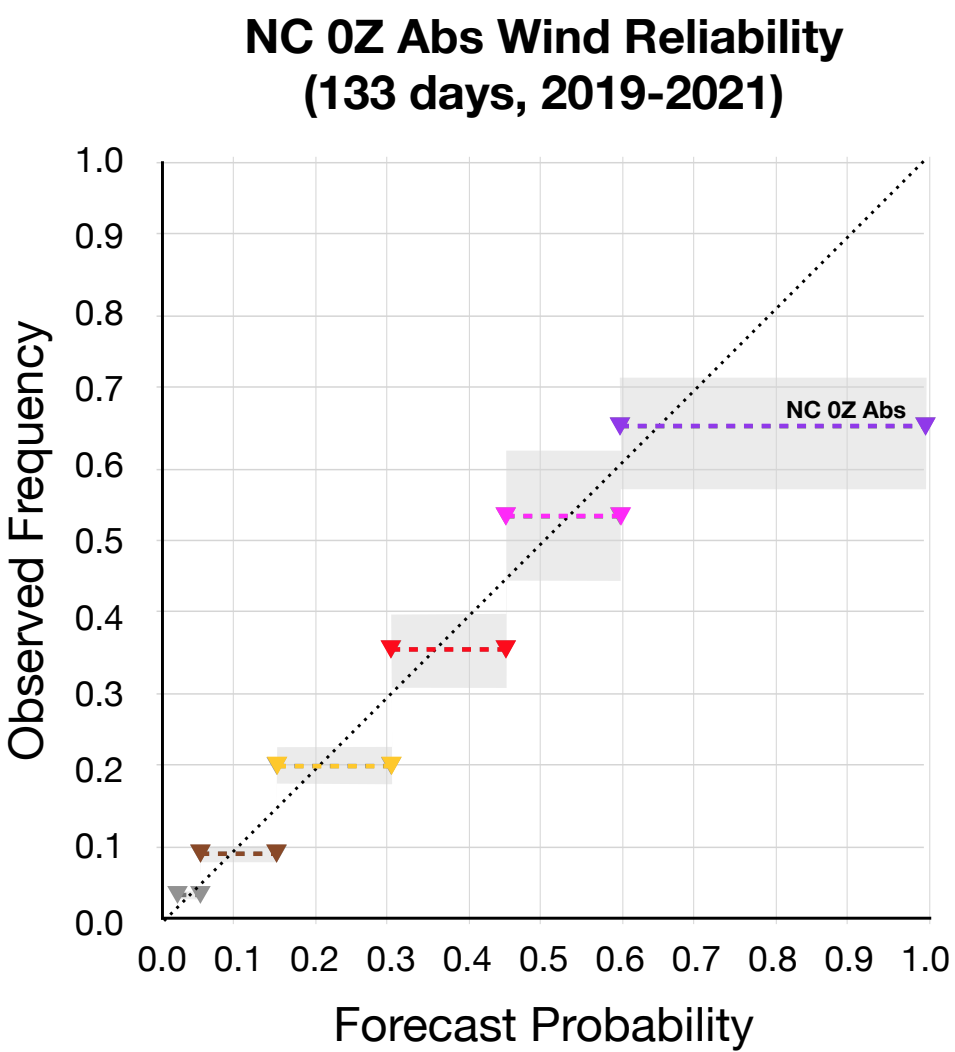
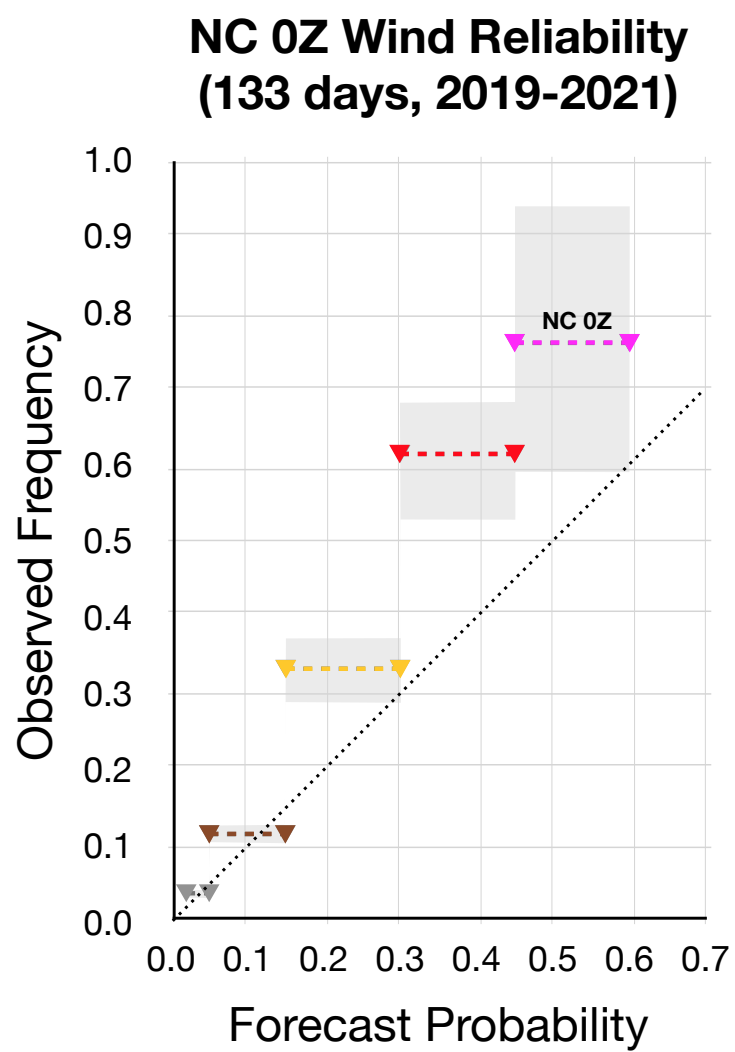
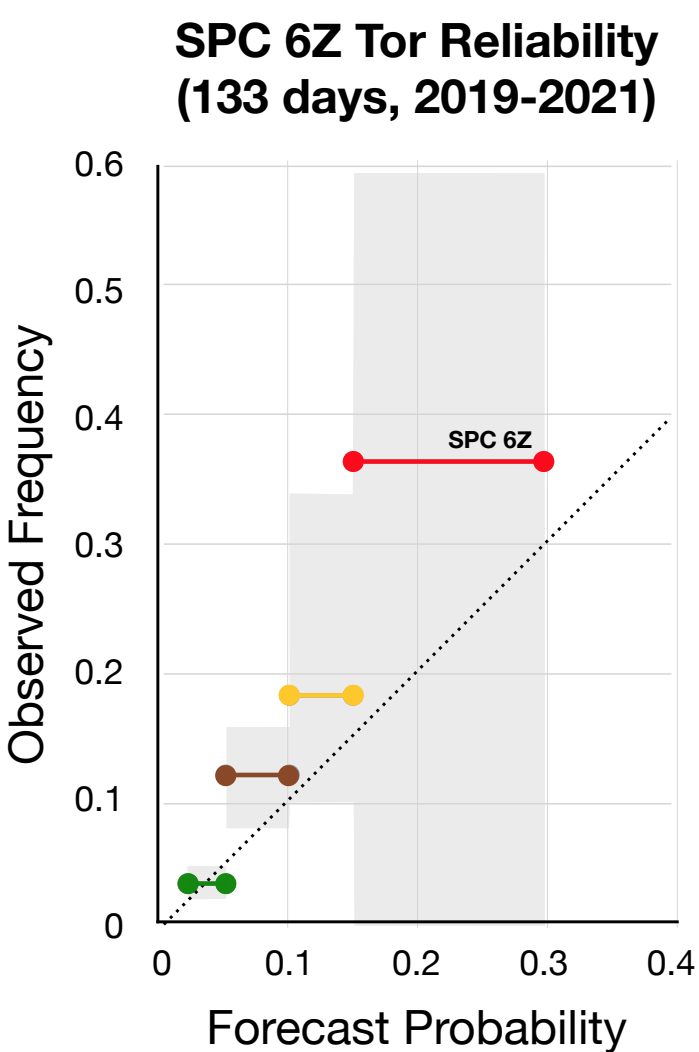
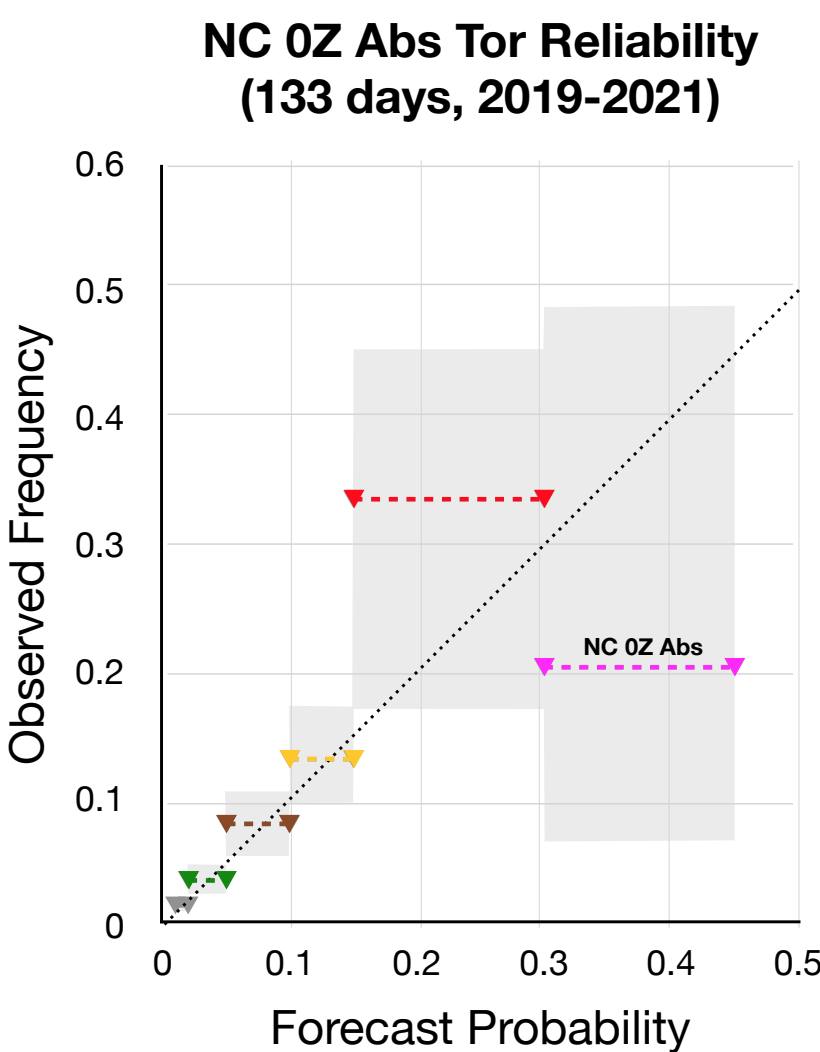


As before, but for severe hail. The gray triangle is the NC 2% threshold. The SPC 6Z outlooks did not include any 45% or 60% hail days in this dataset. The NC naive and bounded sig models produce the same results on this dataset.

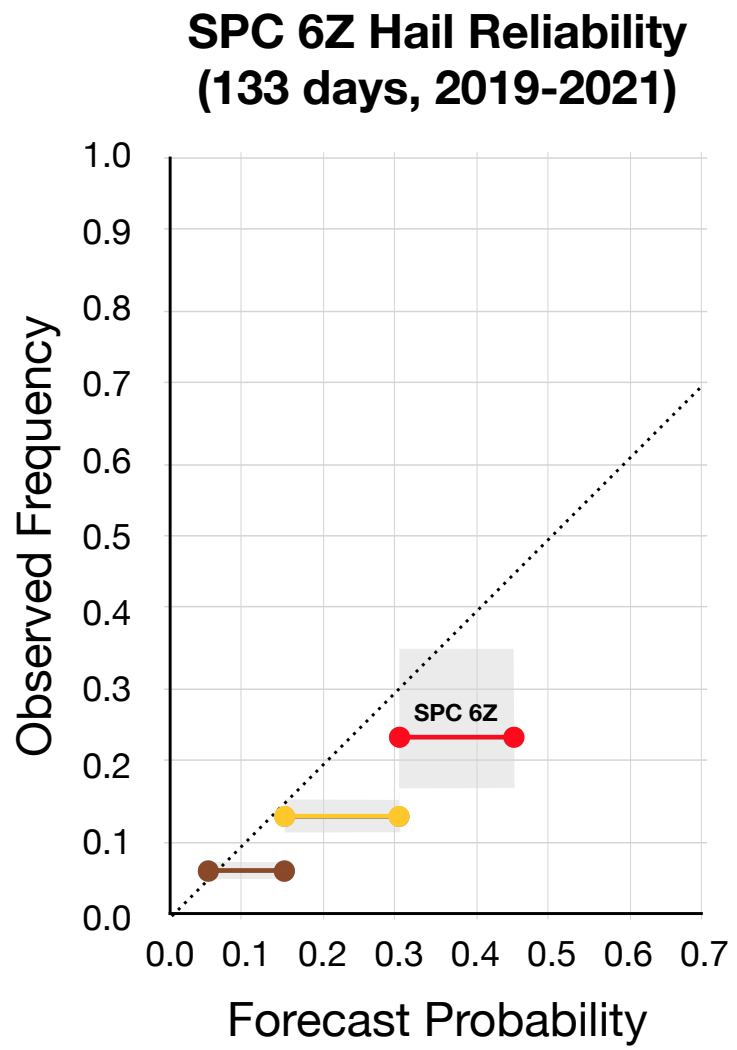
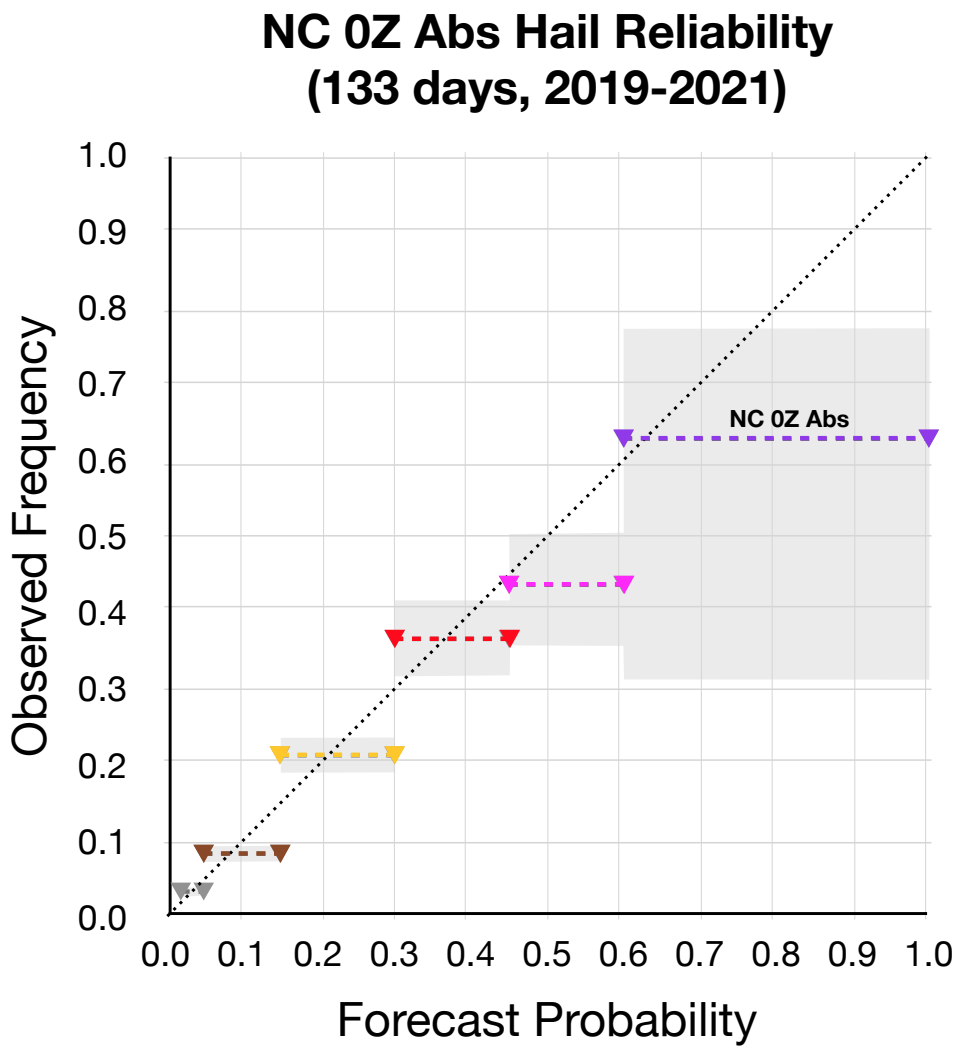
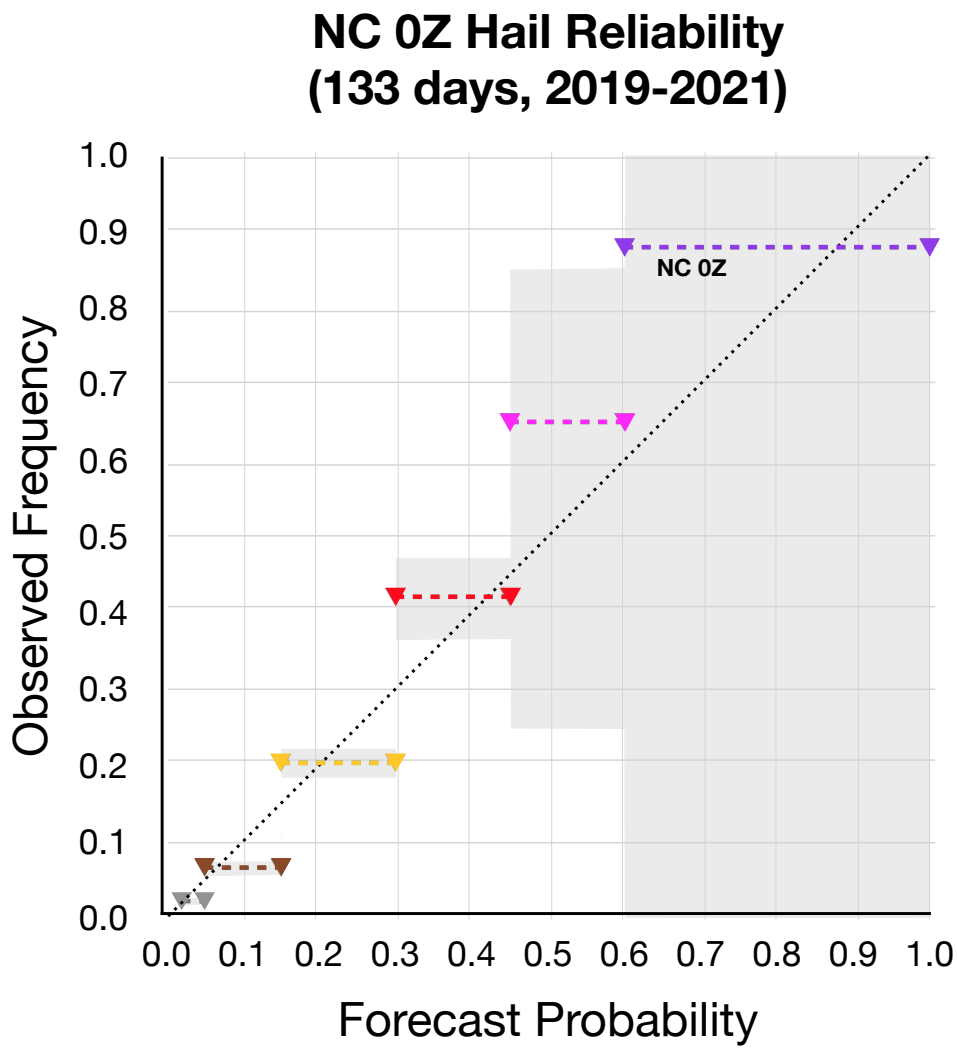
### Reliability diagrams



Reliability diagram for 0Z Nadocast (NC) 24hr tornado outlook with SPC-like calibration (left, triangles) and absolute calibration (center, triangles), compared to 6Z SPC day 1 outlooks (right, circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. The plots compare the forecast probability of an event to the observed frequency of at least one report within 25 miles. The bins for 2-5%, 5-10%, 10-15%, 15-30%, and 30-45% forecast probabilities are each shown as horizontal lines. A bin for NC 1%-2% is included as well (gray triangles). The 95% confidence interval for each bin is shaded, computed by bootstrapping over the 133 days. Ideal reliability is drawn as a dotted diagonal line.

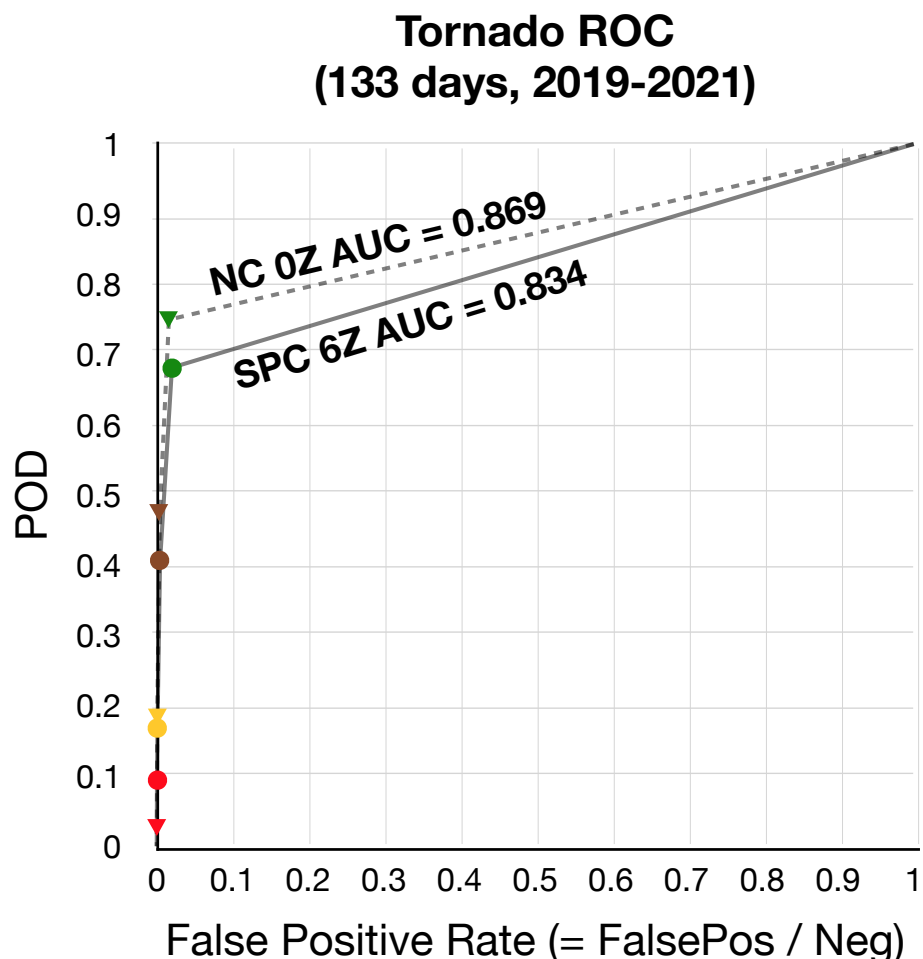


As before, but for severe wind. Bins are 2%-5% (NC only), 5-15%, 15-30%, and 30-45%, 45-60%, and 60%-100%.

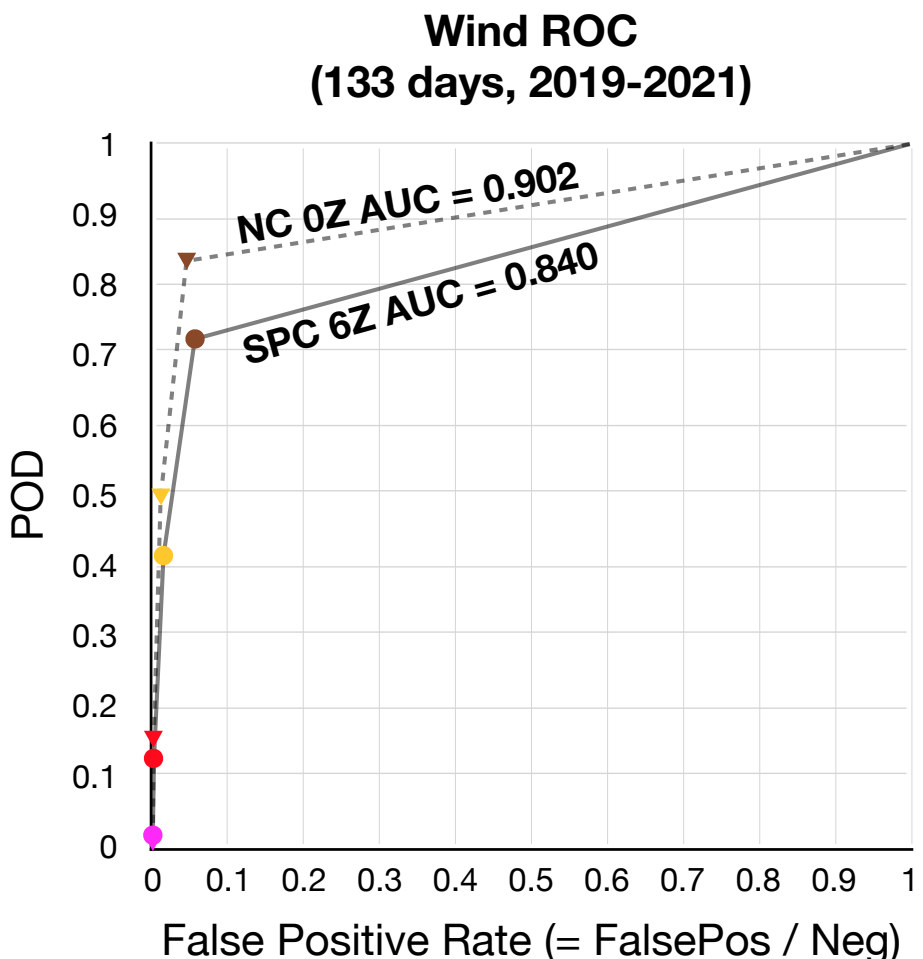


As before, but for severe hail. Bins are 2%-5% (NC only), 5-15%, 15-30%, and 30-45%, and 45-60%, and 60%-100%. The SPC 6Z outlooks did not include any 45% or 60% hail days in this dataset.

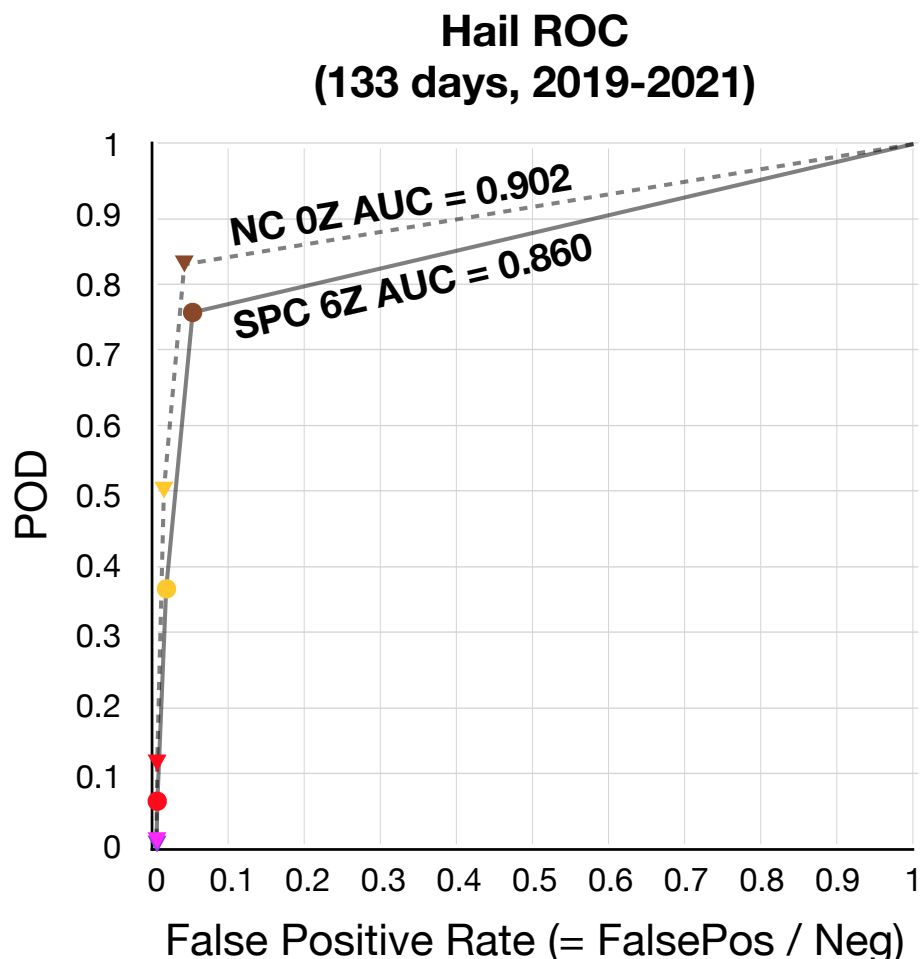
### ROC curves



Receiver operator characteristic curve (ROC) for 0Z Nadocast (NC) 24hr tornado outlook (triangles) and 6Z SPC day 1 outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Area under the curve (AUC) is shown, a metric which is dominated by POD at the lowest threshold.



As before, but for severe wind.



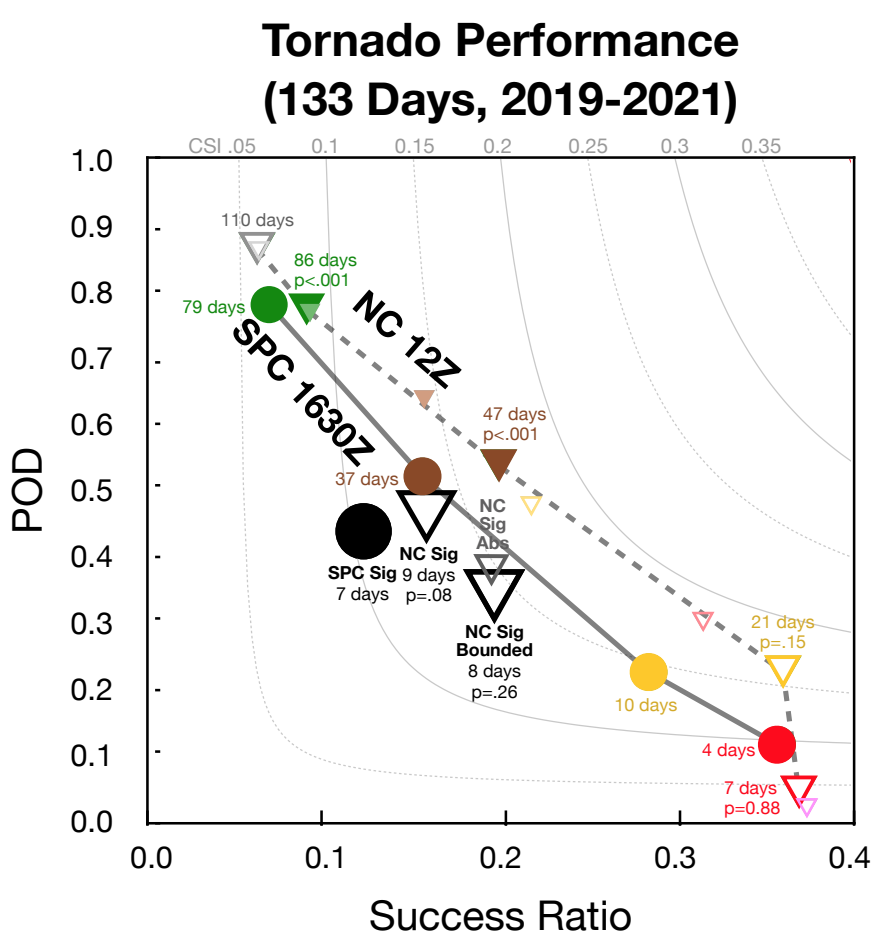
As before, but for severe hail.



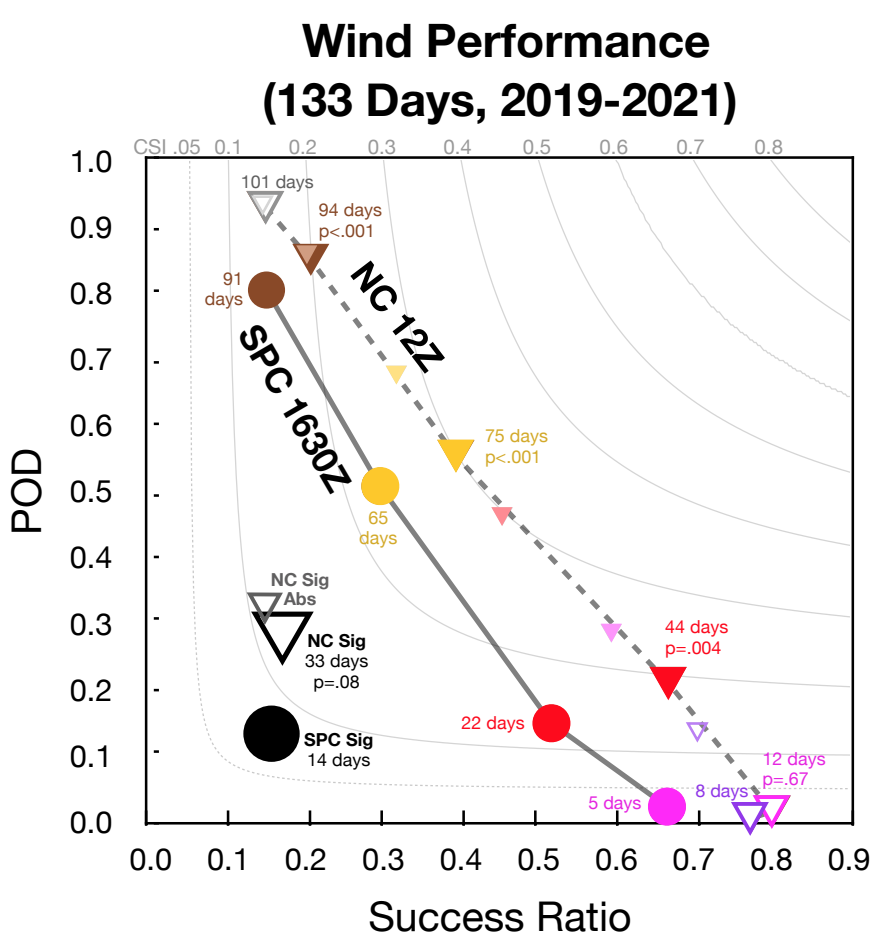
# 12Z Nadocast / 1630Z SPC

## Day 1 Outlooks

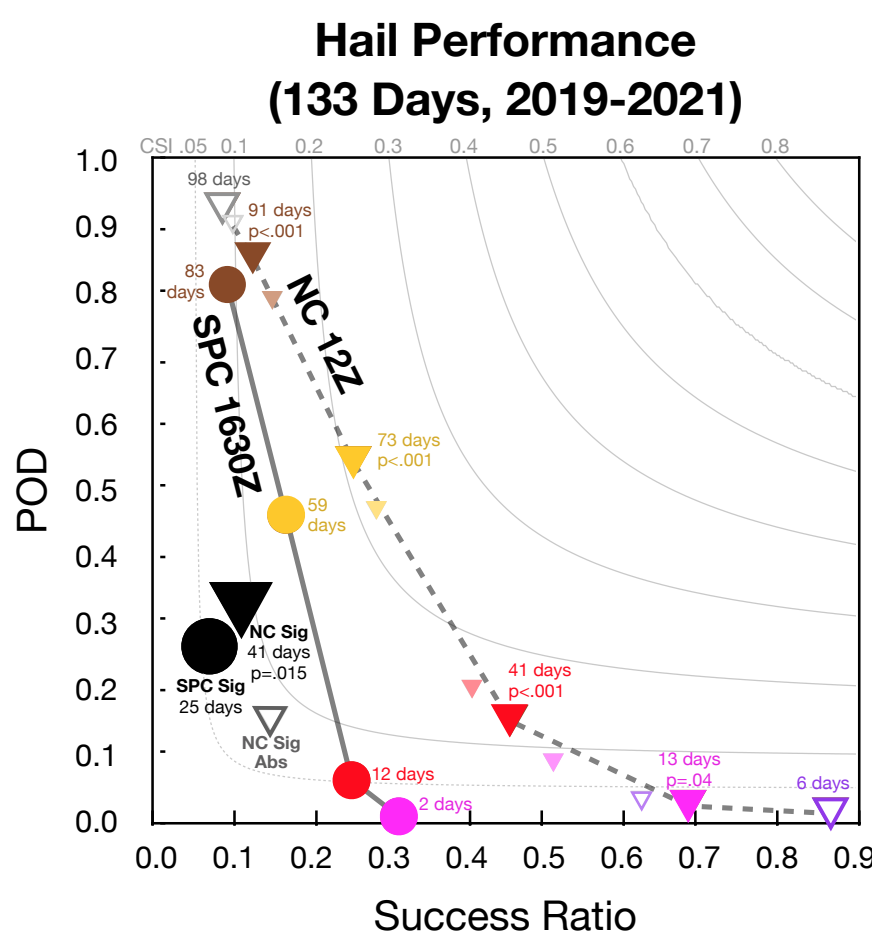
### Perf diagrams



Performance diagram for 12Z Nadocast (NC) day 1 tornado outlook (triangles) compared to 1630Z SPC day 1 outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Fainter, smaller triangles are NC with absolute probability calibration instead of SPC-like calibration. Lines of constant critical success index (CSI) are drawn as light gray curves. The performance for the NC 1% threshold is included as a gray triangle. Significant tornado (EF2+) 10% thresholds are drawn in black; two symbols are shown for NC SPC-like sig predictions: the naive sig model, and the sig model bounded to never output probabilities higher than the non-sig probability. Absolutely calibrated NC naive and bounded sig predictions are similar—only one symbol is shown. Labels indicate with the number of days the threshold appears. To assess statistical significance, for each threshold the difference between the NC and SPC CSI is compared across 1 million bootstrap resamples of the 133 days. The p value for NC CSI exceeding SPC CSI is shown; the NC triangle is drawn with solid fill when  $p < .05$  (i.e. when NC is statistically more skillful).

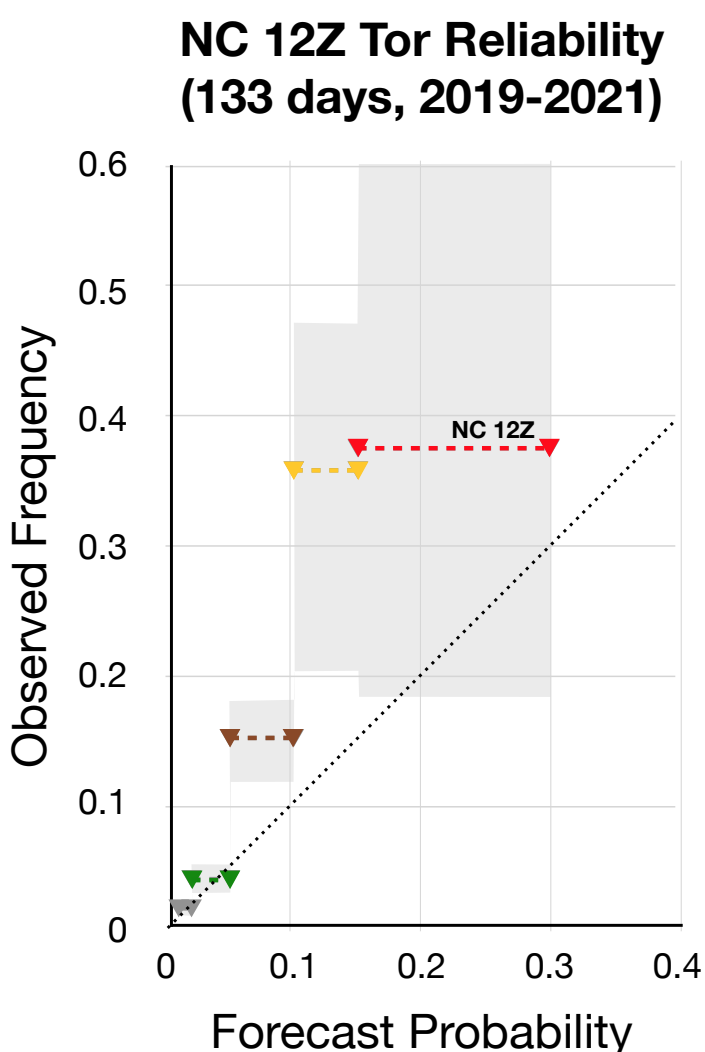


As before, but for severe wind. The gray triangle is the NC 2% threshold. The NC naive and bounded sig models produce similar results on this dataset.

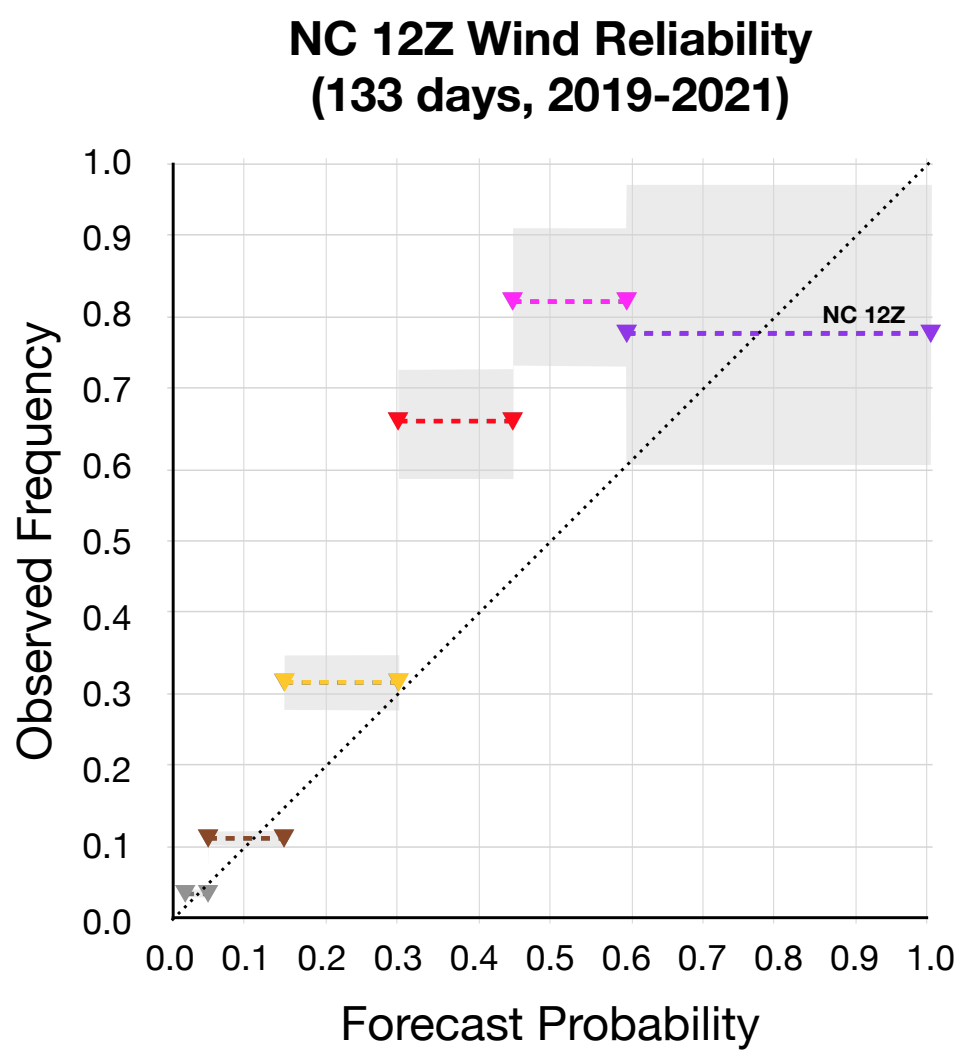
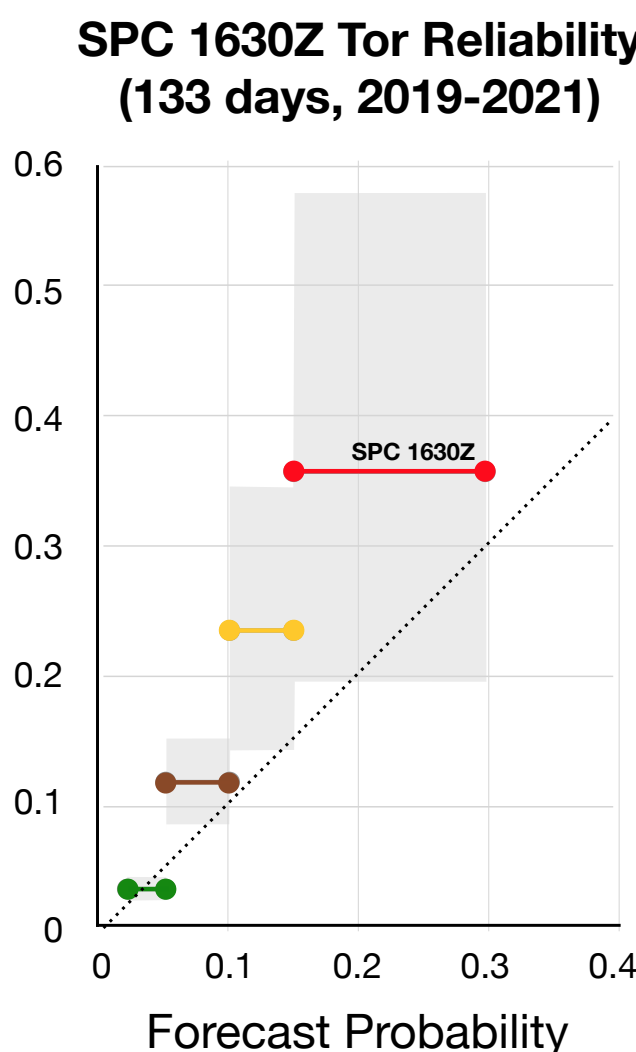
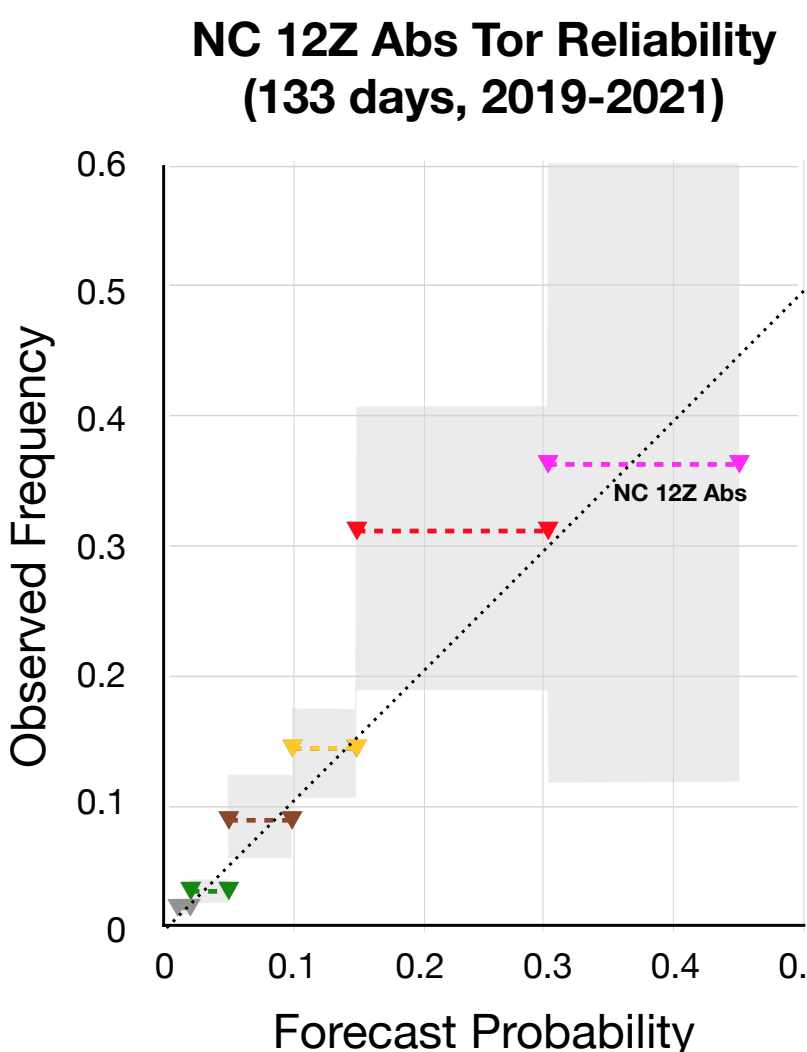


As before, but for severe hail. The gray triangle is the NC 2% threshold. The NC naive and bounded sig models produce the same results on this dataset.

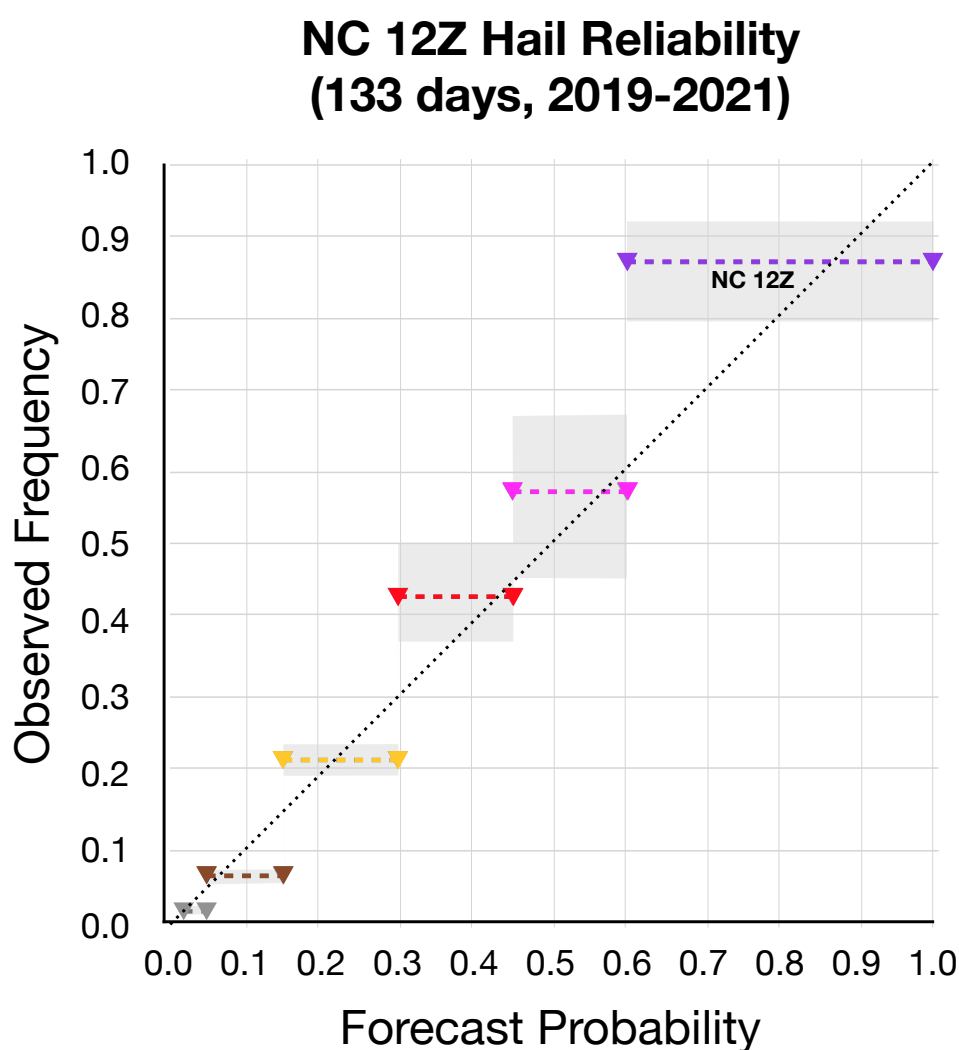
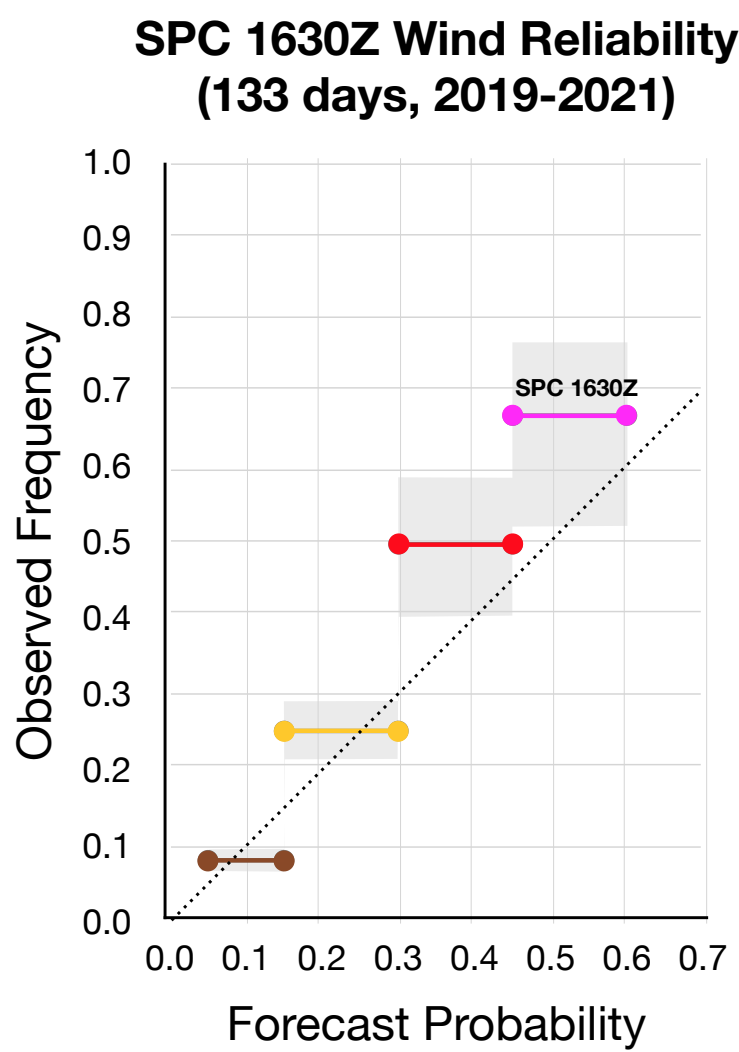
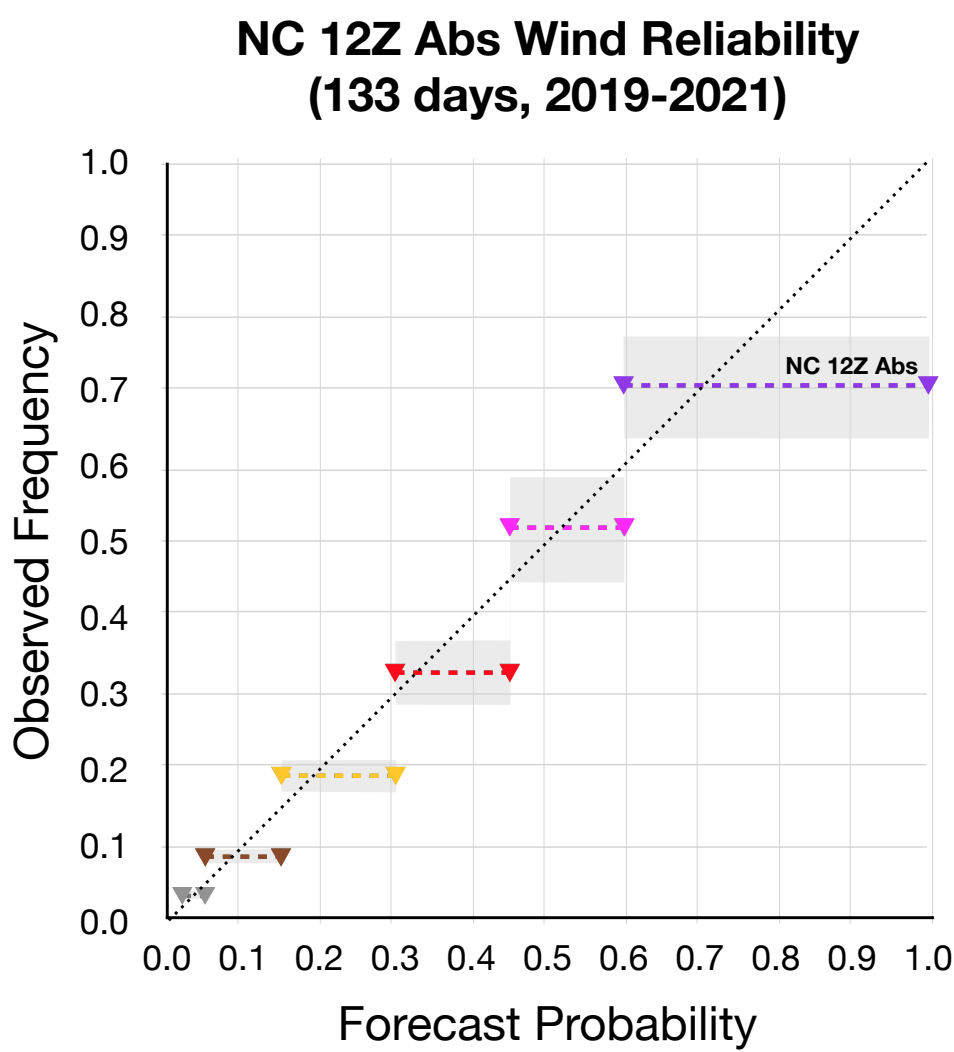
### Reliability diagrams



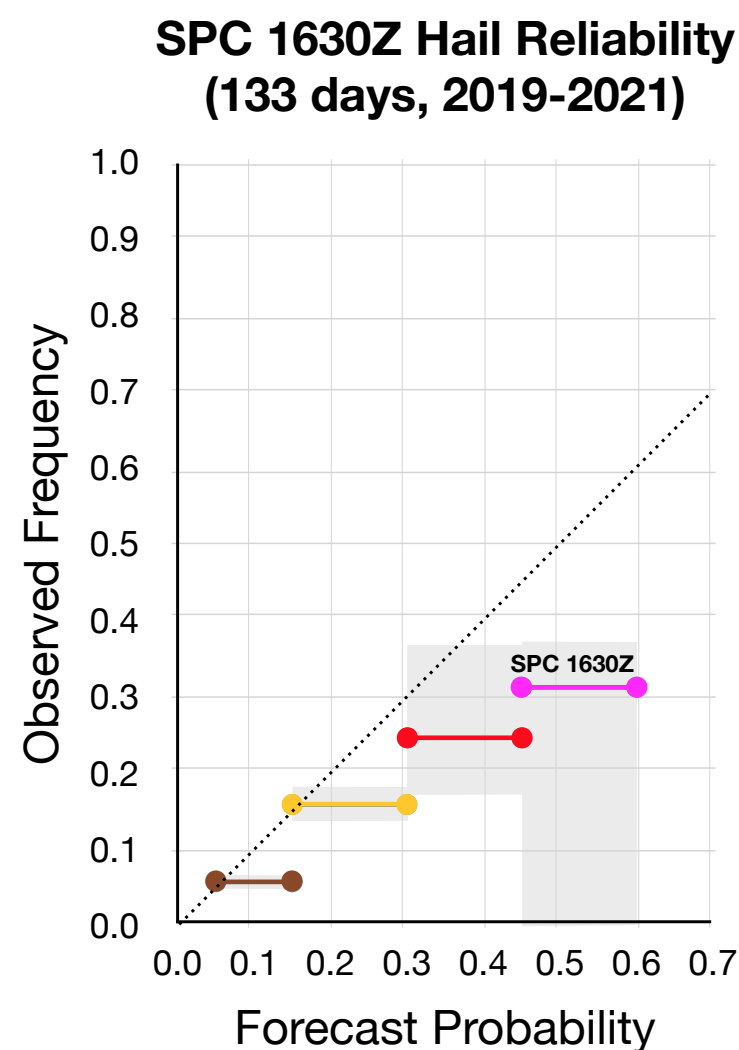
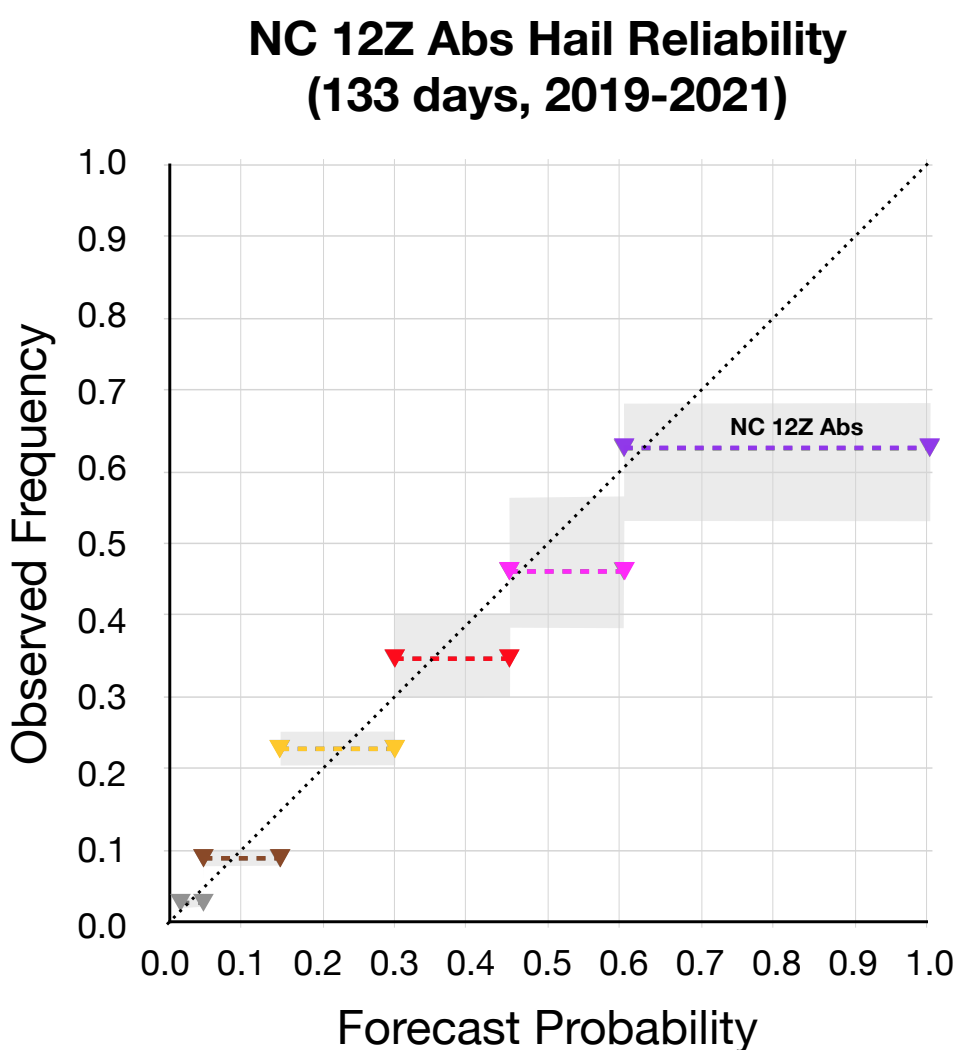
Reliability diagram for 12Z Nadocast (NC) day 1 tornado outlook with SPC-like calibration (left, triangles) and absolute calibration (center, triangles), compared to 1630Z SPC day 1 outlooks (right, circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. The plots compare the forecast probability of an event to the observed frequency of at least one report within 25 miles. The bins for 2-5%, 5-10%, 10-15%, 15-30%, and 30-45% forecast probabilities are each shown as horizontal lines. A bin for NC 1%-2% is included as well (gray triangles). The 95% confidence interval for each bin is shaded, computed by bootstrapping over the 133 days. Ideal reliability is drawn as a dotted diagonal line.



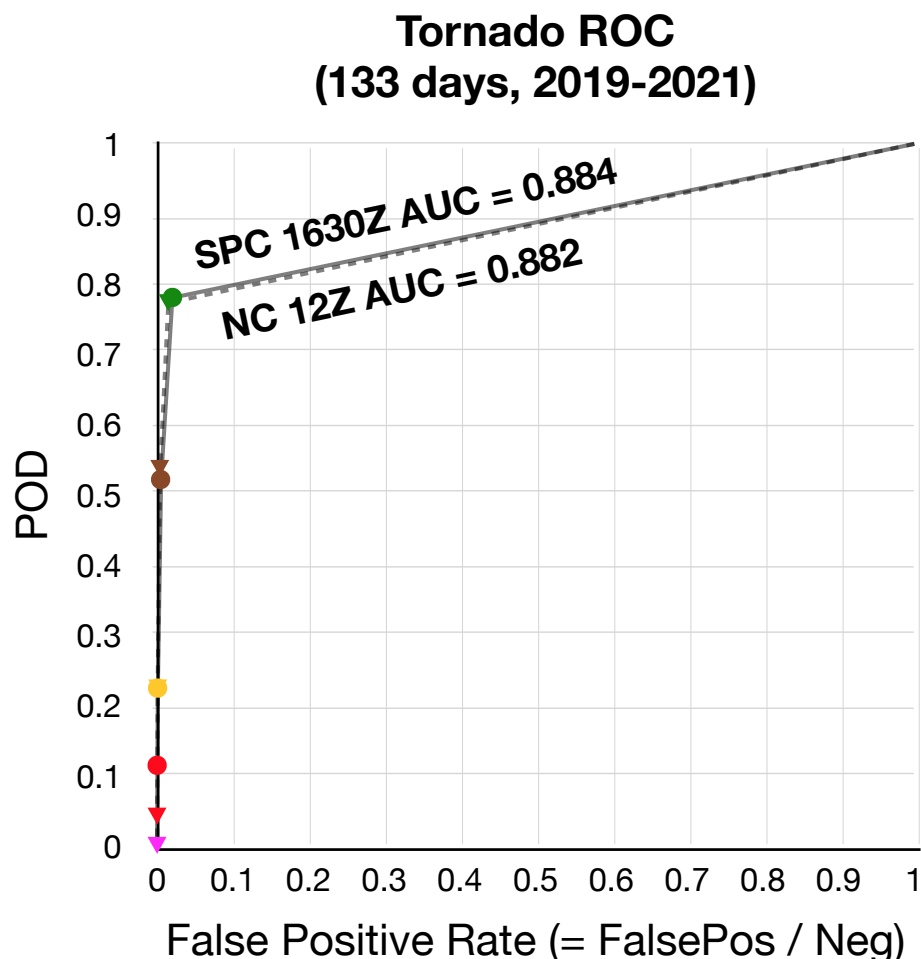
As before, but for severe wind. Bins are 2%-5% (NC only), 5-15%, 15-30%, and 30-45%, 45-60%, and 60%-100%.



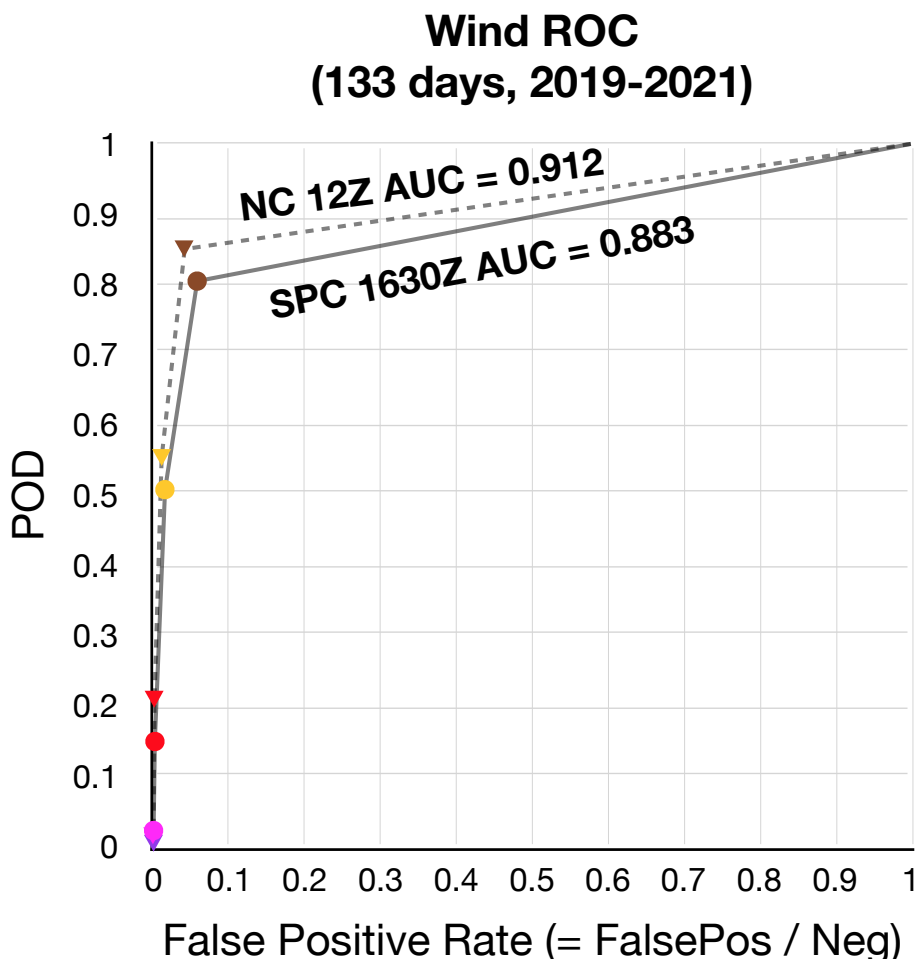
As before, but for severe hail. Bins are 2%-5% (NC only), 5-15%, 15-30%, and 30-45%, and 45-60%, and 60%-100%.



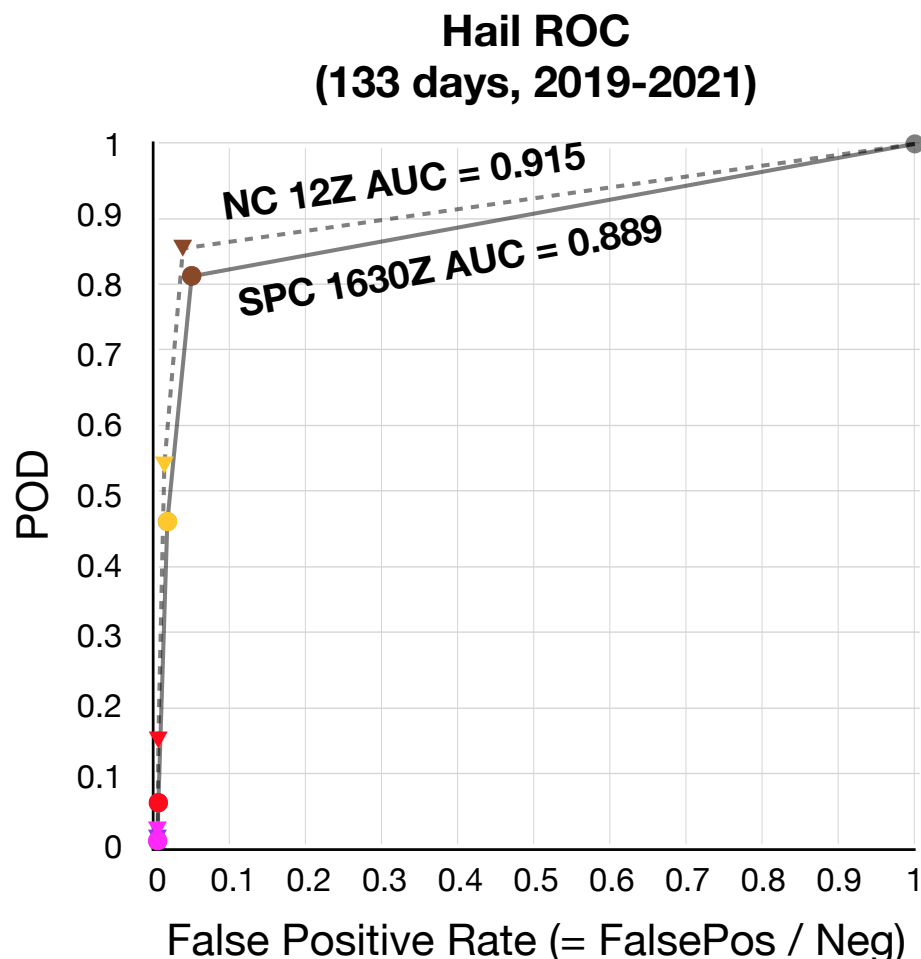
### ROC curves



Receiver operator characteristic curve (ROC) for 12Z Nadocast (NC) day 1 tornado outlook (triangles) and 1630Z SPC day 1 outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Area under the curve (AUC) is shown, a metric which is dominated by POD at the lowest threshold.



As before, but for severe wind.

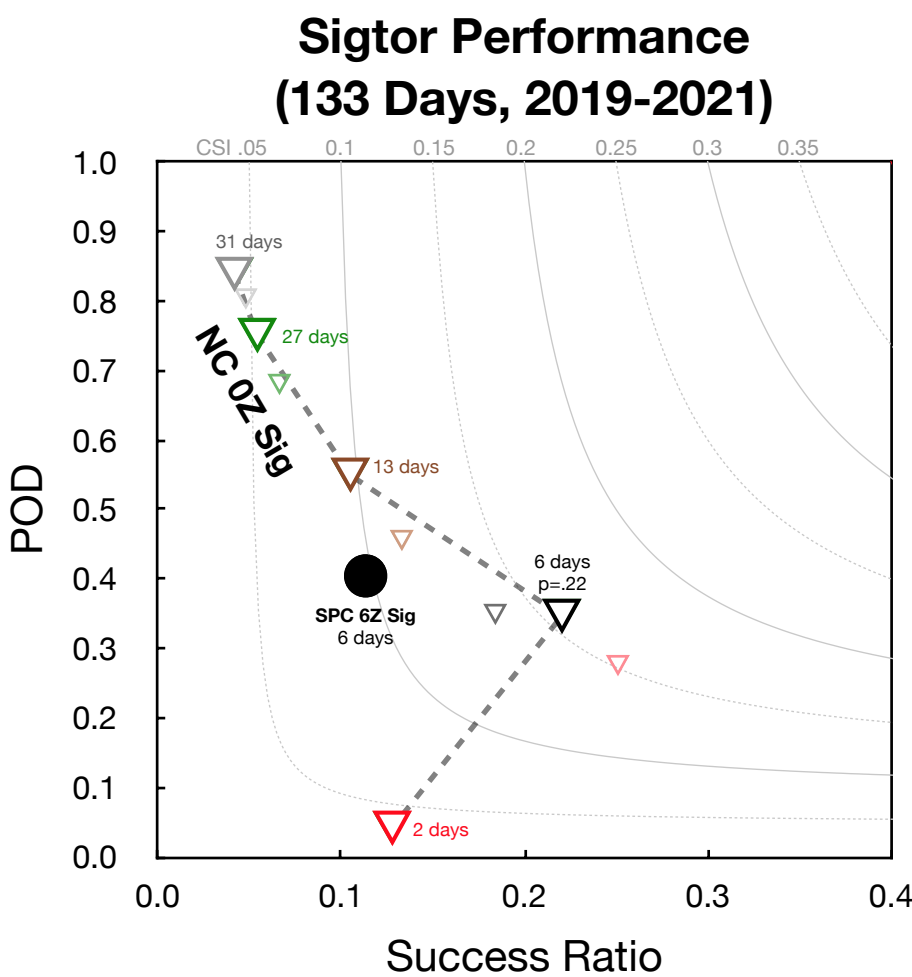


As before, but for severe hail.

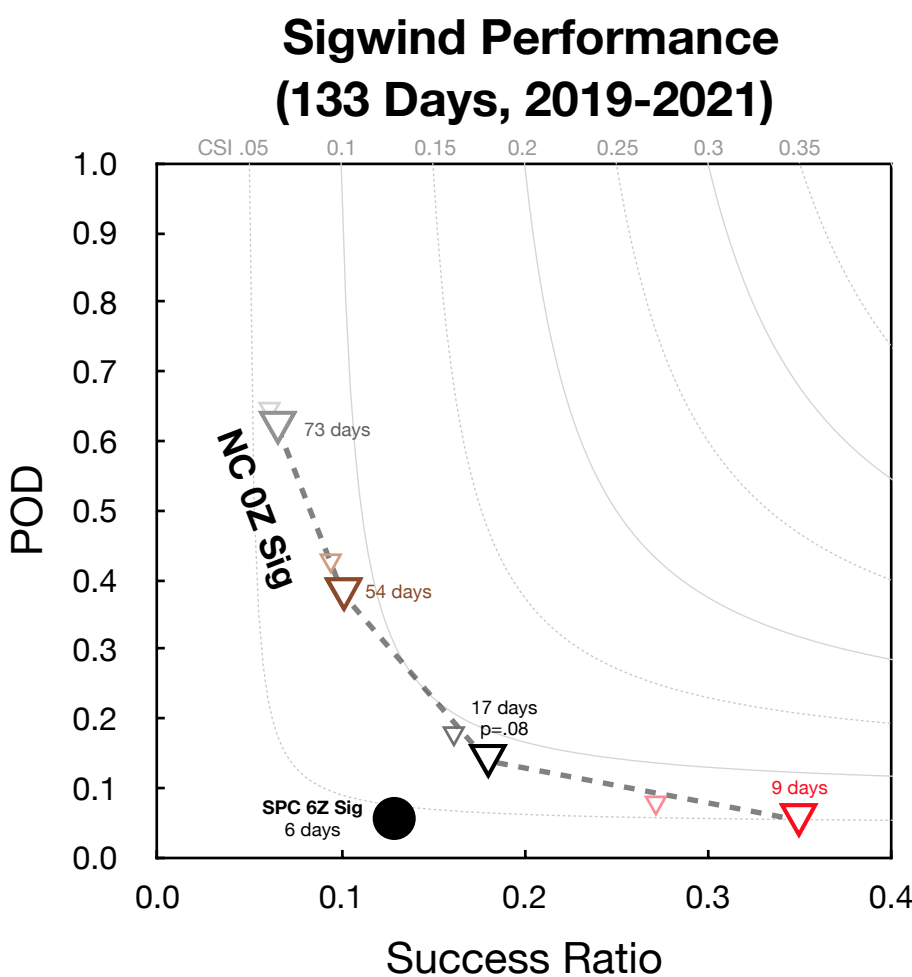
# 0Z Nadocast / 6Z SPC

## Day 1 Sigsevere Outlooks

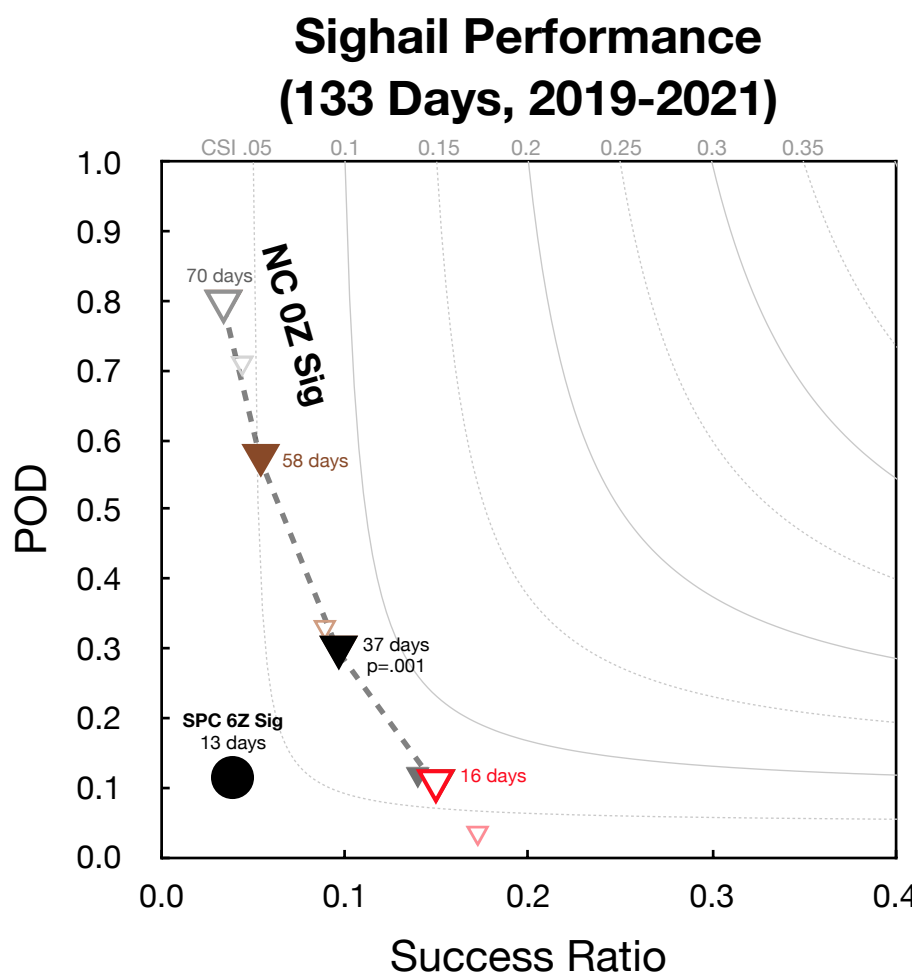
### Perf diagrams



Performance diagram for 0Z Nadocast (NC) 24hr significant tornado outlook (triangles) compared to 6Z SPC day 1 sigtor outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Fainter, smaller triangles are NC with absolute probability calibration instead of SPC-like calibration. Lines of constant critical success index (CSI) are drawn as light gray curves. The performance for the NC 1% threshold is included as a gray triangle. The NC sigtor model is bounded to never output probabilities higher than the NC non-sig prediction. To assess statistical significance, for the 10% threshold the difference between the NC and SPC CSI is compared across 1 million bootstrap resamples of the 133 days. The p value for NC CSI exceeding SPC SCI is shown; the NC triangle is drawn with solid fill when  $p < .05$  (i.e. when NC is statistically more skillful).

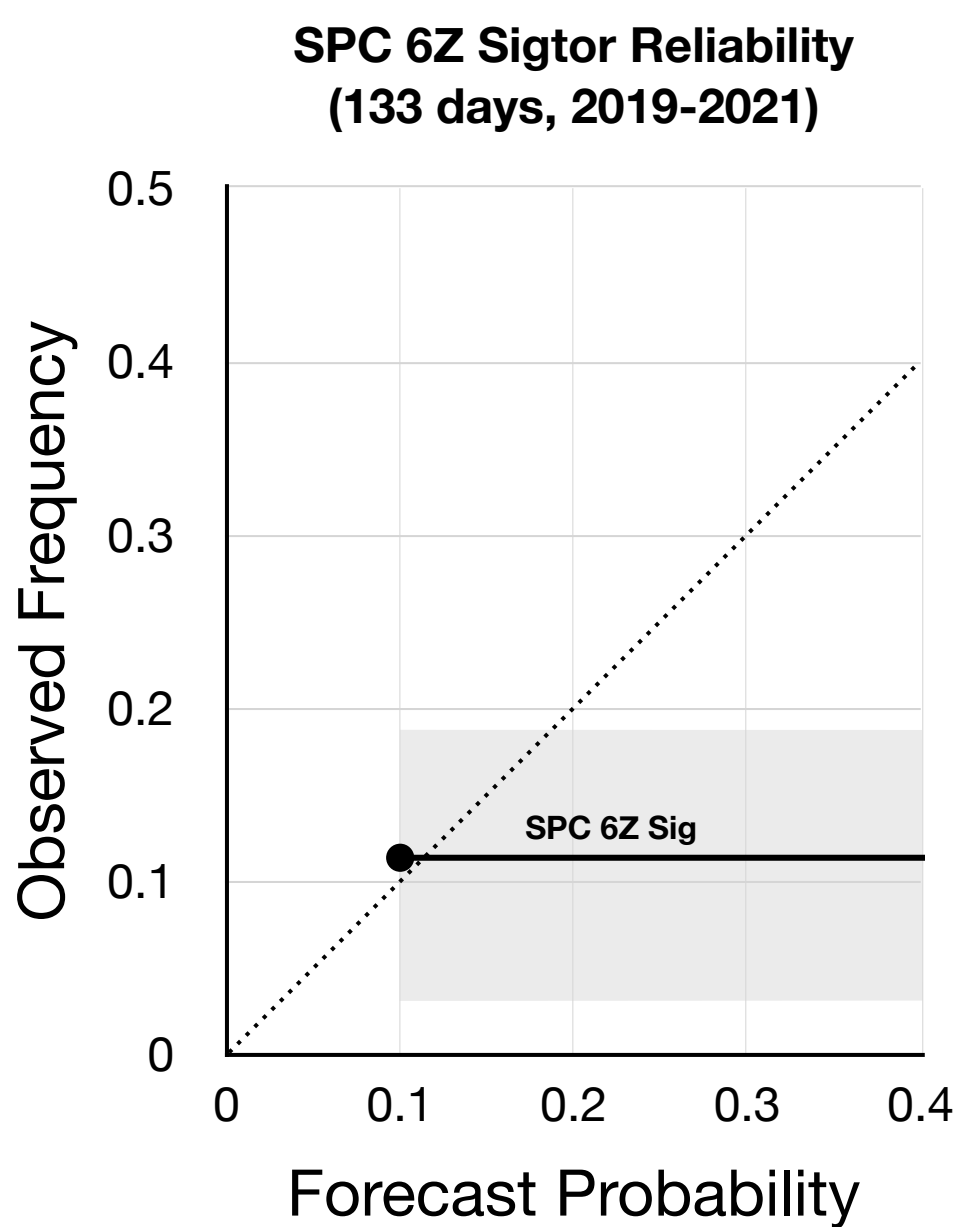
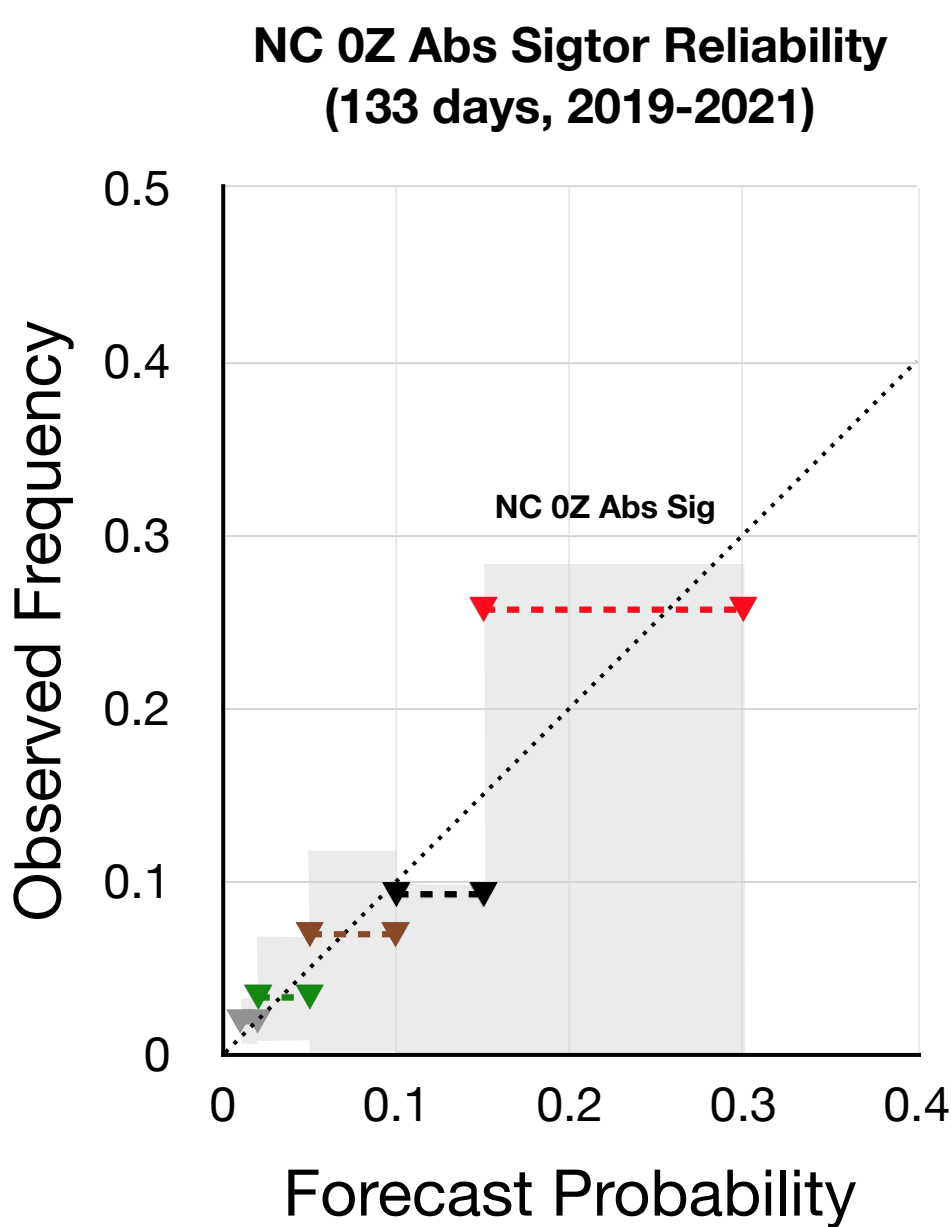
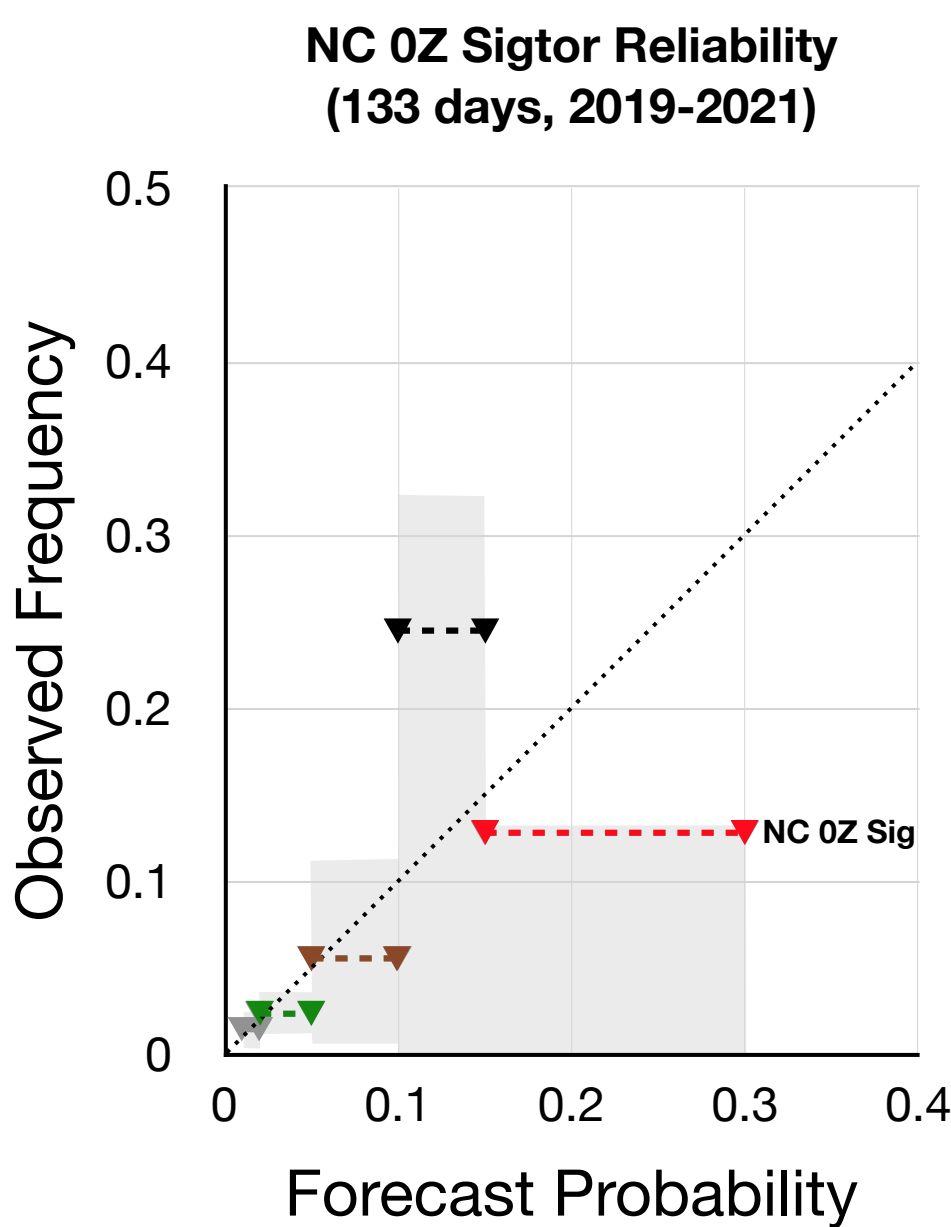


As before, but for significant severe wind. The gray triangle is the NC 2% threshold.

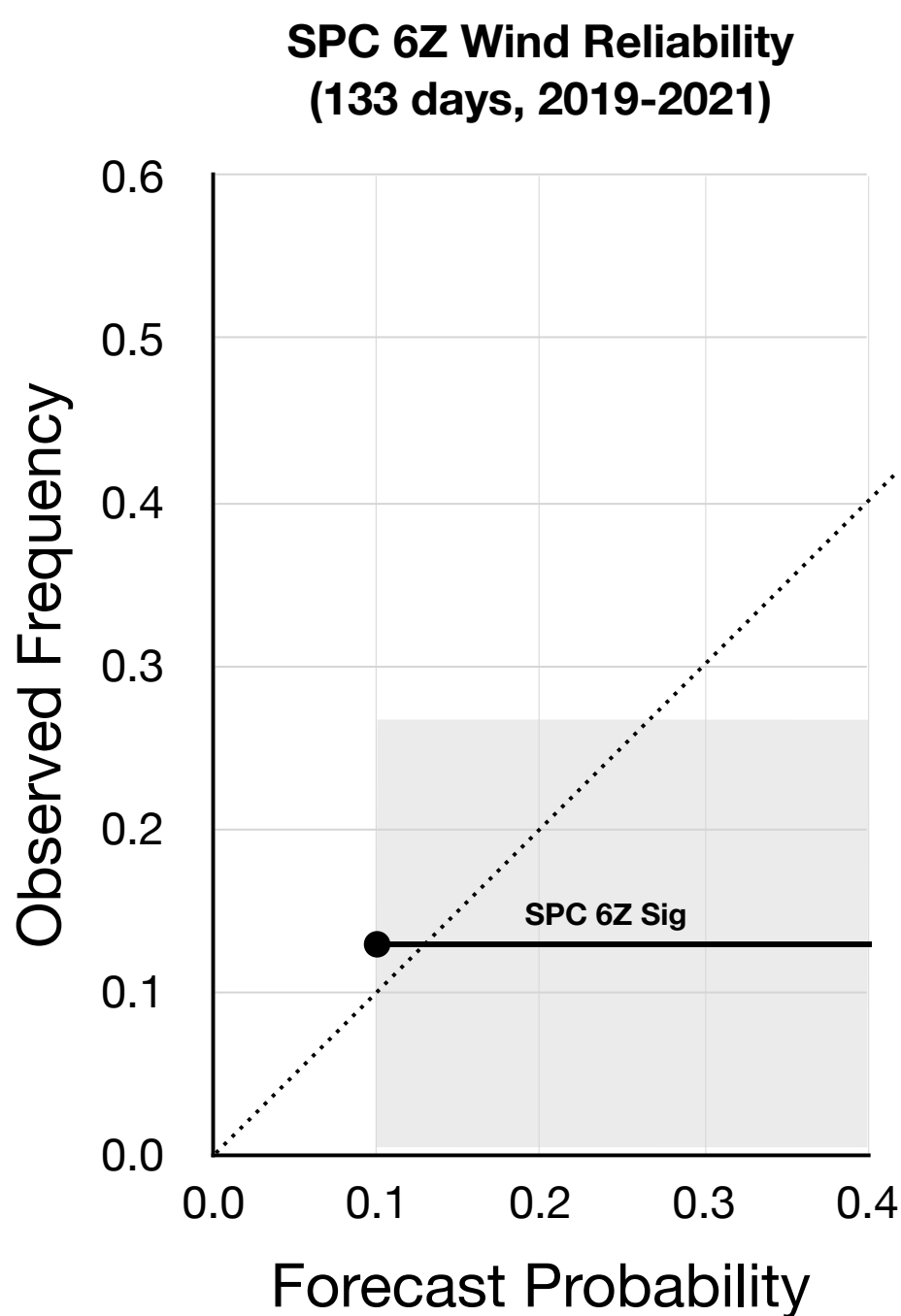
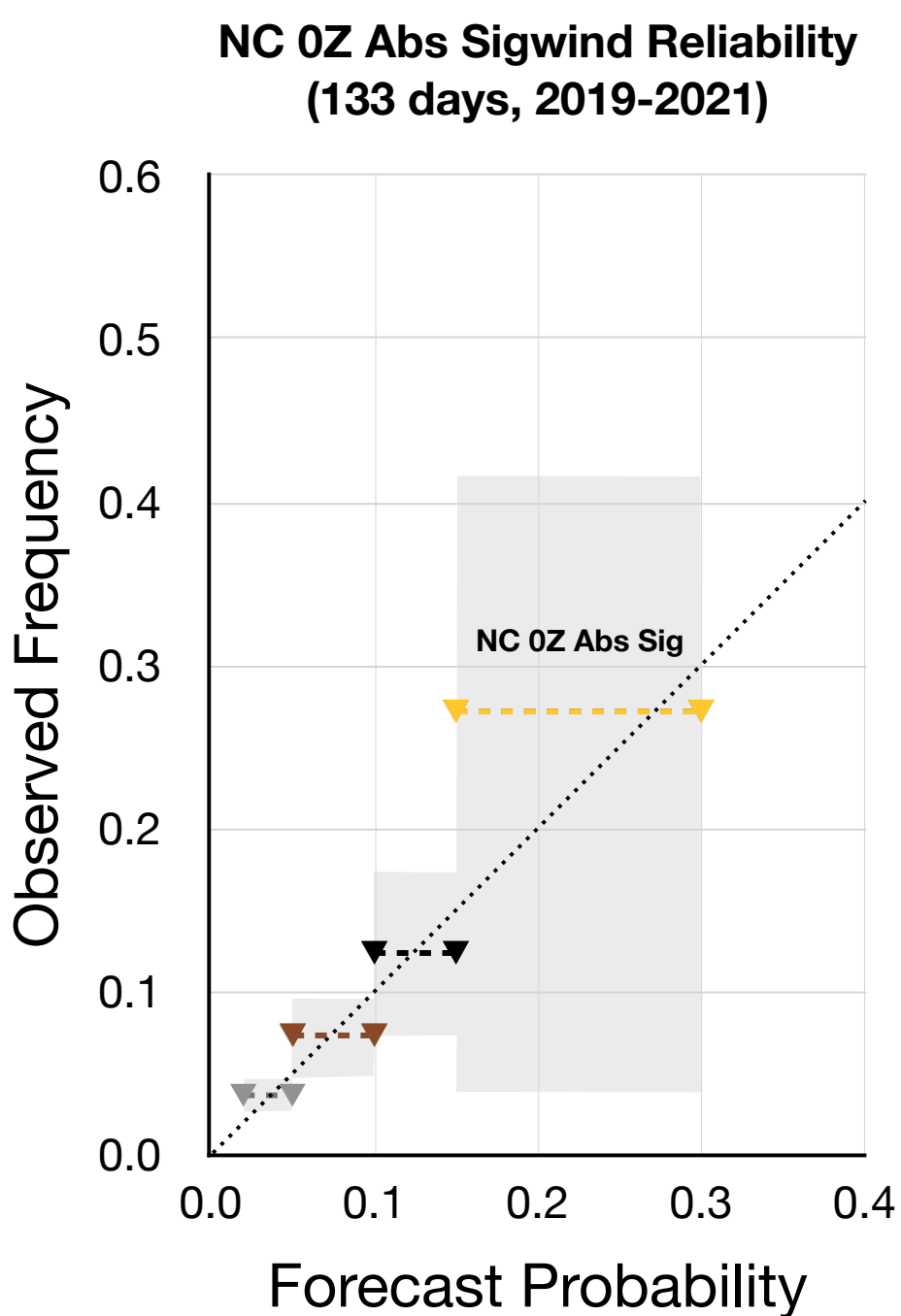
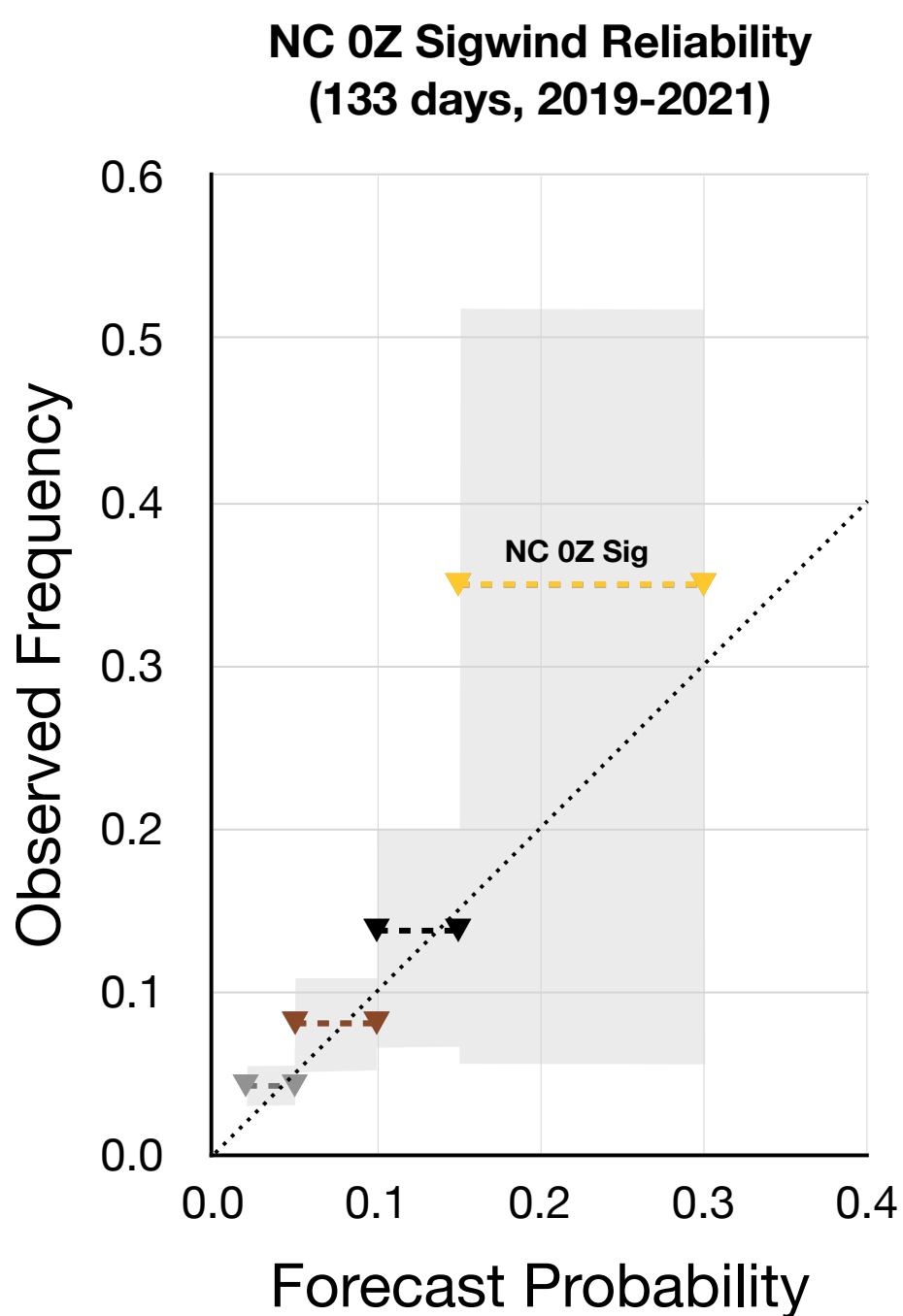


As before, but for severe hail. The gray triangle is the NC 2% threshold.

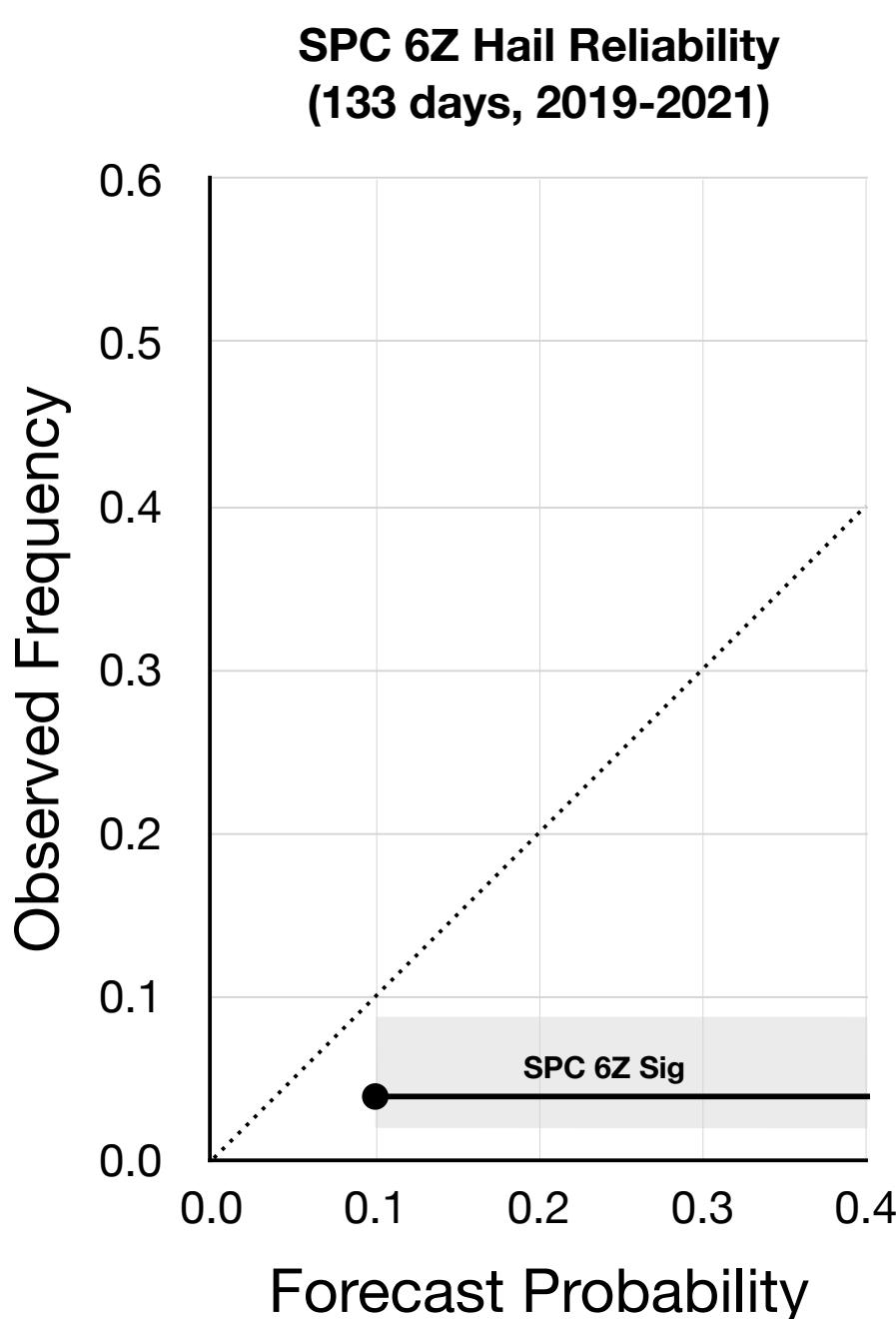
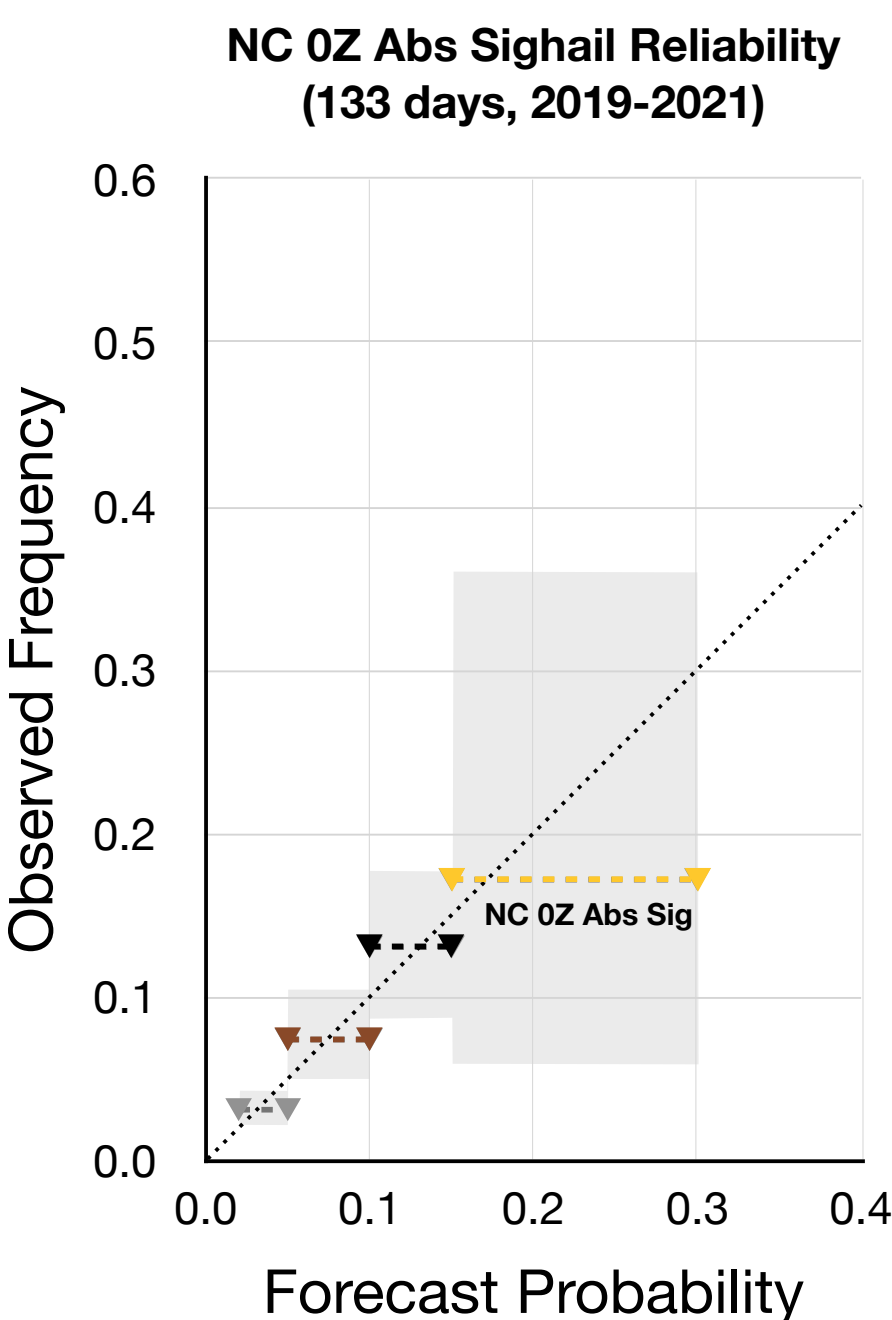
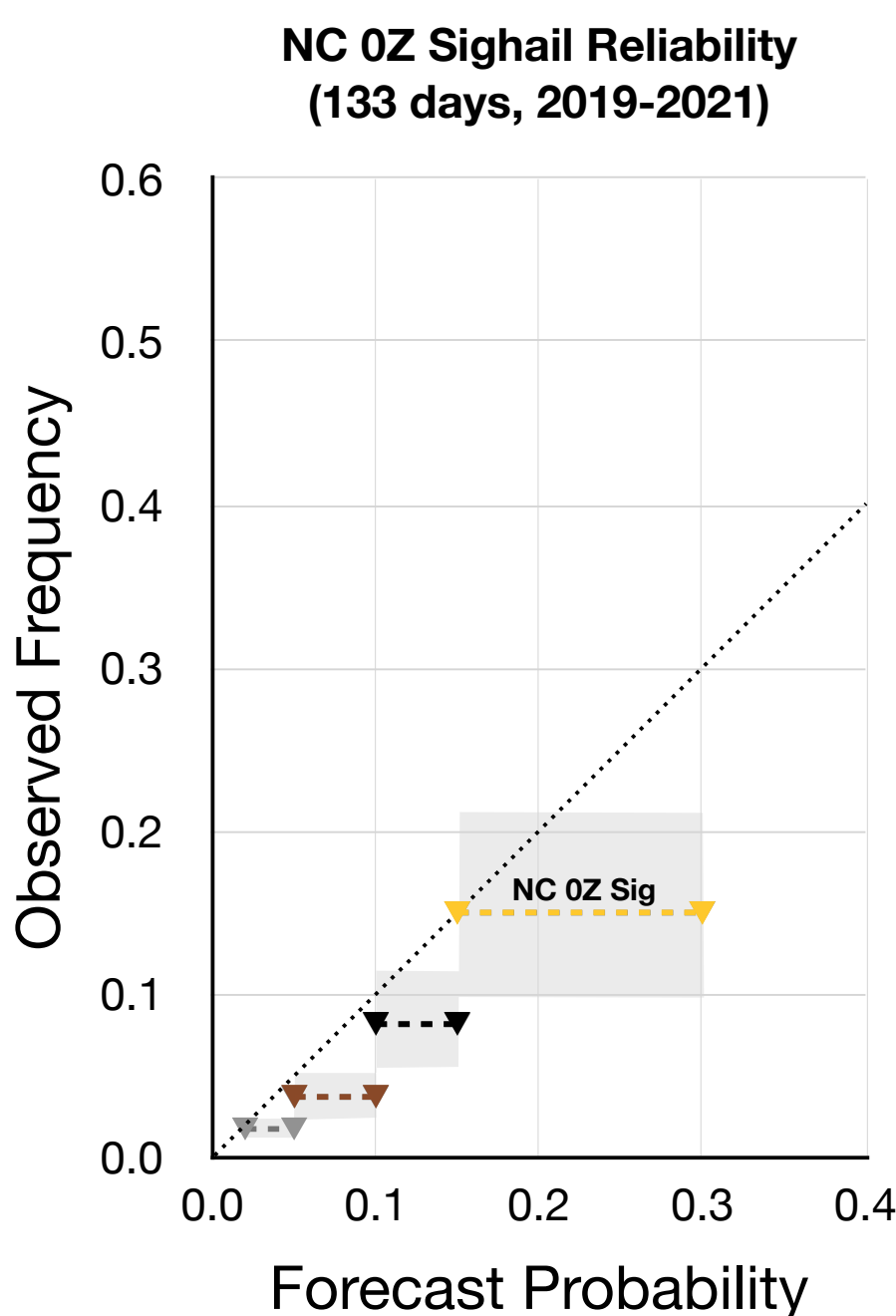
### Reliability diagrams



Reliability diagram for 0Z Nadocast (NC) 24hr significant tornado outlook with SPC-like calibration (left, triangles) and absolute calibration (center, triangles), compared to 6Z SPC day 1 outlooks (right, circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. The plots compare the forecast probability of an event to the observed frequency of at least one report within 25 miles. The bins for 1%-2%, 2-5%, 5-10%, 10-15%, 15-30%, and 30-45% forecast probabilities are each shown as horizontal lines. The 95% confidence interval for each bin is shaded, computed by bootstrapping over the 133 days. Ideal reliability is drawn as a dotted diagonal line.



As before, but for significant severe wind. Bins are 2%-5%, 5-10%, 10-15%, and 15-30%.



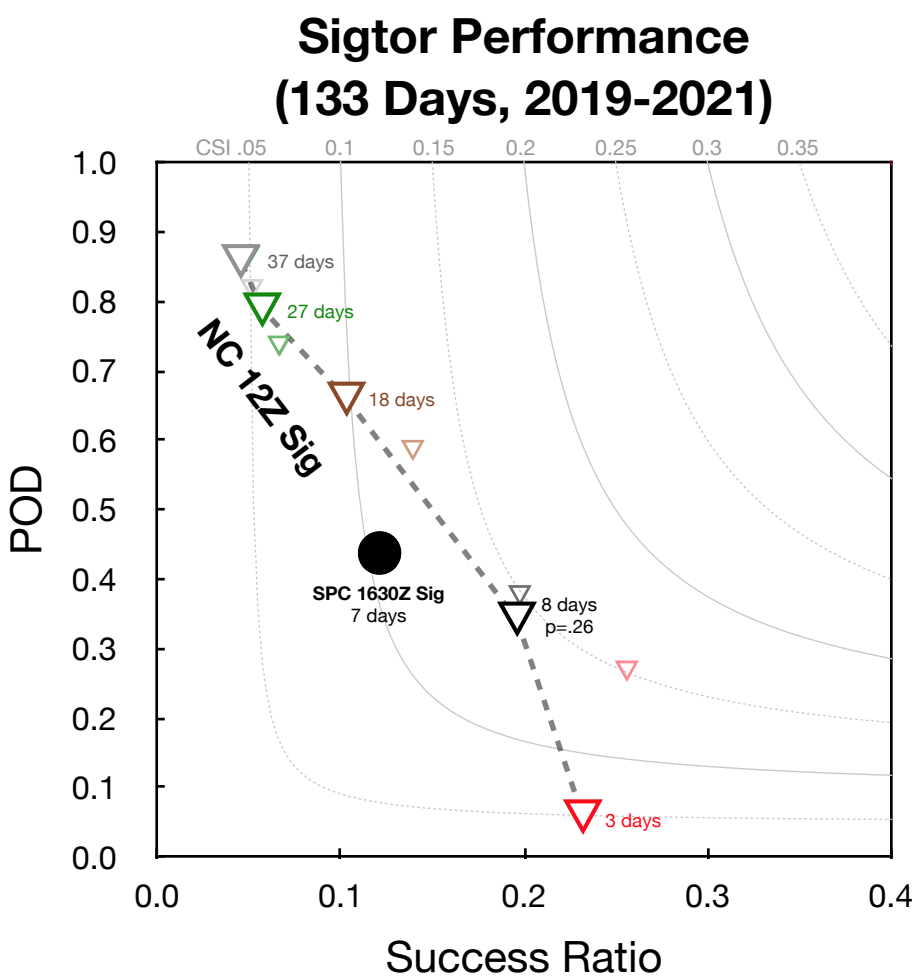
As before, but for significant severe hail. Bins are 2%-5%, 5-10%, 10-15%, and 15-30%.



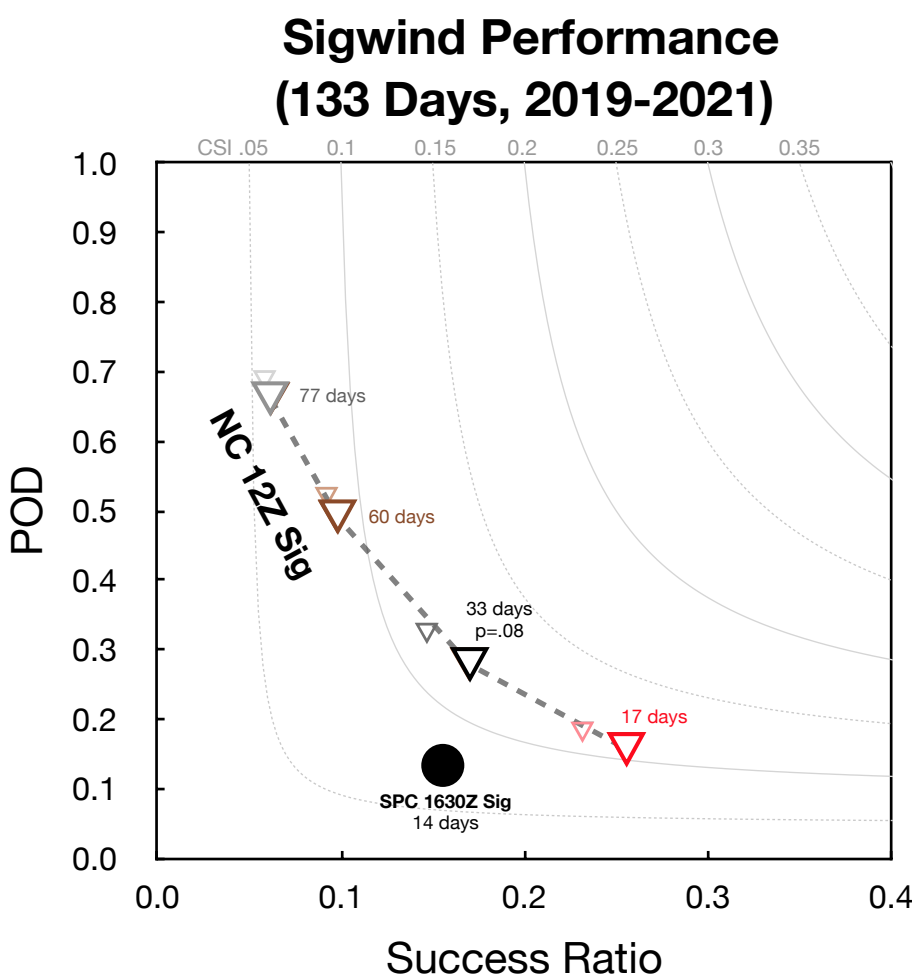
# 12Z Nadocast / 1630Z SPC

## Day 1 Sigsevere Outlooks

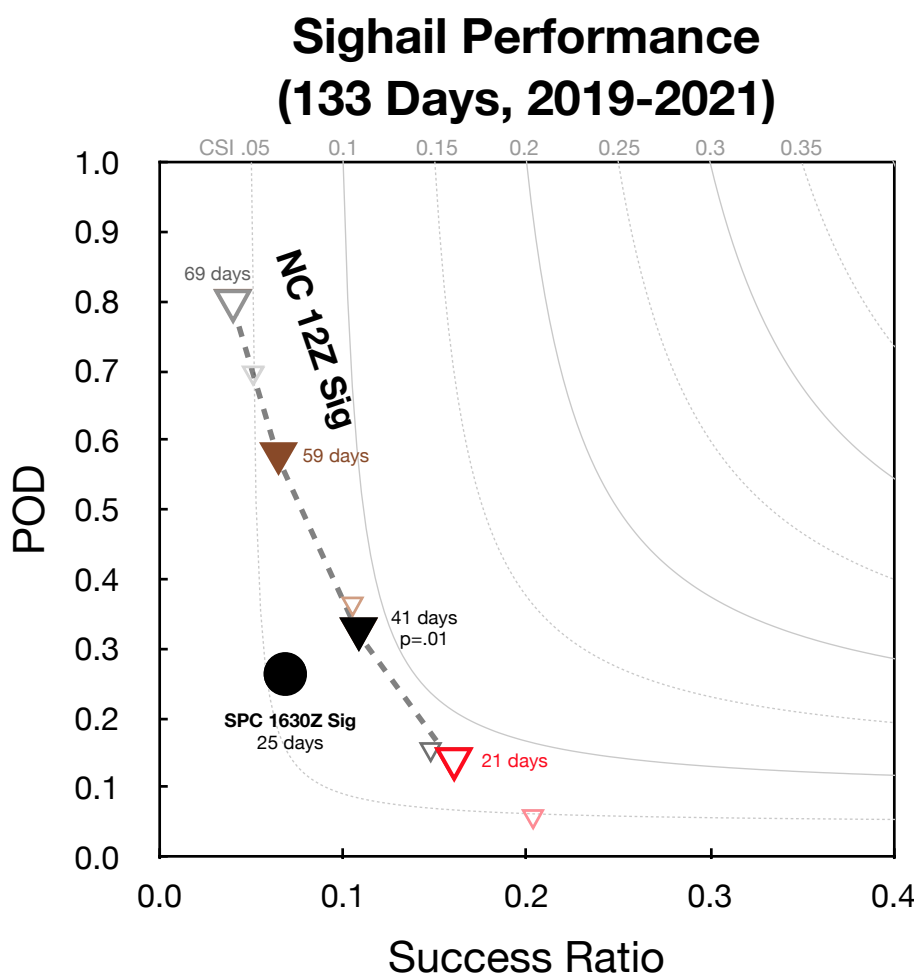
### Perf diagrams



Performance diagram for 12Z Nadocast (NC) day 1 significant tornado outlook (triangles) compared to 1630Z SPC day 1 sigtor outlooks (circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. Fainter, smaller triangles are NC with absolute probability calibration instead of SPC-like calibration. Lines of constant critical success index (CSI) are drawn as light gray curves. The performance for the NC 1% threshold is included as a gray triangle. The NC sigtor model is bounded to never output probabilities higher than the NC non-sig prediction. To assess statistical significance, for the 10% threshold the difference between the NC and SPC CSI is compared across 1 million bootstrap resamples of the 133 days. The p value for NC CSI exceeding SPC CSI is shown; the NC triangle is drawn with solid fill when  $p < .05$  (i.e. when NC is statistically more skillful).

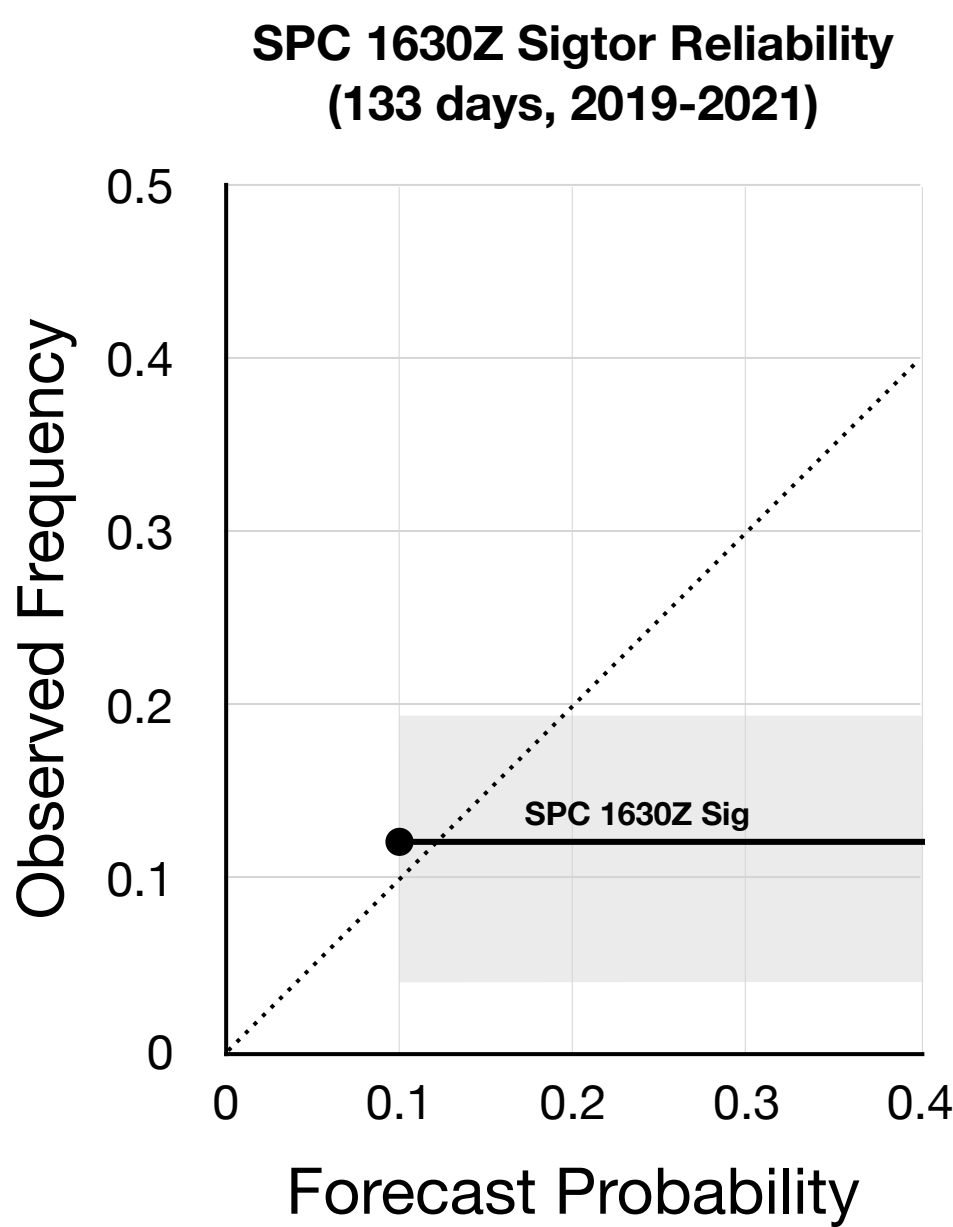
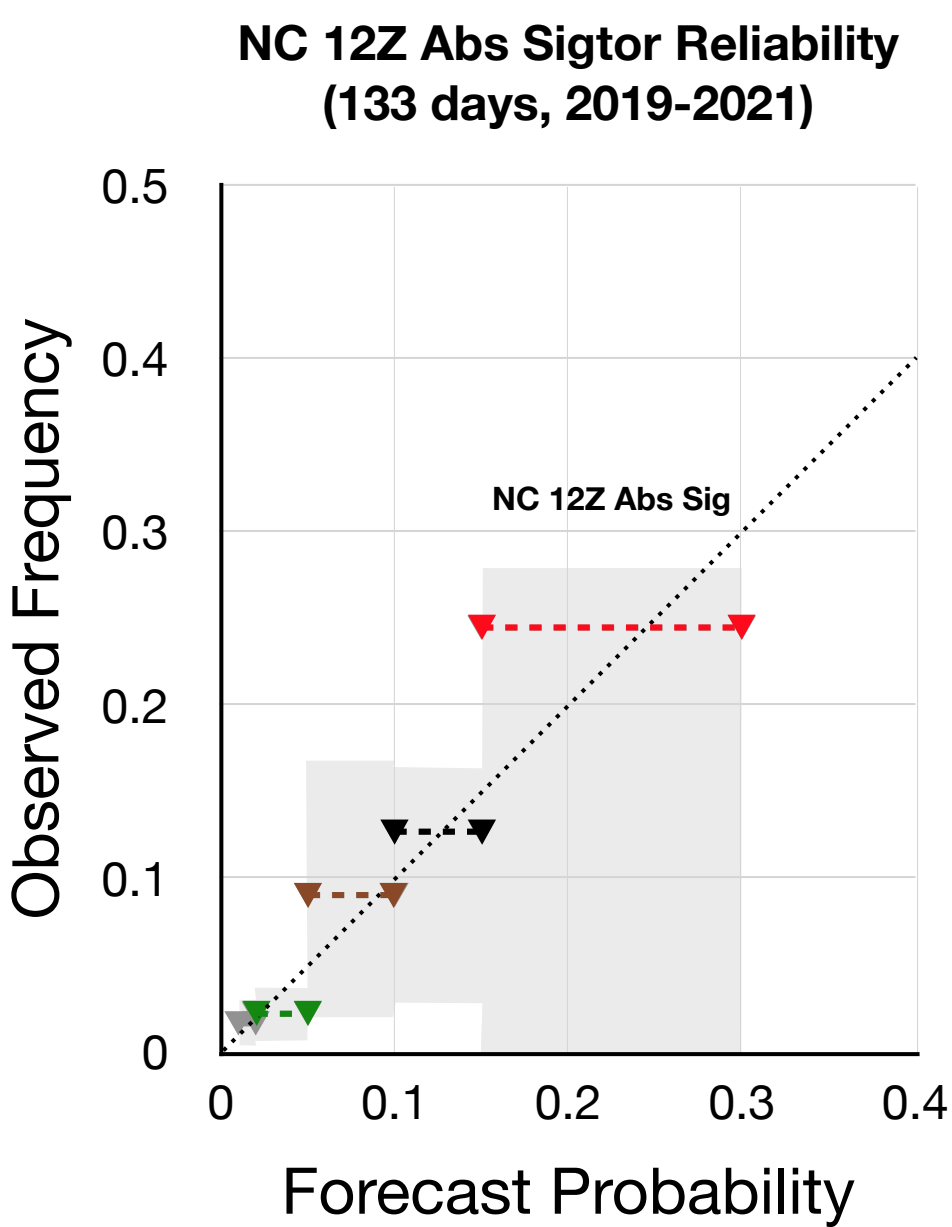
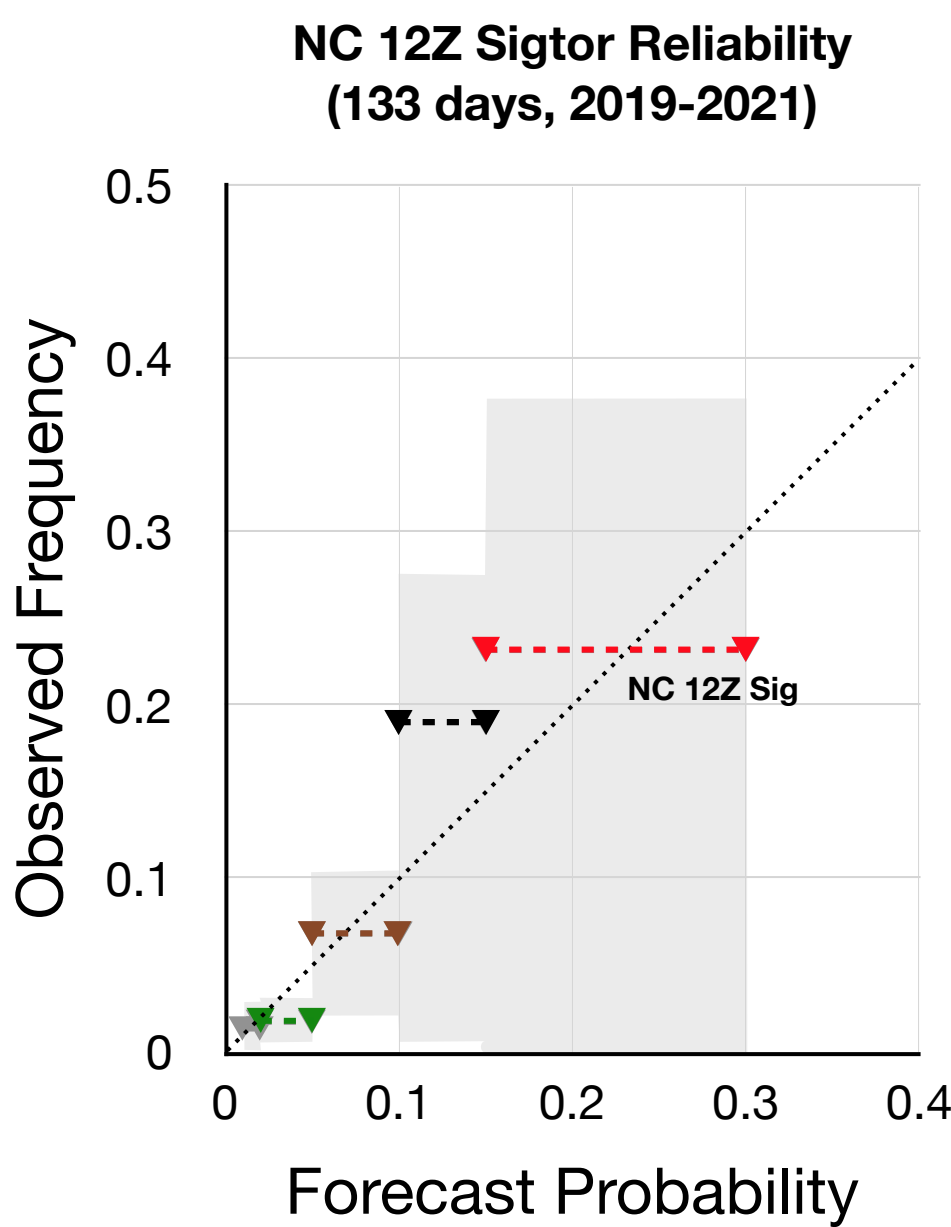


As before, but for significant severe wind. The gray triangle is the NC 2% threshold.

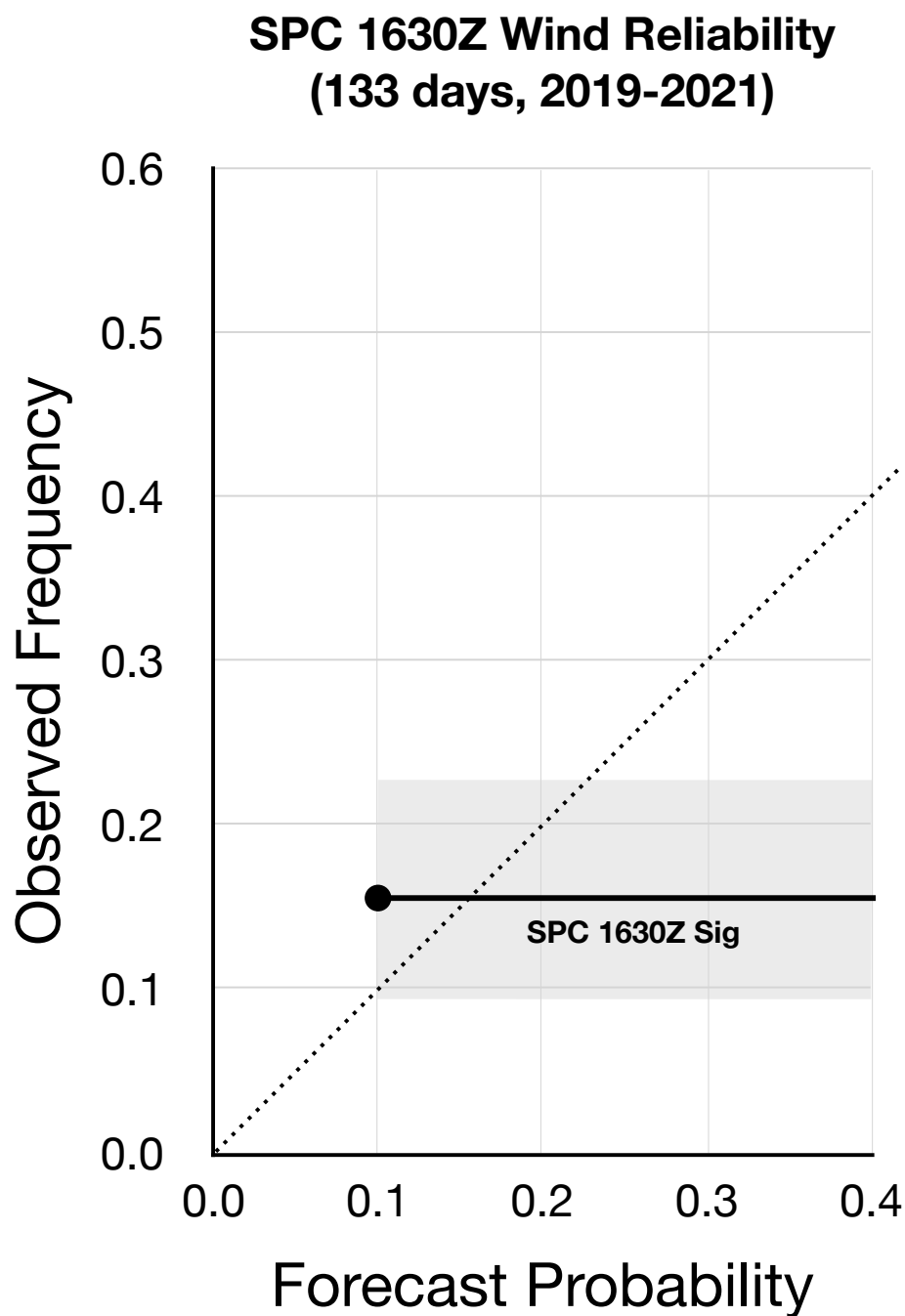
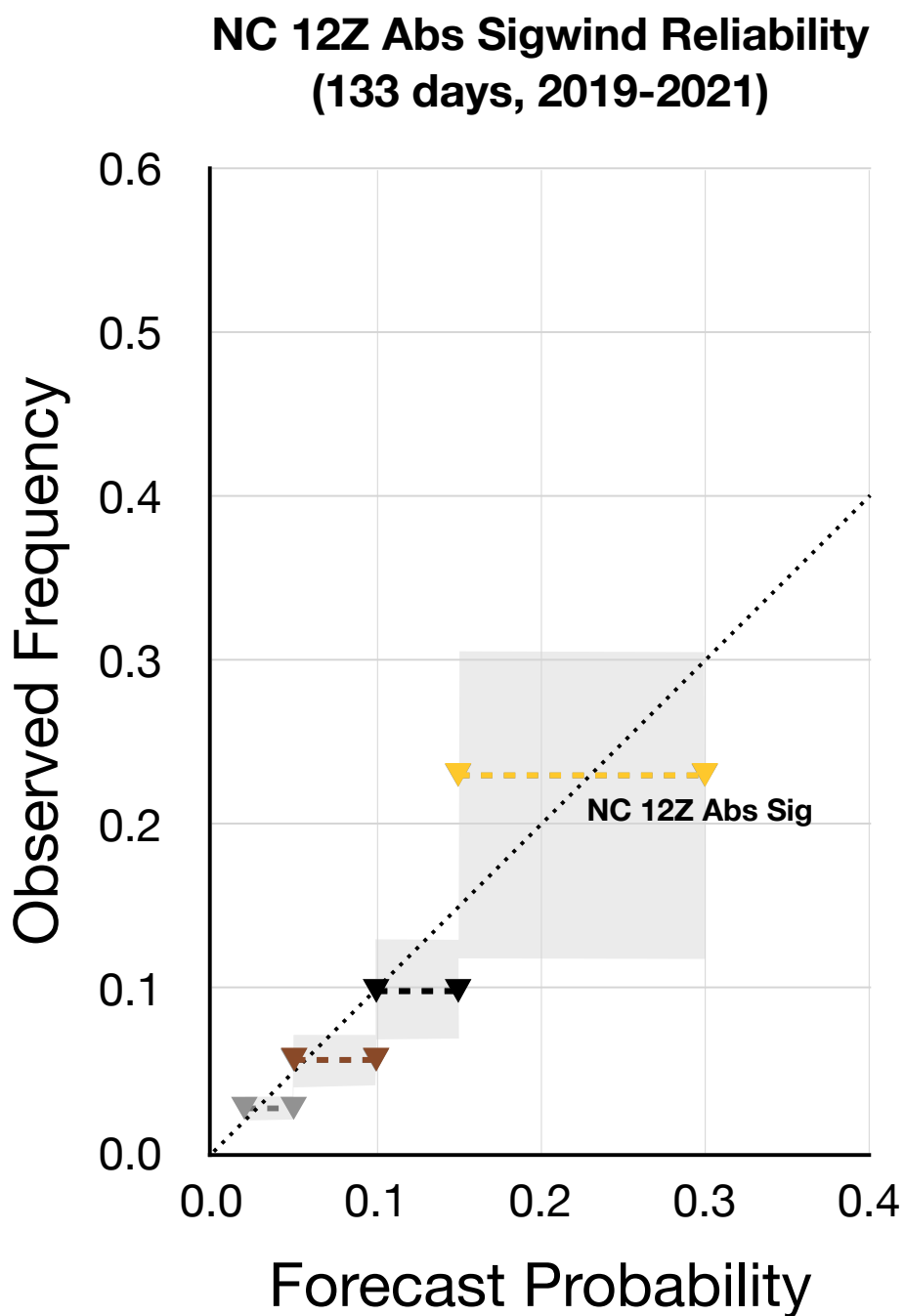
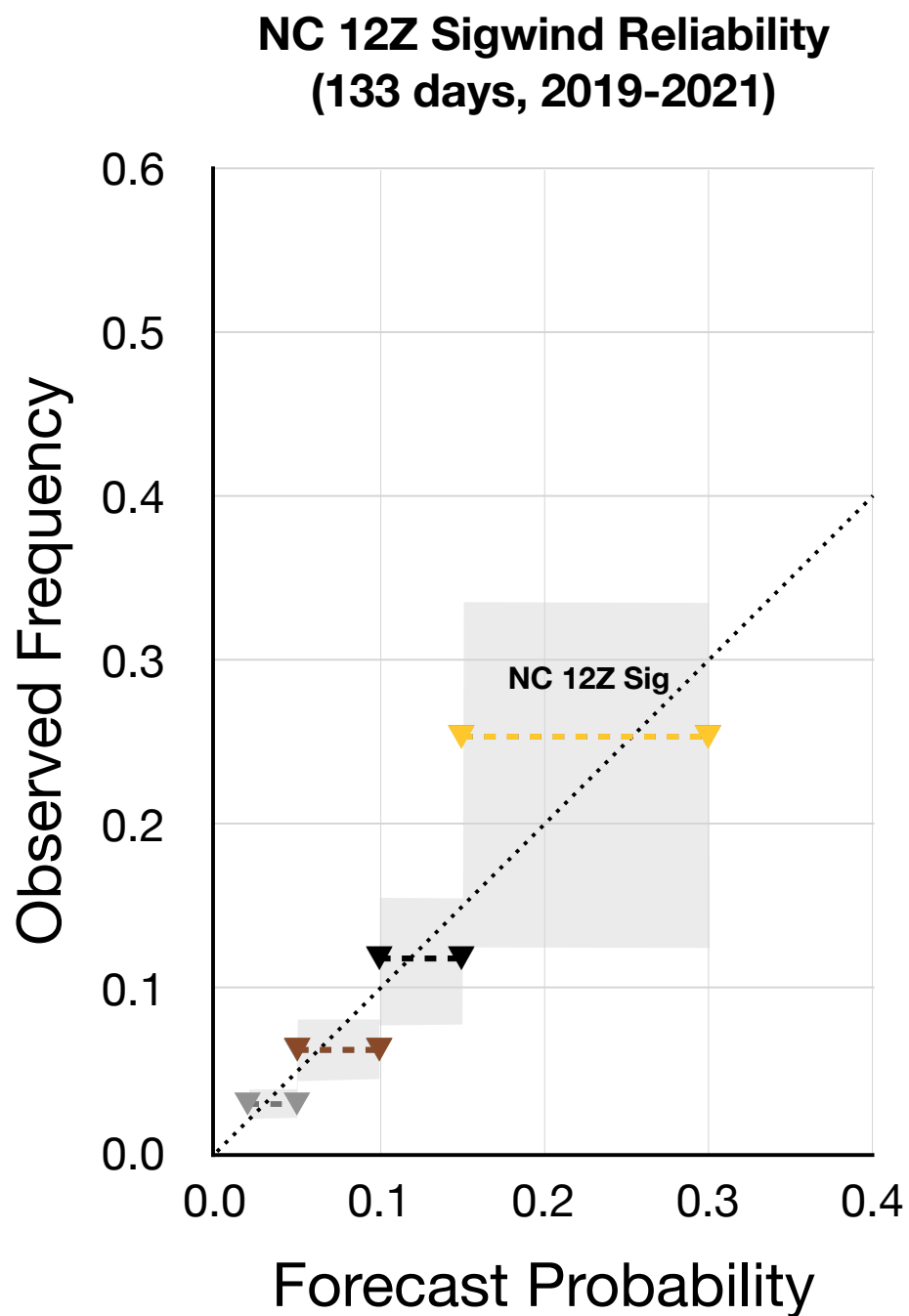


As before, but for severe hail. The gray triangle is the NC 2% threshold.

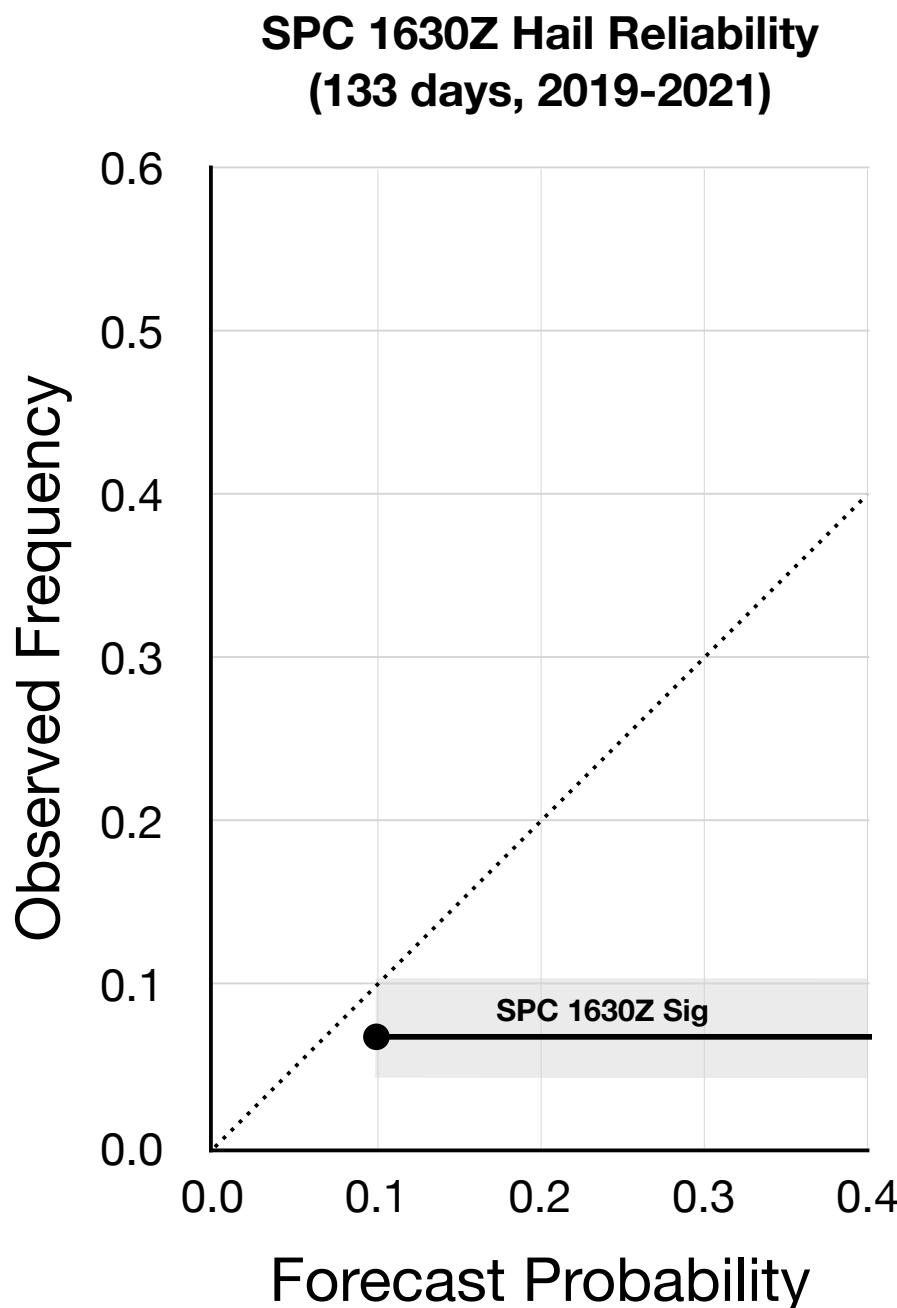
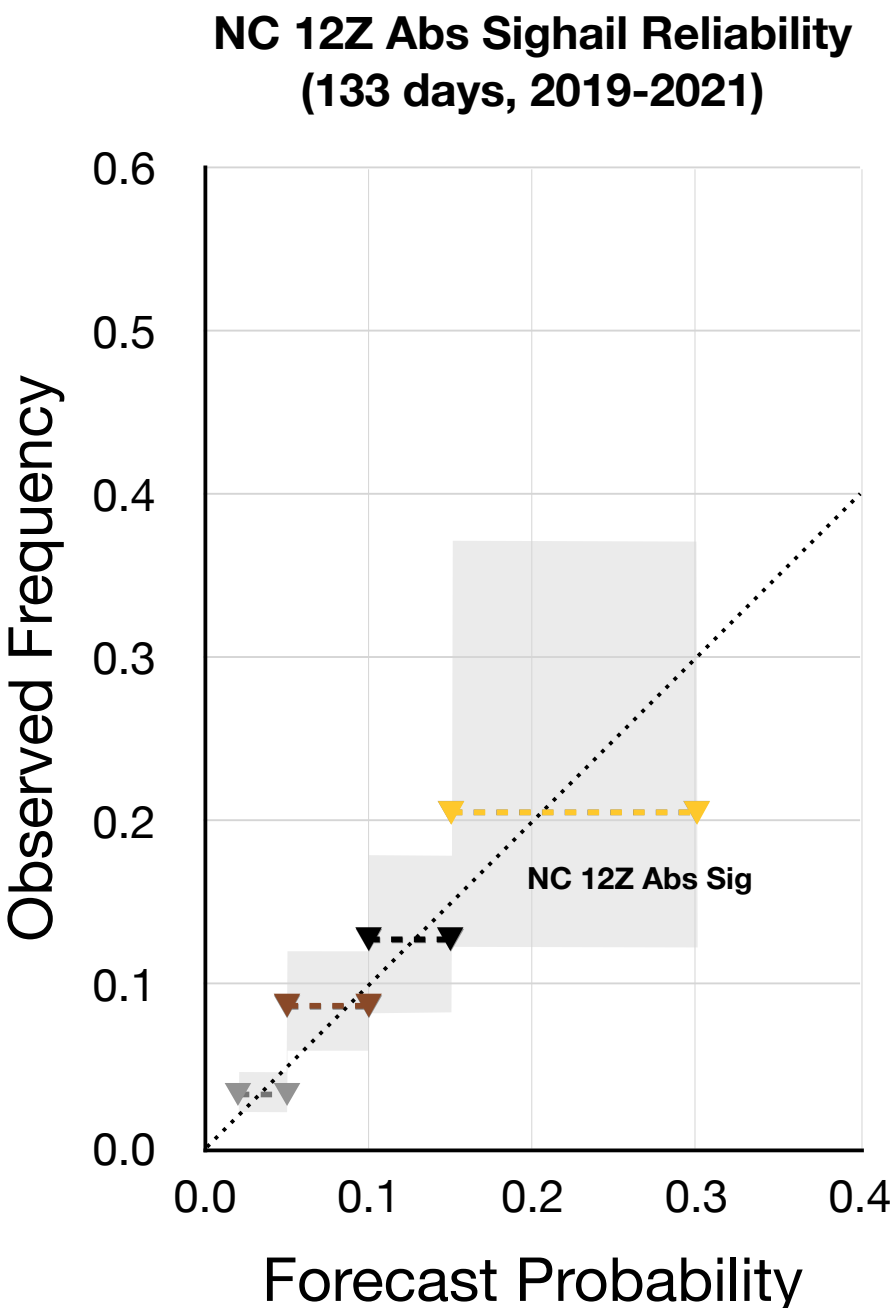
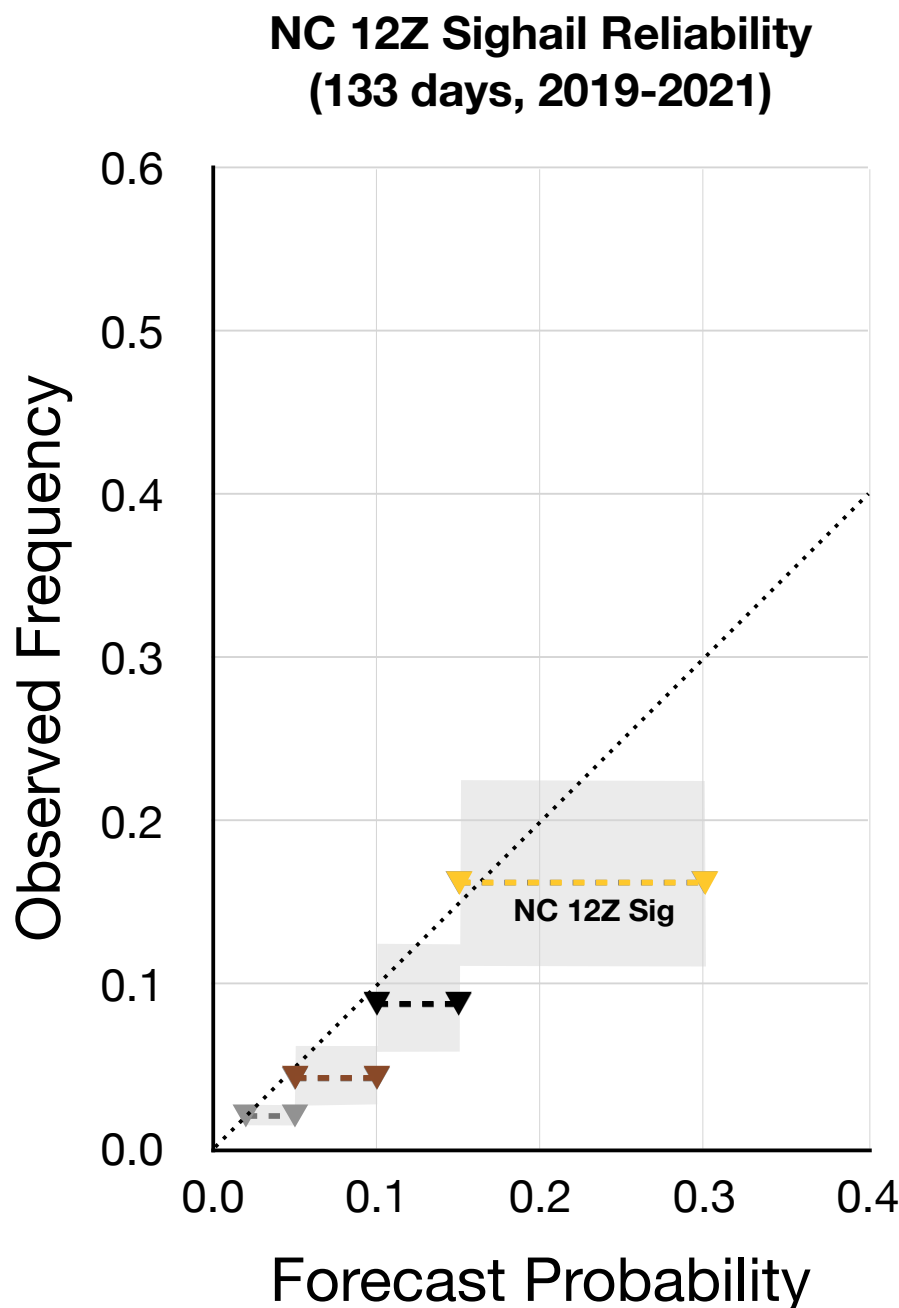
### Reliability diagrams



Reliability diagram for 12Z Nadocast (NC) day 1 significant tornado outlook with SPC-like calibration (left, triangles) and absolute calibration (center, triangles), compared to 1630Z SPC day 1 outlooks (right, circles), for the 133 days (Sundays) not seen by the models during training and calibration throughout the 2019-2021 period. The plots compare the forecast probability of an event to the observed frequency of at least one report within 25 miles. The bins for 1%-2%, 2-5%, 5-10%, 10-15%, 15-30%, and 30-45% forecast probabilities are each shown as horizontal lines. The 95% confidence interval for each bin is shaded, computed by bootstrapping over the 133 days. Ideal reliability is drawn as a dotted diagonal line.



As before, but for significant severe wind. Bins are 2%-5%, 5-10%, 10-15%, and 15-30%.



As before, but for significant severe hail. Bins are 2%-5%, 5-10%, 10-15%, and 15-30%.