Supplement of

Robust predictors for seasonal Atlantic hurricane activity identified with causal effect networks

Peter Pfleiderer et al.

Correspondence to: Peter Pfleiderer (peter.pfleiderer@climateanalytics.org)

Supplement

Definition of favorable conditions during the season

For the definition of most relevant regions for favorable conditions of SST and VWS we calculate point correlation maps between basin wide ACE and SST (and VWS). We then cluster the 3% most significantly correlated ocean grid-cells using a DBSCAN algorithm implemented in sklearn (Pedregosa et al. 2011) with the eps parameter set as 500km and a minimal sample size of 15. We finally select the most prominent regions in the Atlantic main development region as SST_{MDR} and VWS_{MDR}.

Identification of potential precursors

In order to obtain robust results, we work with flexible significance thresholds for the precursor selection. For SST (MSLP) we cluster the 5% (7.5%) most significantly correlated grid-cells into regions using the DBSCAN clustering with the eps parameter set as 500km (1500km) and a minimal sample size of 15 (20). When searching for potential precursors in MSLP in March we create larger regions by considering more grid-cells with clustering parameters that lead to larger regions as we don't expect small features in MSLP to have a relevant influence on hurricane activity 4 months later. For potential precursors in SST and MSLP we exclude grid-cells northward of 60N and southward of -60S.

Identification of causal precursors

When applying the CEN, a significance threshold can be set to define the level at which a link in the network is considered as significant. For our automated application of the CEN in the cross-validation framework we work with a flexible threshold to obtain robust results: We gradually increase the significance threshold until we obtain at least one (for the precursors of SST_{MDR}) or two (for precursors of VWS_{MDR}) most relevant precursors, for which the effect on the target variables is strongest.

Receiver Operating Characteristic (ROC)

Besides correlations, the receiver operating characteristic (ROC), can be calculated to quantify the forecast skill. Here, events are classified into two classes (e.g. above median and below median) and the forecast is evaluated with regard to its skill to predict the two classes by altering the threshold that separates the forecasted values into the two classes. The result can be visualized as a ROC curve.

Logistic regression classifier

We create logistic regression classification models for three types of events: above 33rd percentile seasons, above median seasons and above 66th percentile seasons. We use the logistic regression implementation from sklearn (Pedregosa et al. 2011).

Reliability diagrams

We show reliability diagrams to evaluate the skill of our logistic regression model. Forecast probabilities are binned into 5 groups and each group is plotted as a function of observed frequency (y-axis) versus mean forecast probability (x-axis). Ideally these points would lie on the diagonal. The green area (in Fig. 6 & 9) shows where the forecast is superior to a climatological forecast and the brier skill score (BSS) is greater than zero (Mason 2004). In the light-green area, forecasts have skill

compared to random guessing. These diagrams have been suggested to evaluate the usefulness of forecasts (Weisheimer and Palmer 2014).

References

Mason SJ (2004) On using "climatology" as a reference strategy in the Brier and the ranked probability skill scores. Mon Weather Rev 132:1891–1895. doi: 10.1175/1520-0493(2004)132<1891:OUCAAR>2.0.CO;2

Pedregosa F, Varoquaux G, Gramfort A, et al (2011) Scikit-learn: Machine Learning in {P}ython. J Mach Learn Res 12:2825–2830

Weisheimer A, Palmer TN (2014) On the reliability of seasonal climate forecasts. J R Soc Interface 11:20131162. doi: 10.1098/rsif.2013.1162

Additional Figures
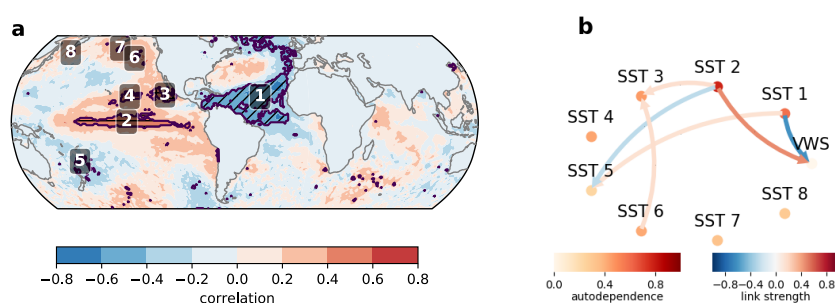
Figure S1 | **Precursors of VWS$_{MDR}$ in a training set containing the years 1979-2015:** *as Figure 3 but with the full causal effect network in (b).*
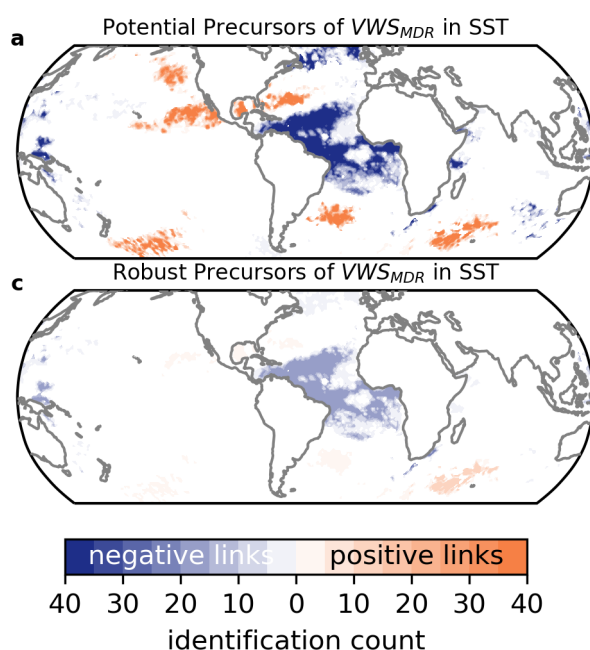

Figure S2 | **Potential and robust precursors of VWS$_{MDR}$ in March.** *As Figure 7a-c but for SST precursors.*
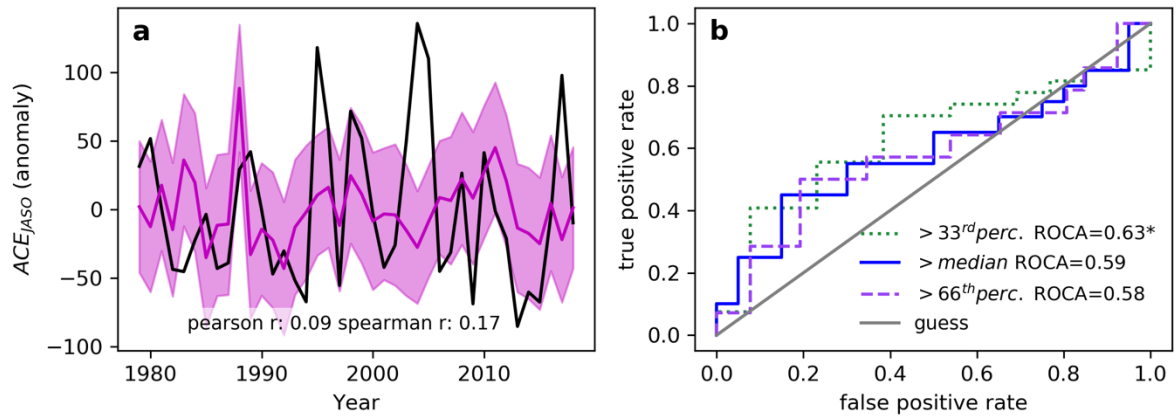
*Figure S3 | **Hindcast skill based on March ERA reanalysis with detrended ACE and detrended precursors.** Same as Figure 8 but with detrended ACE and detrended predictors.*
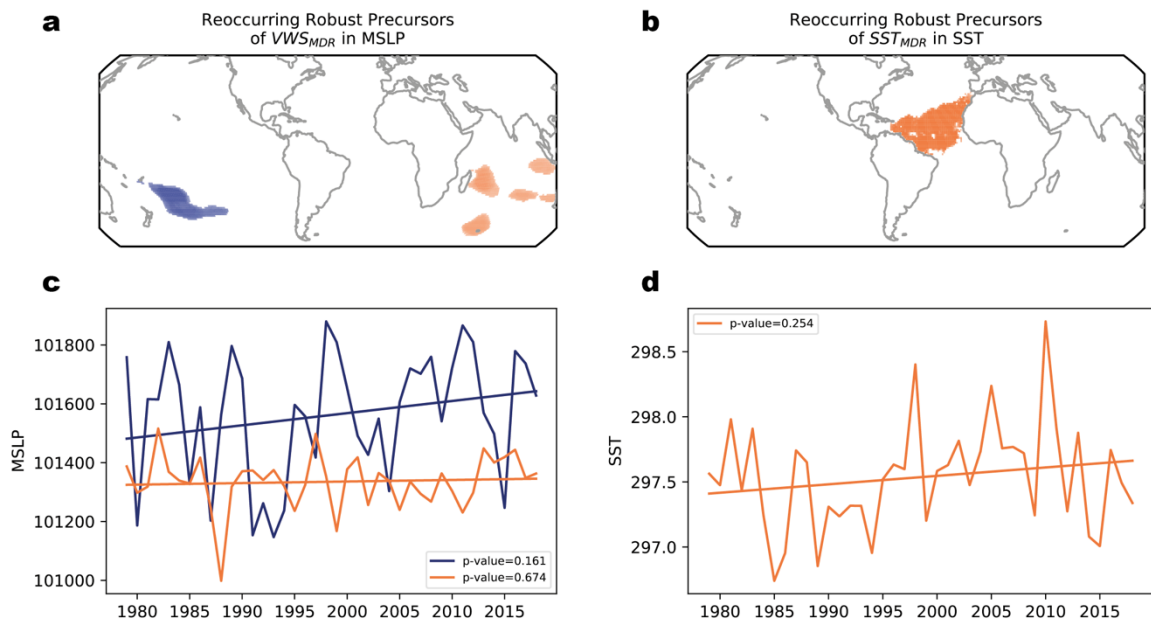

*Figure S4 | **Trends in March precursors.** March precursors of VWS_{MDR} (a) and SST_{MDR} (b) in ERA5. Here grid-cells that are identified as robust precursors in the majority of training sets are considered. Time series in the regional average of these precursors and their trends are shown in the second row for VWS_{MDR} precursors (c) and SST_{MDR} precursors (d). In the legend p-values of trend tests are indicated.*
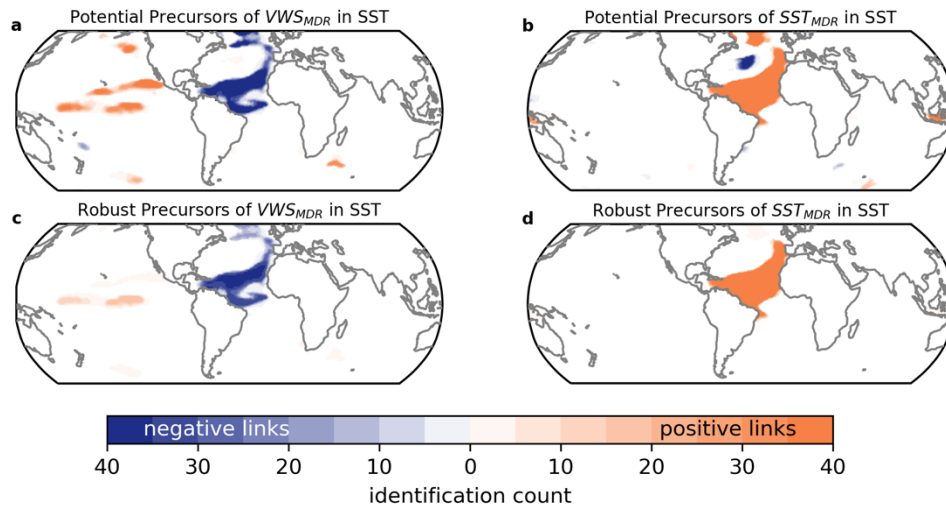
*Figure S5 | **Potential and robust precursors of VWS$_{MDR}$ and SST$_{MDR}$ in May.**
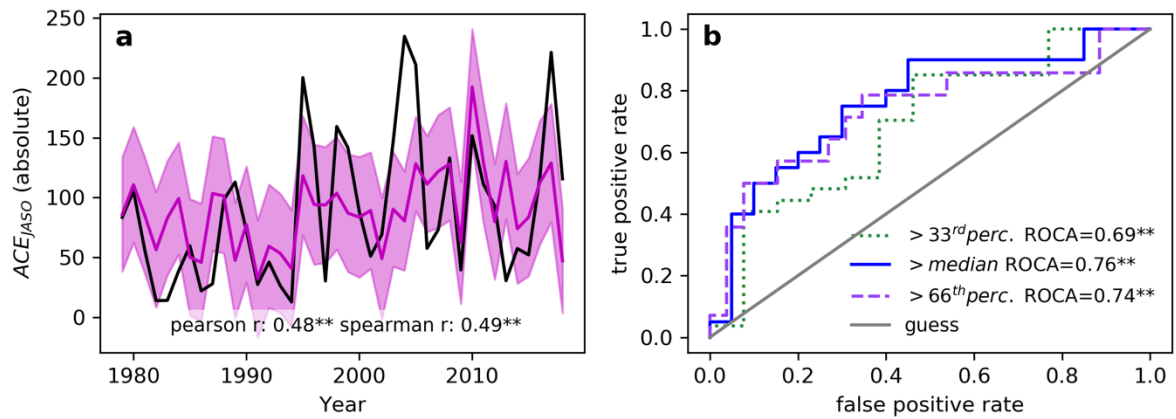Same as Figure 4 but with the JRA55 reanalysis.*

*Figure S6 | **Hindcast skill based on May JRA55 reanalysis.** Same as Figure 5 but
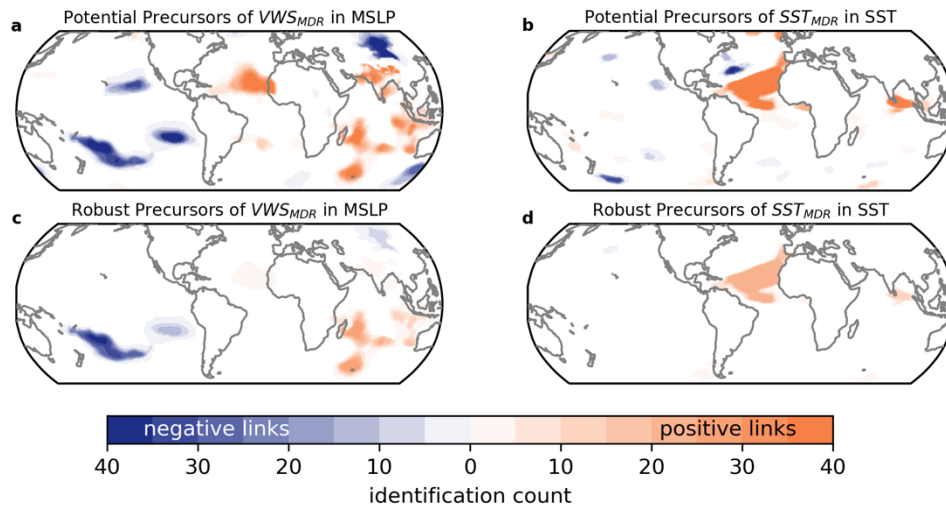with the JRA55 reanalysis.*

*Figure S7 | **Potential and robust precursors of VWS_MDR and SST_MDR in March.** Same as Figure 7 but with the JRA55 reanalysis.*
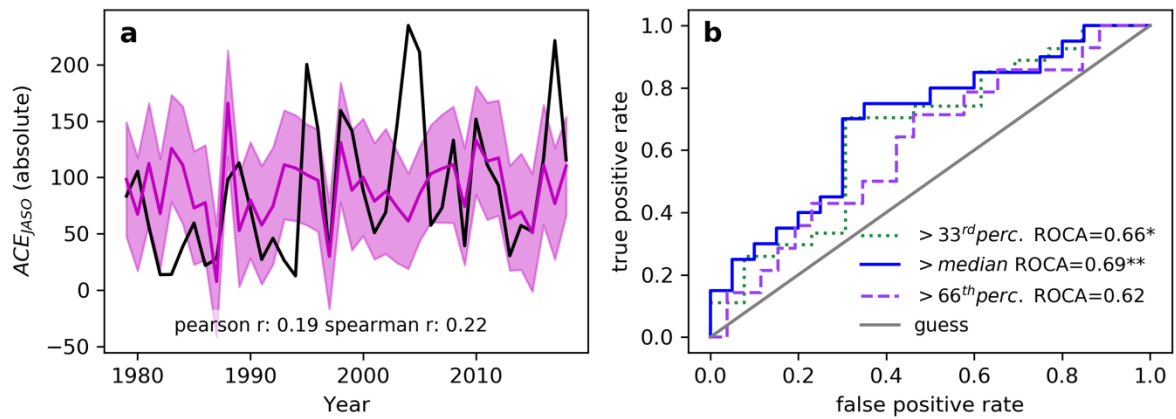

*Figure S8 | **Hindcast skill based on March JRA55 reanalysis.** Same as Figure 8 but with the JRA55 reanalysis.*
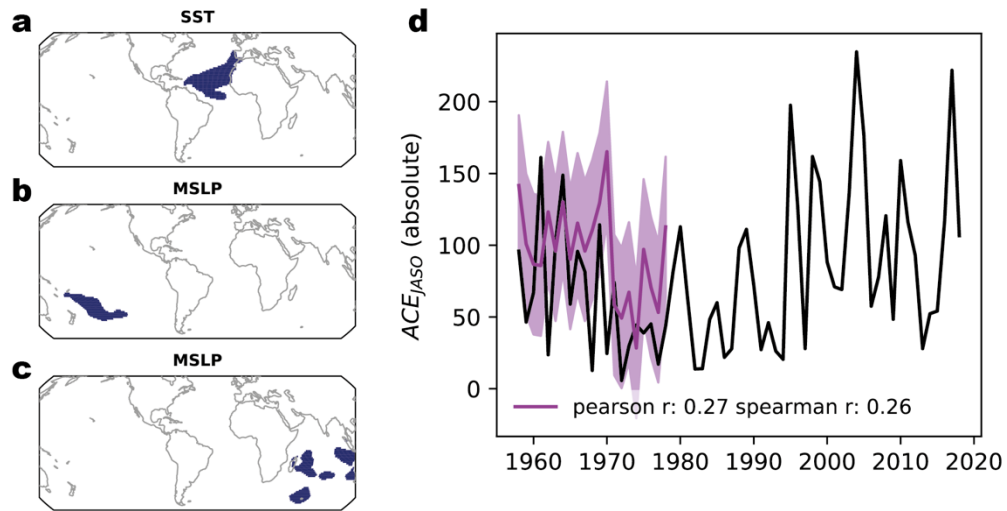
Figure S9 | **Skill of the March forecast model in the early period of JRA55 (1958-1978).** *Using the period 1980-2018 as training period, a forecast model for ACE in July-October is constructed based on a robust precursor of SST_{MDR} (shown in a) and two robust precursors of VWS_{MDR} (b and c) in March. The hindcast performed on the testing period (1958-1978) is plotted in magenta in panel (d). The Pearson correlation coefficient and the Spearman correlation coefficient are indicated in the legend for the test period. Both correlation coefficients are not significant.*