Geophysical Research Letters

Supporting Information for

Potential for early forecast of Moroccan wheat yields based on climatic drivers

J. Lehmann[1], M. Kretschmer[2], B. Schauberger[1], and F. Wechsung[1]

[1]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany

[2]University of Reading, Reading, United Kingdom

Corresponding author: Jascha Lehmann (jlehmann@pik-potsdam.de)

Content:

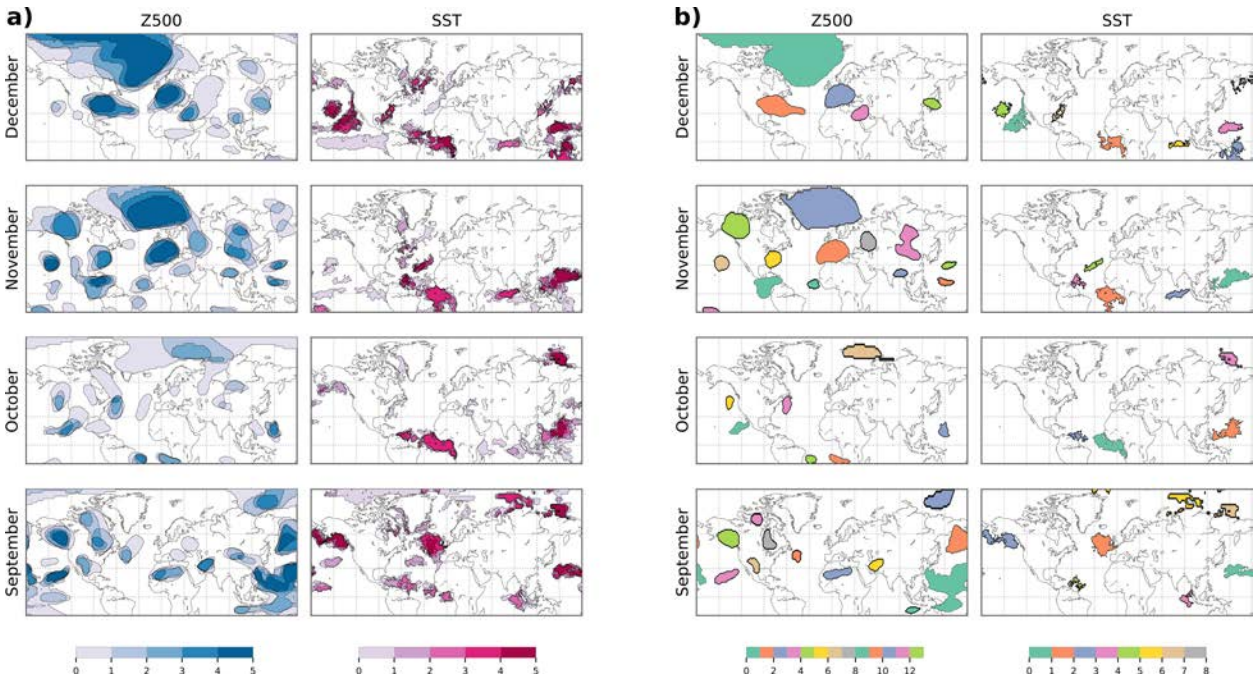

Fig. S 1: Robustness of potential precursors. (a) Colors indicate how often a grid cell is attributed to a potential precursor region. (b) Grid cells which are selected more than 50% of the time in (a) are grouped to robust potential precursors which show strong agreement with potential precursors found for the full time period (Fig. 1 in main manuscript). The significance threshold was raised to p-value < 0.04 to compensate for the shorter time series length of the subsamples.

Potential precursors are calculated based on subsamples of 30 years which are derived by iteratively removing 7-year periods from the full time series with each year removed only once. Dark blue and dark red regions in Fig. S1a indicate that similar potential precursors are detected throughout the studied time period. For better comparison with results from the full time series (Fig. 1 in main manuscript) we agregate results from all subsamples into one figure by showing only those regions that were detected more than 50% of the time. This gives an impression of the most robust potential precursor regions. Varying the threshold level at which significance is defined (p-value $< 0.02, 003, 0.04, 0.05$) leads to similar potential precursor regions, however, with small effects on the overall number and spatial extend of the regions as one would expect. We also tested the robustness based on 30-year running time periods and found consistent results.

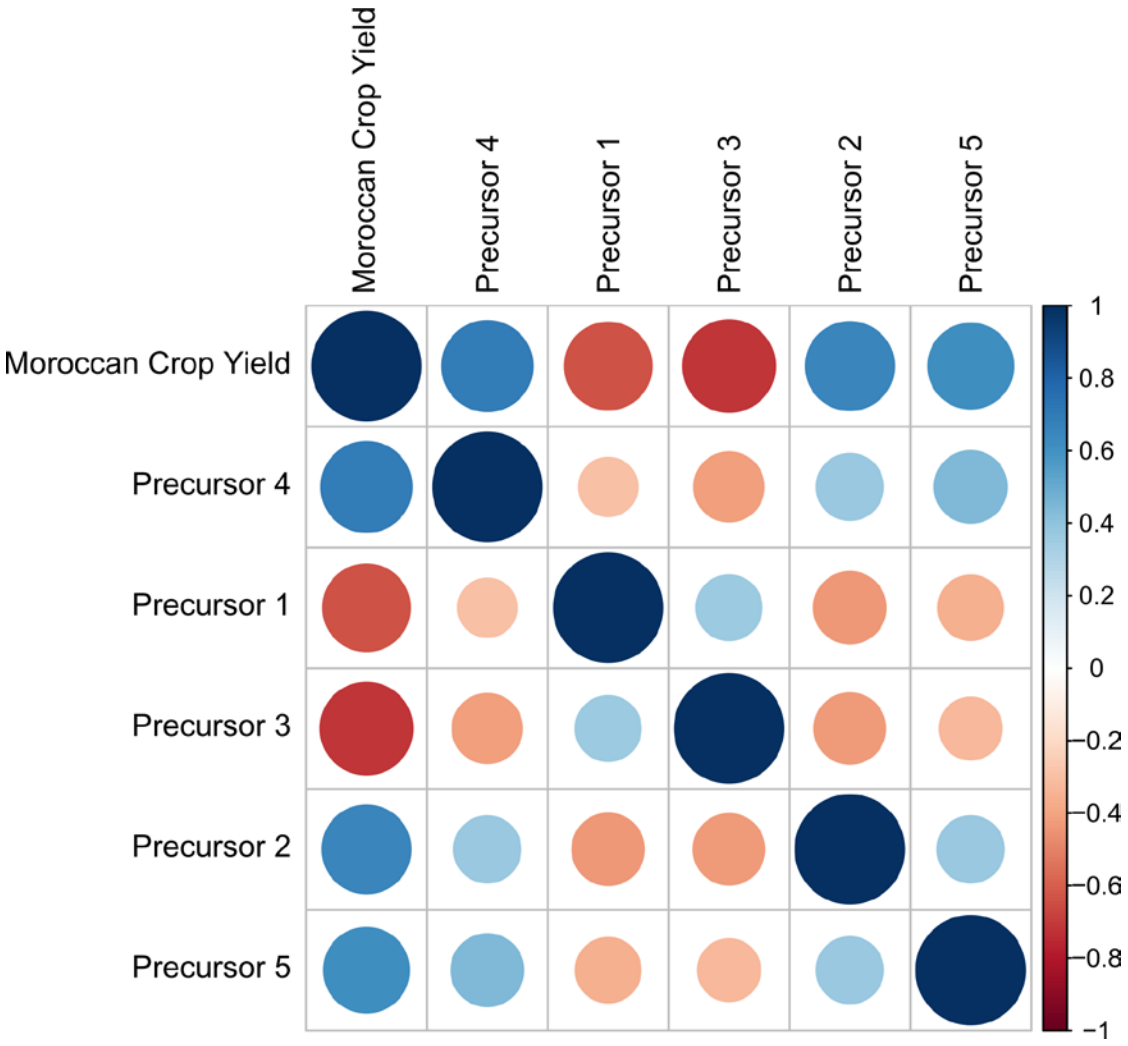

Fig. S 2: Correlogram of causal precursor time series and Moroccan wheat yields anomalies.

Causal precursor anomalies show strong correlation with yield anomalies as indicated by the size and color of the circles. Correlation between individual causal precursors is much smaller as expected from the partial independence tests (see step 2 in Methods section of main manuscript).

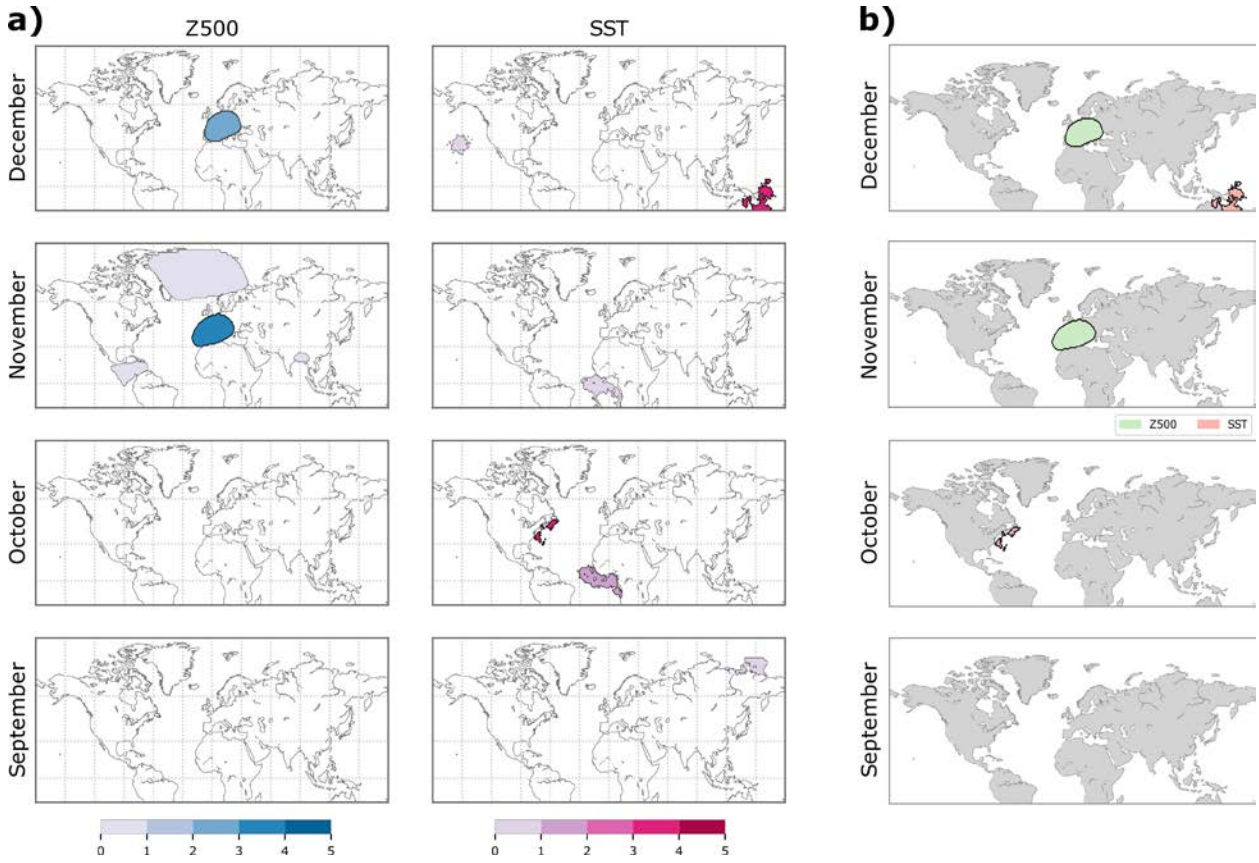

Fig. S 3: Robustness of causal precursors. (a) Colors indicate how often a grid cell is attributed to a causal precursor region. (b) Causal precursors which are selected at least 50% of the time in (a) show strong overlap with causal precursors extracted from the full time period (Fig. 2 in main manuscript).

To test the robustness of the causal selection step we extract causal precursors from the given set of potential precursors (Fig. 1 of the main manuscript) by using only data from 30-year subsamples. Similarly to the approach in Fig. S 1, subsamples are derived by iteratively removing a 7-year period from the full time series with each year removed only once. On average, each set consists of five causal precursors with no set having less than two or more than six (Fig. S3a). Different sets of causal precursors could be due to statistical shortcomings based on the limited (and in case of the subsamples reduced) amount of data but may also reflect actual changes in the relationship between precursors and wheat yields over the studied time period. Note that these changes may arise from physical changes in ocean-atmosphere feedbacks impacting on Moroccan rainfall or from the link between rainfall and wheat yields. The most

robust causal precursors strongly overlap with causal precursor regions found for the full time period. This also holds for different significance thresholds in the partial independence test ($\alpha = 0.05, 0.10, 0.20$) and also when the set of potential precursors is replaced by the robust set depicted in Fig. S1b.

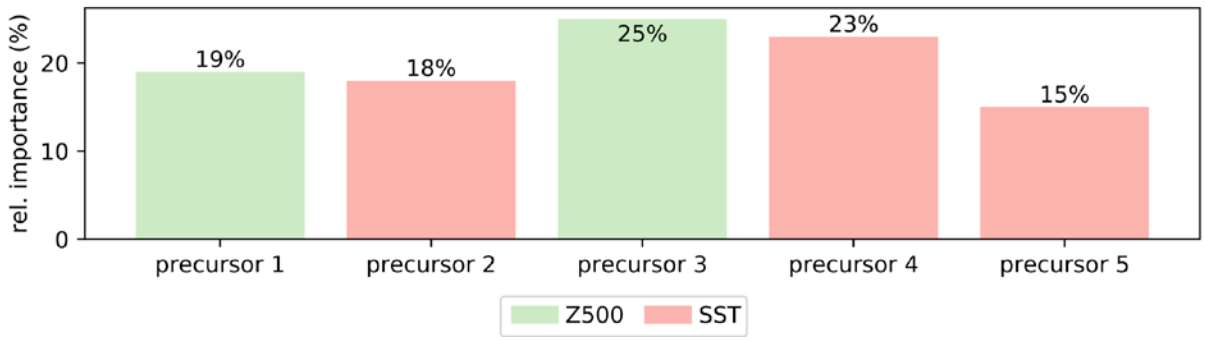

Fig. S 4: Relative importance of causal precursors for overall explained variance. Relative importance is calculated from variance decomposition of the multiple-linear regression model.

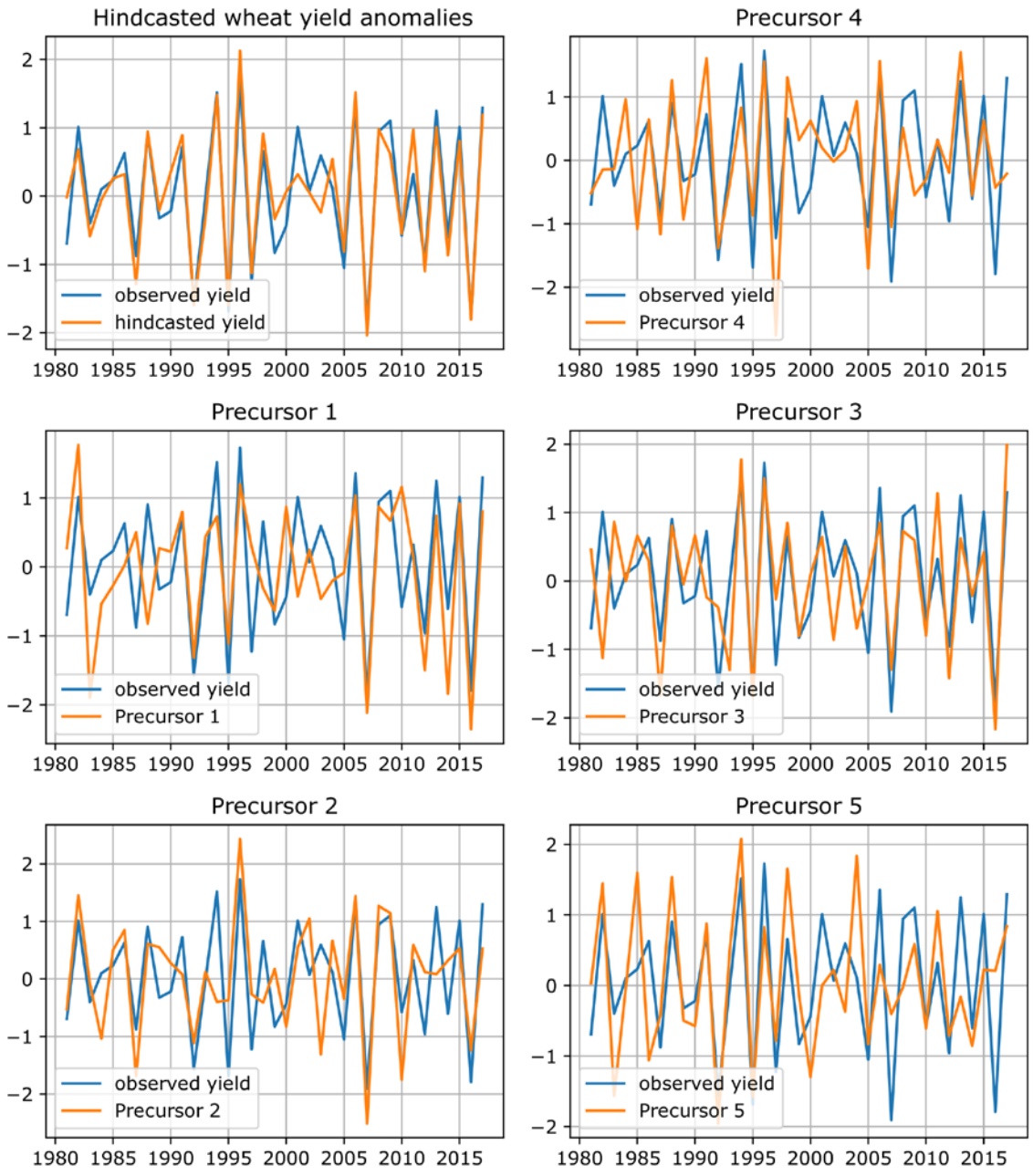

Fig. S 5: Variability of causal precursors. Observed wheat yield anomalies (blue line) are overlaid with time series of hindcasted yield anomalies and causal precursor time series for qualitative comparison.

For better comparison time series are divided by their standard deviation and in case of precursor 1 and 3 inversed in sign to account for their anti-correlation with yield anomalies.

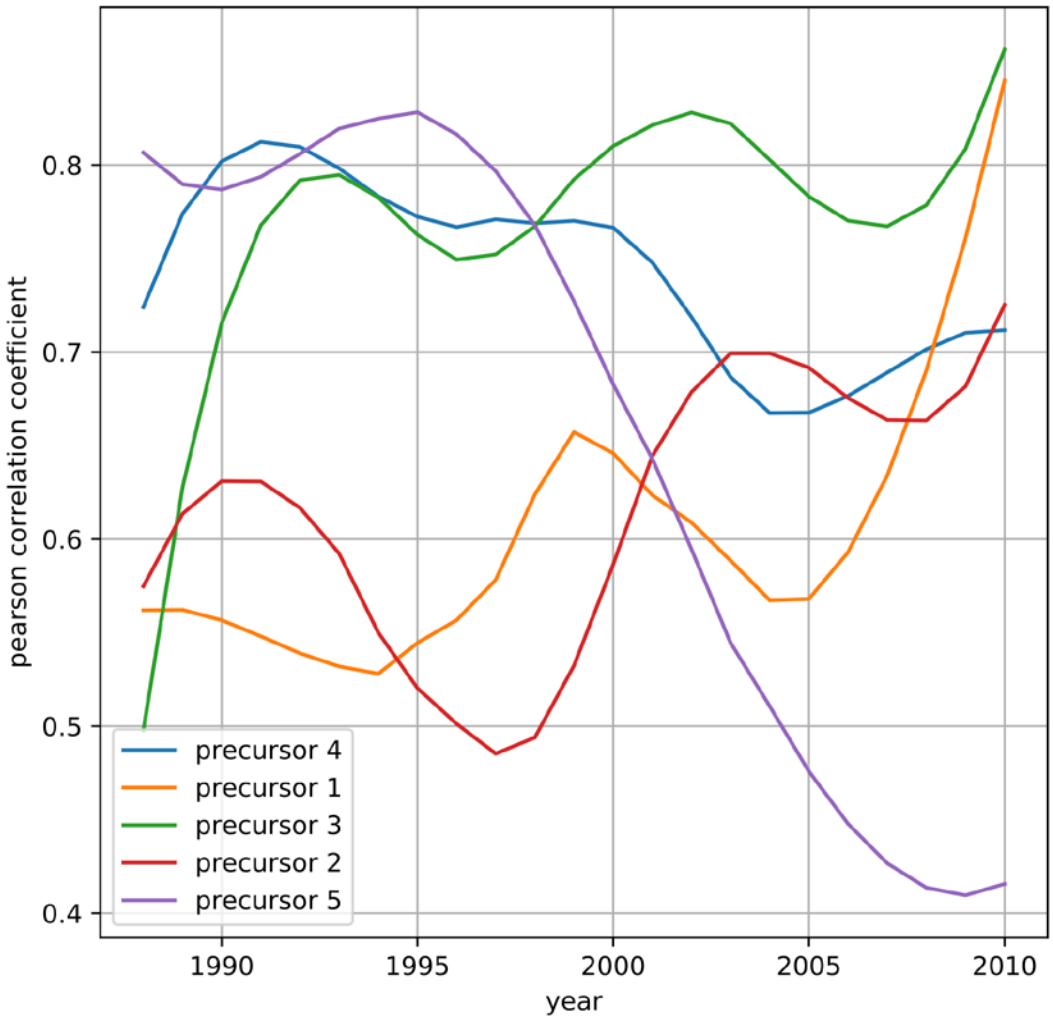

Fig. S 6: Correlation strength between causal precursors and Moroccan wheat yield anomalies. Changes in correlation strength over time are calculated using a rolling window of 15 years.

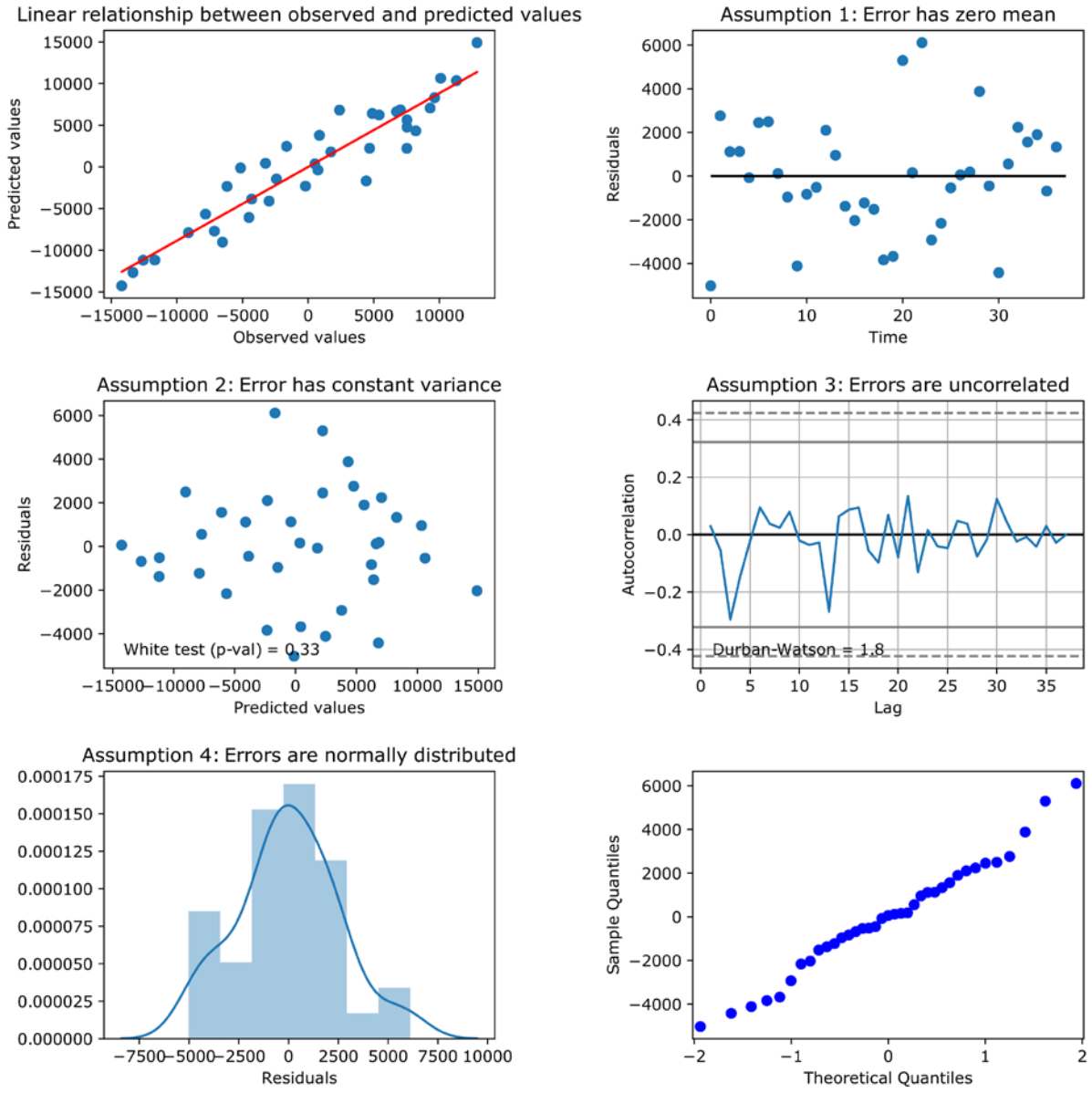

Fig. S 7: Analysis of the residuals from the hindcast model. All requirements of a multiple-linear regression model are fulfilled.

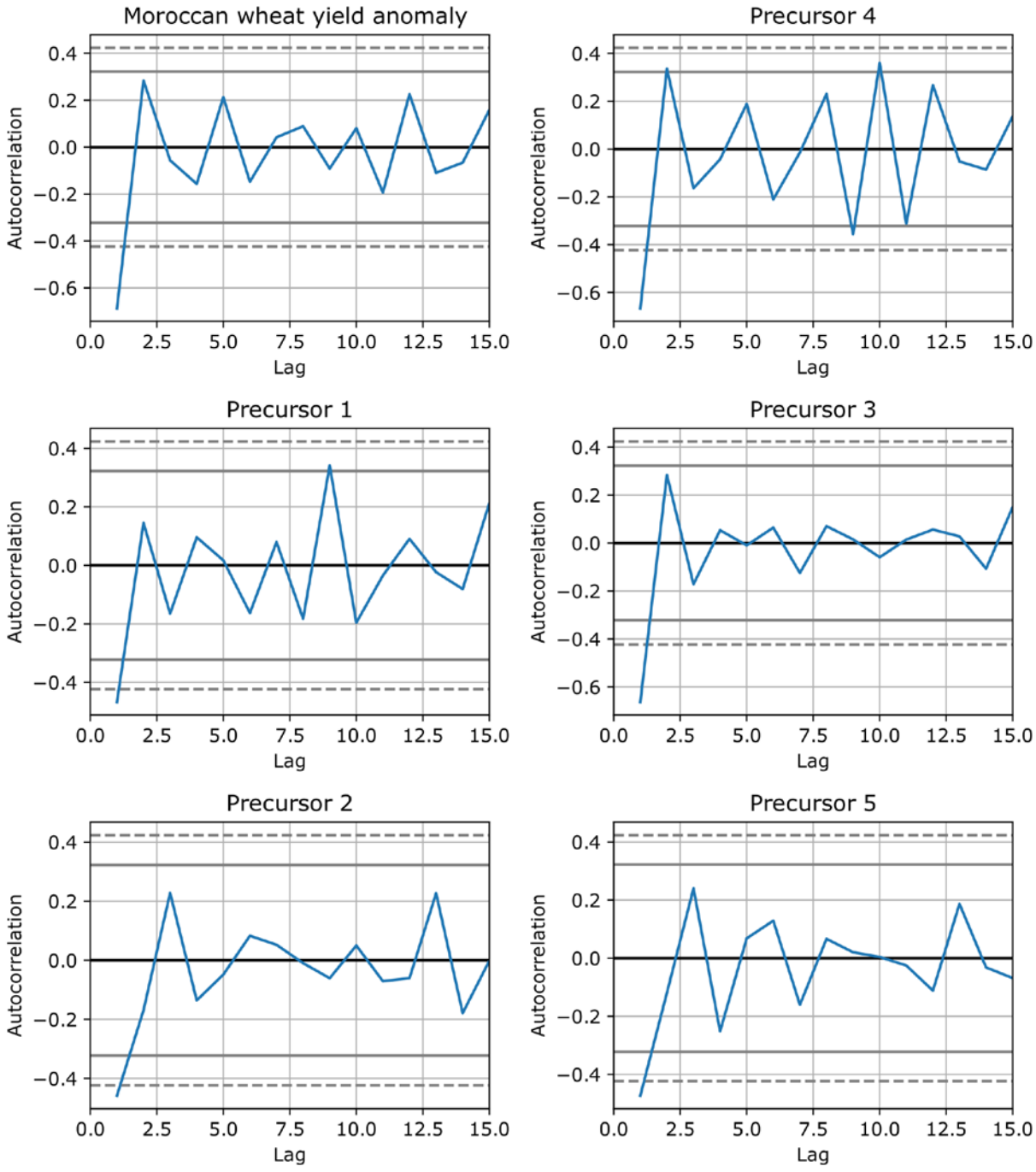

Fig. S 8: Autocorrelation of Moroccan wheat yield anomalies and causal precursors. Autocorrelation is mostly insignificant except for Lag 1, i.e. from one year to the next, where all time series show a significant autocorrelation.

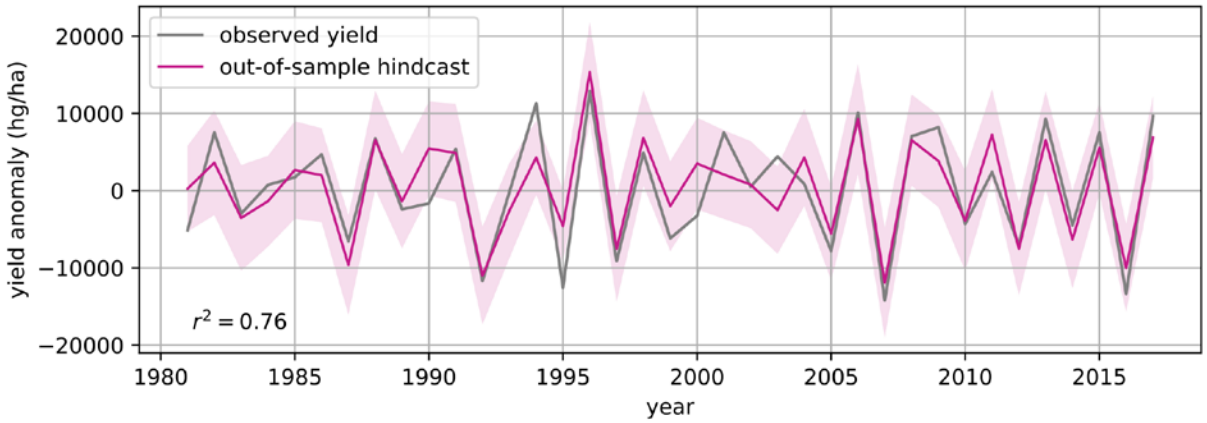

Fig. S 9: Out-of-sample cross validation with prescribed potential precursors. Potential precursors are calculated from the full studied time period (1979-2017) whereas causal precursors and regression parameters are computed using data from the training period only with two years omitted in each training period.

Fig. S 9 shows the result of the cross validation with a significance threshold of $\alpha = 0.10$ in the causal precursor selection (step 2). The value was raised from $\alpha = 0.05$ used for the full time period to compensate for the shorter time series length of the training period.
Accordingly, Fig. 3c of the main manuscript shows results with p-value < 0.03 in step 1 and $\alpha = 0.10$ in step 2 of the model building process since in this case potential and causal precursors are calculated from the training period. Testing different thresholds leads to overall consistent results.

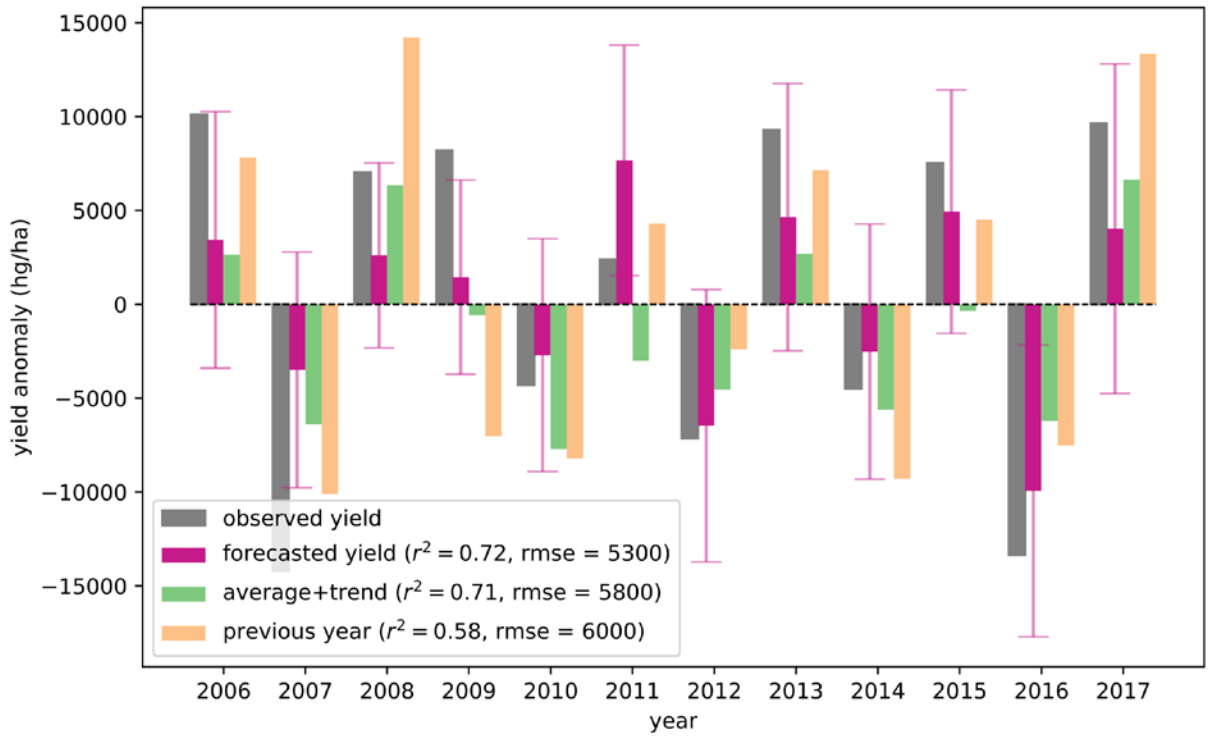

Fig. S 10: Comparison of different forecast models. One-step-ahead forecasted yield anomalies from our model (pink bars) are compared to forecasts from two simple models; one which assumes that the forecasted yield is equal to the average of historical yield totals (not anomalies) plus a linear trend (green bars) and another which sets all forecasts to be the anomaly of the previous year but inversed in sign (orange bars). Observed yield anomalies are shown as grey bars. For each model the explained variance (r^2) and the root mean squared error (rmse) are given. Vertical lines indicate the 95% prediction interval.

The average+trend model explains 71% of the observed variability over the last 12 years but tends to forecast too low yields. The root mean squared error (rmse = 5800 hg/ha) is thus considerably higher than in our model (rmse = 5300 hg/ha) although explained variance is almost the same. The previous-year model has its strength in episodes of alternating yield anomalies but, by default, fails when anomalies deviate from this pattern.

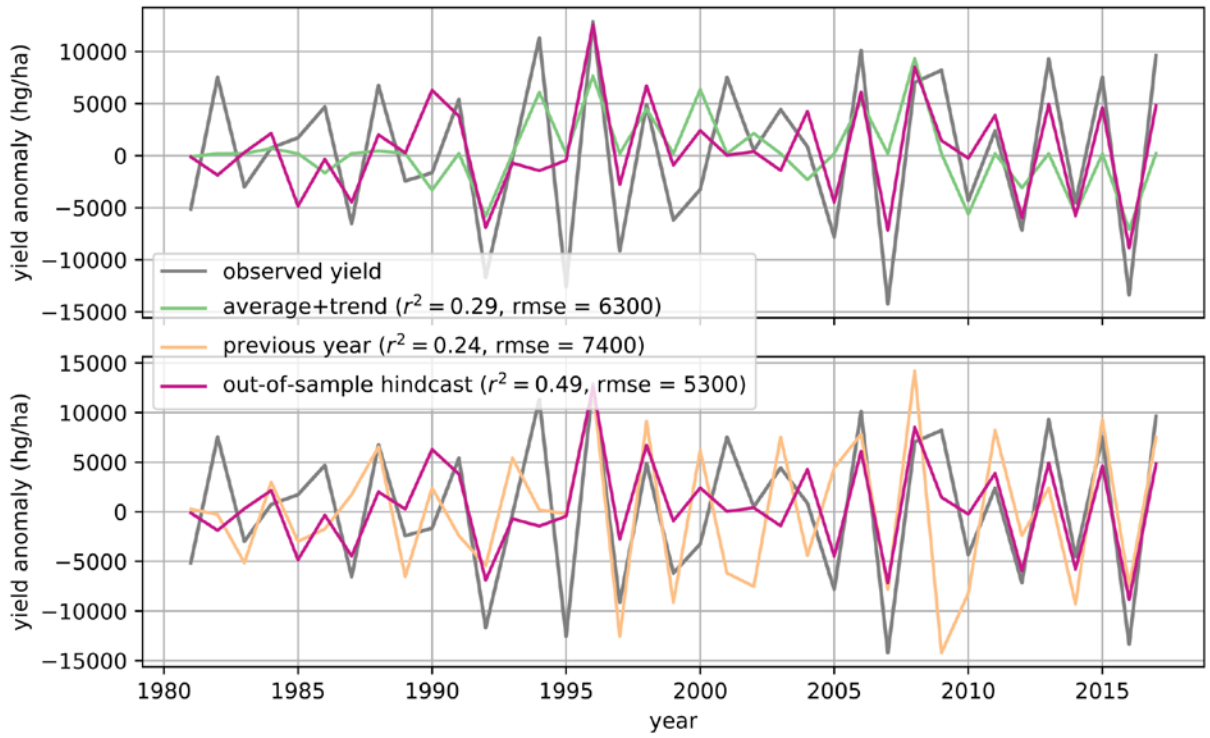

Fig. S 11: Comparison the hindcast skill. Out-of-sample hindcasts from our model (pink line) are compared to out-of-sample hindcasts from the same two simple models described in Fig. S10. Observed yield anomalies are shown as a grey line. Two years are left out in each training period.

Our hindcast model outperforms the two simple models both in terms of explained variance as well as root mean squared error. Both simple hindcast models show drastic reductions in r^2 and rmse in the out-of-sample cross validation compared to their one-step-ahead forecast mode indicating that most of the skill in the forecast mode comes from the strong year-to-year autocorrelation of MWY and causal precursor anomalies.