

Supplementary Materials for

North Atlantic salinity as a predictor of Sahel rainfall

Laifang Li, Raymond W. Schmitt, Caroline C. Ummenhofer, Kristopher B. Karnauskas

Published 6 May 2016, *Sci. Adv.* **2**, e1501588 (2016)

DOI: 10.1126/sciadv.1501588

The PDF file includes:

- Supplementary Materials and Methods
- fig. S1. SSSA-Sahel rainfall relationship independent of precipitation data sets.
- fig. S2. Evaluation of SSSA-Sahel rainfall relationship using salinity.
- fig. S3. Relationship between South Atlantic SSSA and Sahel precipitation.
- fig. S4. Land surface moisture balance in the Sahel.
- fig. S5. Consistency of soil moisture feedback mechanism as shown by NOAA CPC and ESA CCI products.
- fig. S6. North Atlantic SSSA–Sahel precipitation mechanism verified by remote sensing (SMOS) soil moisture data.
- fig. S7. SSSA-rainfall relationship independent of SSTA.
- fig. S8. Random forest regression using eight predictors.
- fig. S9. Linear model selection to predict Sahel monsoon-season precipitation.
- fig. S10. Linear model prediction of Sahel monsoon-season precipitation.
- table S1. Four precipitation data sets used in this study.
- table S2. SSTA-based predictors used to construct the random forest regression model for Sahel monsoon-season precipitation.
- table S3. Pairwise cross correlation between SSS mode, precipitation mode, North Atlantic MAM salinity, and Sahel JJAS precipitation.
- References (66–80)

Supplementary Materials and Methods

Land-surface moisture balance in the Sahel

According to the water balance, the land-surface moisture budget can be formulated as

$$\frac{\partial SM}{\partial t} = P - E + Residual = -\underbrace{\frac{1}{g} \nabla \cdot \int_0^{p_s} q \mathbf{V} dp}_{\text{Moisture Flux Convergence}} + Residual \quad (S1)$$

where SM is soil moisture, P is precipitation, E is evaporation, and the residual term includes infiltration and lateral moisture transport. Further, the difference between precipitation and evaporation ($P-E$) is balanced by the atmospheric moisture flux convergence (MFC) at time scales longer than 10 days (66).

The land surface moisture budget terms (eq. S1) in the Sahel are shown in fig. S4. Averaged over the Sahel region, the moisture storage term is mainly balanced by the convergence of atmospheric moisture flux (MFC) since the two terms show a close one-to-one relationship (fig. S4). Comparatively, the residual term is relatively small (fig. S4). Thus, the land-surface moisture condition is approximately a balance between soil moisture content change and atmospheric MFC

$$\frac{\partial SM}{\partial t} \sim -\underbrace{\frac{1}{g} \nabla \cdot \int_0^{p_s} q \mathbf{V} dp}_{\text{Moisture Flux Convergence}} \quad (S2)$$

Soil moisture signal as diagnosed using ESA CCI data set and a case study using SMOS satellite remote sensing data set

We acknowledge that the long-term NOAA CPC soil moisture data used in this study is model-based, which will introduce uncertainties in evaluating the key mechanisms proposed in this study. Here, we further verify the contribution of soil moisture using the ESA CCI data set (60). The ESA CCI data set merges and homogenizes six passive microwave and active microwave soil moisture products. The data set covers the period from 1980 to present, i.e. the second half of the time series analyzed in this study.

We applied composite analysis to the ESA CCI soil moisture data set (60). The high (low) SSS cases are the years when North Atlantic SSSA is ranked at the top (bottom) decile of the 1980-2009 period. According to the composite soil moisture, the ESA CCI data set generally reproduces the soil moisture signal diagnosed using the model-based NOAA CPC data (fig. S5). Specifically, both data sets show increased soil moisture content in the spring season over the African Sahel region (fig. S5A and S5C) in the high salinity years. In addition, the soil moisture persists and further increases into the monsoon season (fig. S5B and S5D). It is noteworthy that the further increases of soil moisture occurs mainly over the grid cells that are already wet in the spring season, indicating a potentially important role of localized feedback in amplifying the initial soil moisture anomalies (fig. S5). The consistency between the two data sets supports our

conclusion that soil moisture is the likely mechanism that extends the springtime SSSA signal into the monsoon season.

We further verify the results involving soil-moisture feedback by a case study based on soil moisture data from satellite remote sensing (SMOS data). The SMOS data covers only the recent five years (2010-2014), and we selected the highest SSS (2012) and lowest SSS (2013) years as sample cases according to the 2010-2014 MAM North Atlantic SSS time series (fig. S6A). The differences in precipitation and soil moisture between the two sample cases are shown in fig. S6B-D. Consistent with the long-term relationship between MAM North Atlantic SSS and Sahel monsoon-season precipitation, the 2012 case shows about 0.7 mm day^{-1} more precipitation than in the 2013 case (fig. S6B). Furthermore, the differences in soil moisture content qualitatively agree with the analysis based on NOAA CPC and ESA CCI data. In the spring of 2012, soil moisture content is higher over the Sahel, especially over the eastern portion of the domain and the coastal regions (fig. S6C). Over these regions, the soil moisture content further increases in the monsoon season.

Generally, the soil moisture anomalies diagnosed in the case study are weaker than those from the NOAA CPC and ESA CCI. This weak signal is expected as the SSSA anomalies in neither the 2012 nor the 2013 spring exceed one standard deviation. As the smaller SSSA corresponds to less moisture export from the subtropical North Atlantic, the soil moisture in the Sahel shows less deviation due to weaker moisture input. Consequently, the magnitude of soil moisture feedback tends to be weakened leading to lower soil moisture anomalies and precipitation anomalies in the composite maps (figs. S5 and S6).

In conclusion, the case study using the short available record of the SMOS soil moisture data supports the conclusions drawn from the longer NOAA CPC and ESA CCI data sets, which suggests that soil moisture feedback is a likely mechanism that physically links springtime SSS in the North Atlantic with monsoon-season precipitation in the Sahel.

Linkage between South Atlantic SSSA and Sahel precipitation

The leading mode of the SVD analysis indicates that the springtime SSSA over the South Atlantic Ocean also significantly correlates with Sahel precipitation (Fig. 1). The results from the SVD analysis are further verified by the cross correlation between pre-monsoon season South Atlantic SSSA and African precipitation (fig. S3A-B). The correlation coefficients are statistically significant over the Sahel region, although lower than the correlation with North Atlantic SSSA (Fig. 2B and fig. S3B).

This statistically significant correlation between South Atlantic SSSA and Sahel precipitation, however, is not physically meaningful as indicated by the moisture flux from the South Atlantic and soil moisture content over the Sahel. In the spring, with high SSS, the anomalous moisture fluxes originating from the South Atlantic are mainly directed poleward rather than towards the Sahel (fig. S3C). As a result, the variation of South Atlantic SSSA and the associated moisture flux cannot physically influence the moisture budget in the Sahel.

In addition, the positive feedback between local soil moisture and atmospheric MFC that explains the relationship between North Atlantic SSSA and Sahel precipitation (Fig. 3) is absent

in the South Atlantic SSS composites (fig. S3E-F). According to fig. S3E-F, the MFC and soil moisture content anomalies do not differ significantly between high and low SSS composites in the South Atlantic. For example, the springtime soil moisture content is below normal for both high and low SSS composites (fig. S3F). The MFC during low SSS years is generally negative (fig. S3E). However, the large uncertainties in MFC during high SSS years make the differences between the two composites statistically insignificant.

North Atlantic SSSA is independent of SSTA modes that influence Sahel precipitation

Previous studies have shown that Sahel precipitation is sensitive to global sea surface temperature anomaly (SSTA) patterns (42-46), whose influence can be further amplified by local land-atmosphere coupling (34). Thus, the issue of whether North Atlantic SSSA and its relationship with Sahel precipitation add unique value independent of typical SSTA modes must be explored.

Using NOAA Extended Reconstruction SST (ERSST) (67), fig. S7 shows the global SSTA pattern associated with Sahel precipitation and SSSA. Since the linear trend in SSTA has been removed, the correlation coefficients shown in fig. S7 reflect the relationship between precipitation/SSSA and SSTA at interannual-to-interdecadal scale.

SSTA patterns associated with Sahel precipitation indicate that pre-monsoon and monsoon-season global SSTA are significant mainly over the eastern tropical Pacific and *during* the monsoon season (fig. S7C). The SSTA presents a typical La Niña pattern, consistent with previous observations and numerical simulations that La Niña events tend to result in wet monsoon seasons in the Sahel (44, 45). This La Niña-like SSTA signal is absent in the SSTA regression using North Atlantic SSSA. Globally, the North Atlantic springtime SSSA does not show significant correlations with SSTA (fig. S7B and S7D). The insignificant correlation between North Atlantic SSSA and La Niña SSTA is consistent with previous studies showing limited loading of ENSO SSTA on SSSA outside the tropical Pacific and the Indian Ocean (68, 69). Furthermore, SSSA in the North Atlantic is controlled by processes different from those generating local SSTA (70). Thus, these results suggest that the relationship between springtime SSSA and monsoon season Sahel precipitation is independent of SSTA's influence. More importantly, the results suggest that North Atlantic SSSA can provide value for prediction of Sahel monsoon-season precipitation, in addition to the previously recognized ENSO signal.

Multiple linear regression model

Besides the random forest algorithm, we also constructed linear regression models to assess the predictability of Sahel precipitation using North Atlantic SSSA as a predictor. A multiple linear regression aims to model the relationship between two or more predictors and one response variable by fitting a linear equation to observed data. In this study, we select the predictors for the linear regression model using a subset selection method. Specifically, among the eight predictors (SSSA and seven SST-based predictors in table S2), we select a subset with k ($k \leq 8$) predictors that are related to the Sahel precipitation. First, for $k = 1, 2, \dots, 8$, we fitted all $\begin{pmatrix} 8 \\ k \end{pmatrix}$

models that contain exactly k predictors. We then selected among the $\binom{8}{k}$ models the one that has the largest R^2 (i.e. the least squared errors) and denoted it as M_k . Second, we selected a best model among M_k ($k=1, 2, \dots, 8$), according to both indirectly estimated test errors (adjusted R^2 and Mallows's Cp value (71, 72)) and cross-validated prediction errors.

Figure S9A shows the variables selected in each M_k . The results suggest that North Atlantic SSSA should be included in every M_k in order to maximize the explained variance of Sahel precipitation (fig. S9A). The highest importance of SSSA in multiple linear models is consistent with that in the random forest algorithm (Fig. 4A). Thus, both prediction models, based on different assumptions, agree that the newly identified SSSA predictor can improve rainfall prediction in the Sahel.

According to fig. S9A, the second most important predictor is ENSO, which is selected in models with two or more predictors. In contrast, the Pacific Decadal Oscillation (PDO; second most important in the random forest algorithm) is the last predictor to include in linear models. This is likely due to the nonlinear relationship between PDO and Sahel precipitation identified based on the 60-year samples (results not shown here). The random forest algorithm is capable to capture this nonlinear relationship and thus leads to higher rainfall predictability by including PDO.

The best multiple linear regression model is first evaluated using indirectly estimated test errors which take into account the model complexity. The statistics include the adjusted R^2 and

Mallows's Cp (71, 72). The adjusted R^2 is formulated as $1 - \frac{SS_{res}}{SS_{tot}} \frac{N-k-1}{N-1}$, where SS_{res} is the

unexplained variance by the predictions, and SS_{tot} is the total variance of the response. N is sample size and k is the number of predictors selected in the model. A higher adjusted R^2 indicates a model with a smaller test error. In this study, the Mallows's Cp value is also

calculated: $\frac{1}{N}(SS_{res} + 2k\hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the estimated variance of the error associated with

each response. Models with lower Cp value have less prediction errors and thus should be selected. Fig. S9B shows the Adjusted R^2 and Mallows's Cp versus the number of variables included in the multiple linear regression models. According to the figure, the Adjusted R^2 reaches maximum and Mallows's Cp value reaches a minimum for models with two variables (fig. S9B).

The model choice using the indirect criteria can be verified using cross-validation. Specifically, we divided the data sets into 10 folds, using 9 folds to construct linear models and the remaining fold for validation. The same process was repeated 1000 times and the test errors were calculated for each iteration (fig. S9C). Consistent with the adjusted R^2 and Mallows's Cp value shown in Fig. S9B, the model with two variables shows the least test errors (fig. S9C).

Overall, both the indirect and direct estimates of test errors suggest that the linear regression model with North Atlantic SSSA and ENSO (fig. S10) should be used to predict Sahel precipitation. According to our analysis, the multiple linear regression model is formulated as

$$y=5.46SSSA-0.14ENSO+\varepsilon \quad (S3)$$

where y is the precipitation, and ε is the residual of the prediction model which is normally distributed. eq. (S3) suggests that a change of 1 unit (PSS-78) SSS is expected to generate 5.46 mm day⁻¹ anomalies in precipitation. The constructed linear model shows certain skill in predicting Sahel precipitation (fig. S10A), since the R^2 between prediction and observation is 0.40 ($p < 0.001$). More importantly, the majority of the prediction skill results from the incorporation of North Atlantic SSSA (fig. S10B). Specifically, SSSA explains 34% of the variance, while ENSO can only explain 6% (fig. S10B-C). Furthermore, the prediction using springtime SSSA lies close to the $y=x$ line (fig. S10B). In contrast, the predictions based on the ENSO index largely underestimate the variability of Sahel precipitation (fig. S10C).

Supplementary Figures

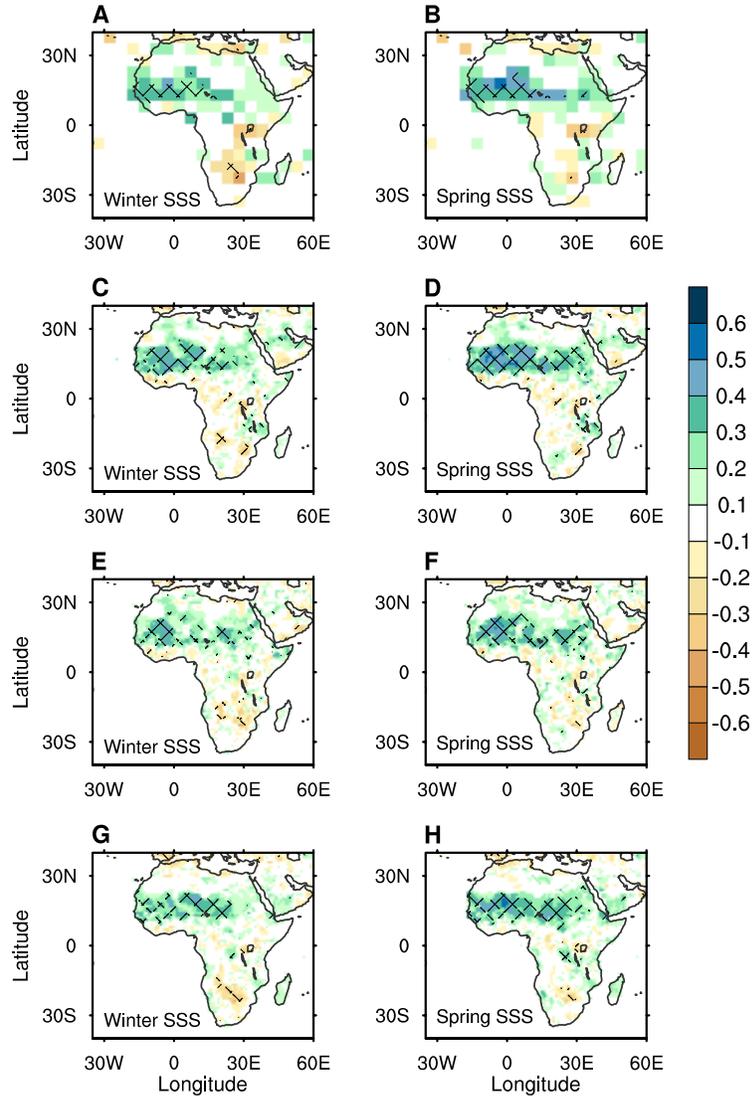


fig. S1. SSSA-Sahel rainfall relationship independent of precipitation data sets. Correlation between June-September African precipitation and wintertime (January-March) SSS (left panels) and springtime (March-May) SSS (right column) over the North Atlantic as shown by four different precipitation data sets: (A) and (B) are for the GHCN data set, (C) and (D) are for the U. Delaware data set, (E) and (F) are for the GPCC, and (G) and (H) are for the CRU. The grid cells with correlation coefficients significant at the $\alpha=0.05$ level are hatched.

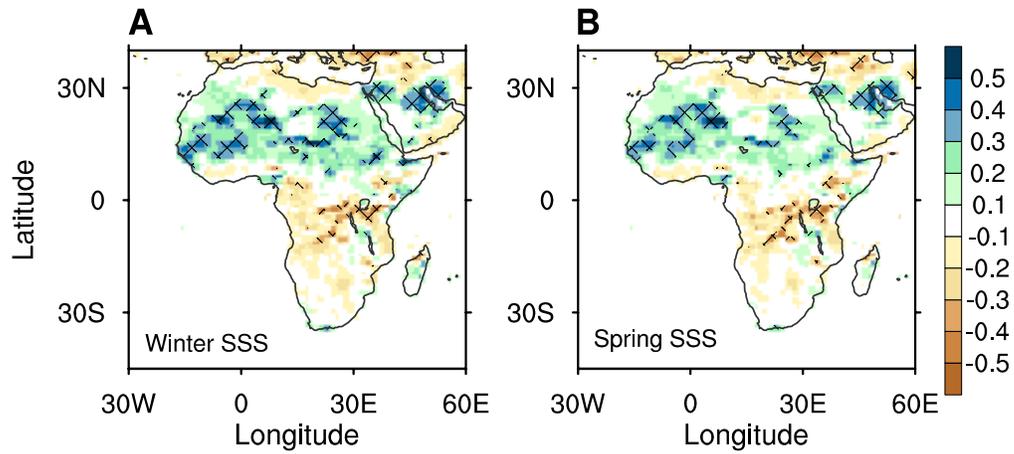


fig. S2. Evaluation of SSSA-precipitation relationship using SODA data. Correlation between June-September African precipitation and wintertime (A) and springtime (B) North Atlantic SSS using SODA data. Grid cells with correlation coefficients significant at $\alpha=0.05$ level are hatched.

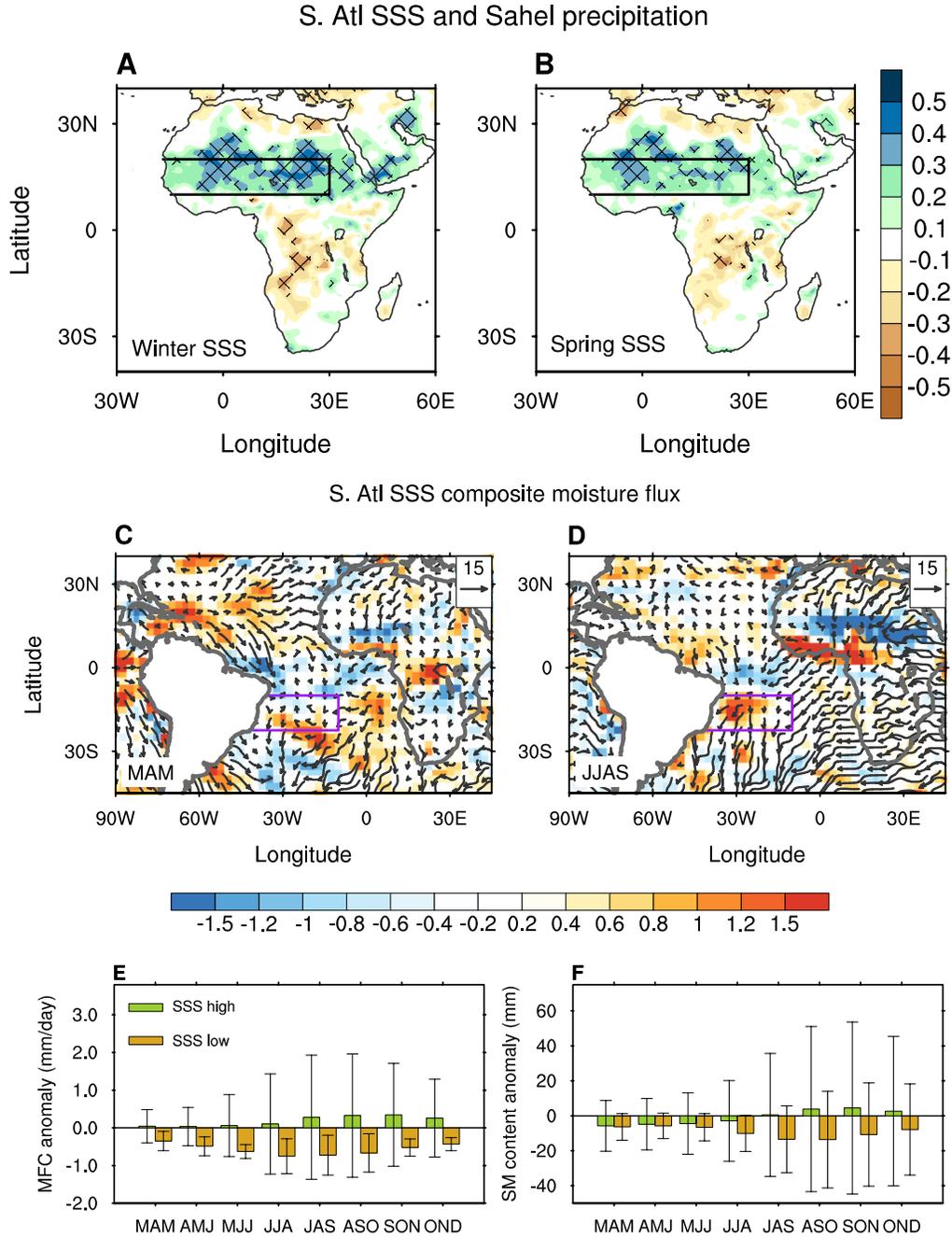


fig. S3. Relationship between South Atlantic SSSA and Sahel precipitation. (A) and (B) are the correlation between June-September African precipitation and wintertime (A) and springtime (B) South Atlantic SSSA. The boxes in (A) and (B) denote the Sahel region. The grid cells with correlation coefficients significant at the $\alpha=0.05$ level are hatched. (C)-(F) are the same as Fig. 3, but for composites on South Atlantic SSS. The boxes in (C) and (D) are the domain used to calculate South Atlantic SSS anomalies. The error bars in (E) and (F) denote the upper and lower bound of soil moisture or moisture flux convergence anomaly defined by one standard deviation.

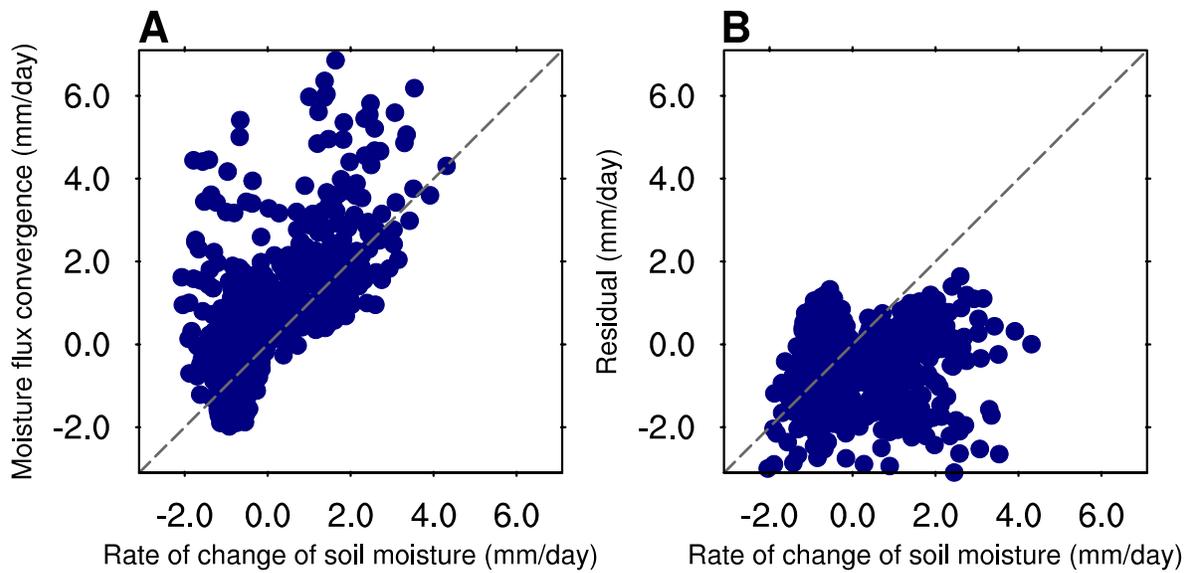


fig. S4. Land surface moisture balance in the Sahel. Scatter plots of soil moisture rate of change versus moisture flux convergence (A), and residual terms (B). The dashed black lines are the $y=x$ line.

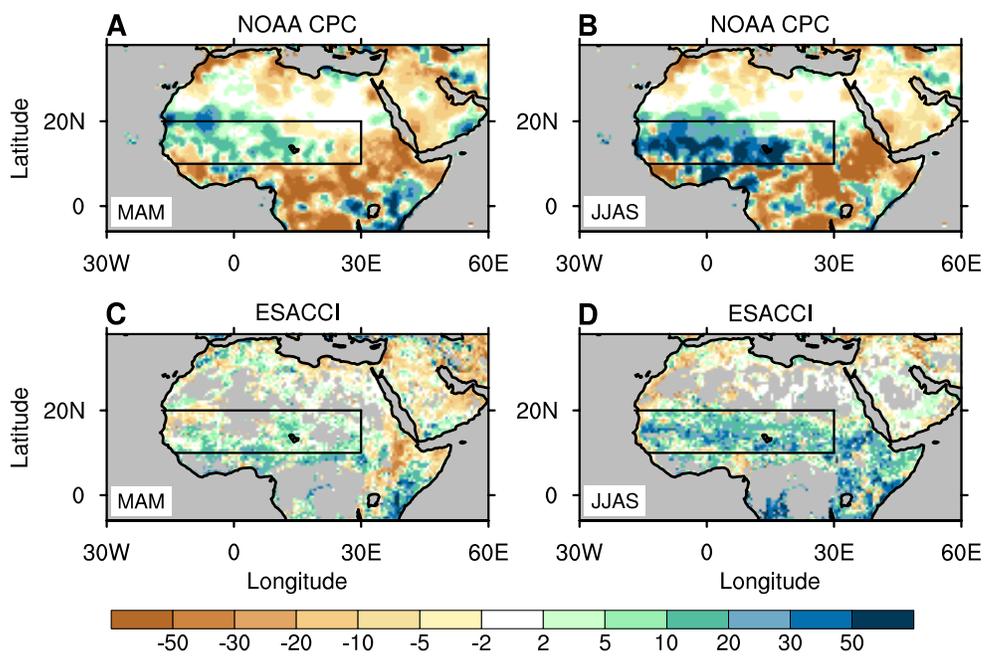


fig. S5. Consistency of soil moisture feedback mechanism as shown by NOAA CPC and ESA CCI products. (A) and (B) are the MAM and JJAS soil moisture anomaly composite upon the 1950-2009 North Atlantic SSSA using the NOAA CPC data (shaded; unit: $10^{-3} \times \text{m}^3 \text{m}^{-3}$). (C)-(D) are the same as (A)-(B), but are using the ESA CCI soil moisture products during the 1980-2009 period.

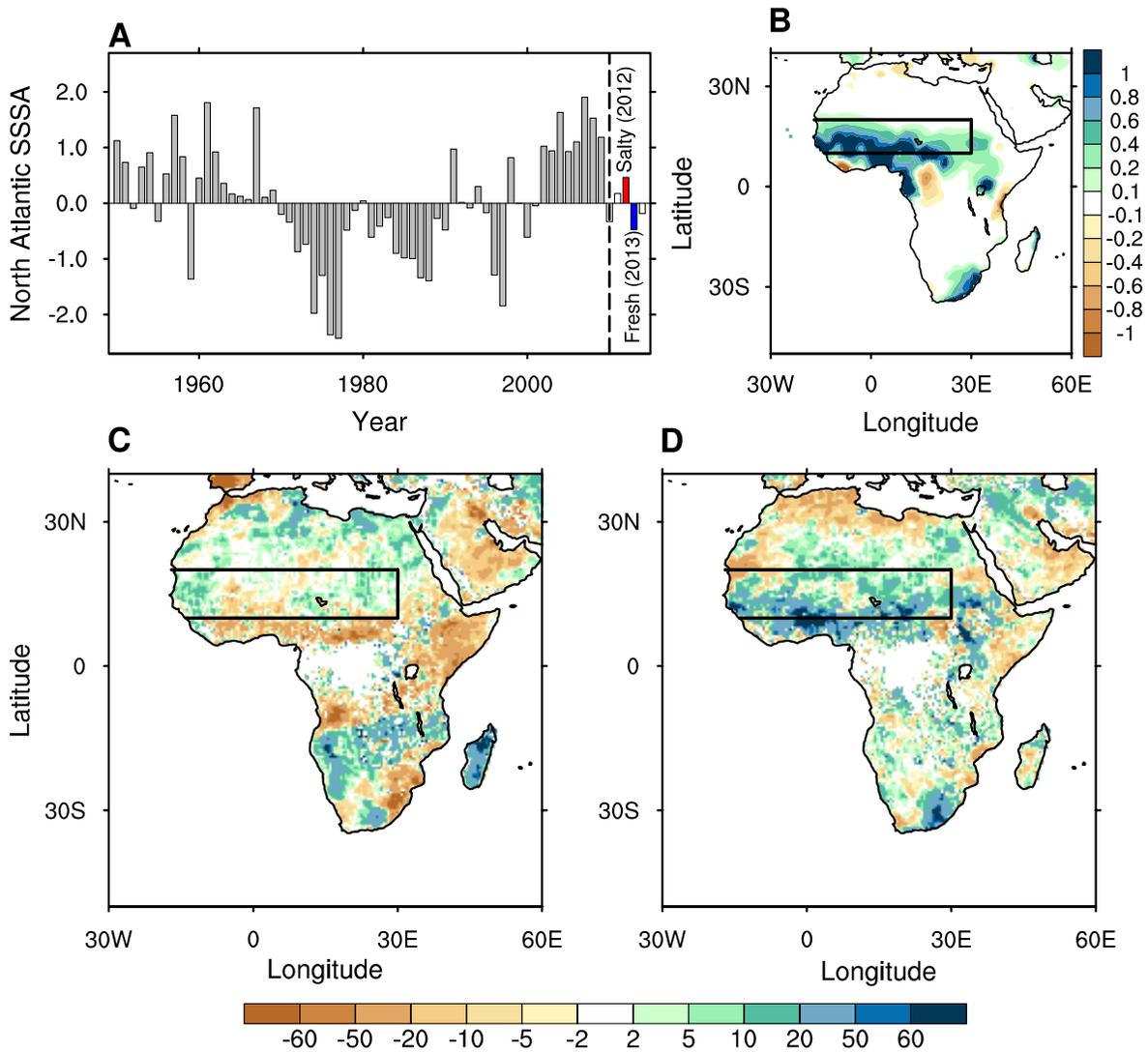


fig. S6. North Atlantic SSSA–Sahel precipitation mechanism verified by remote sensing (SMOS) soil moisture data. (A) Normalized MAM north Atlantic SSSA time series during 1950-2014. The dashed line is the year 2010 after which SMOS soil moisture becomes available. The red (blue) bar denotes the year 2012 (2013), the highest (lowest) salinity case after 2010. (B), (C), and (D) show the differences between 2012 and 2013 for the African monsoon-season precipitation (B; shaded, unit: mm day^{-1}), and for (C) MAM and (D) JJAS soil moisture content (shaded, unit: $10^{-3} \times \text{m}^3 \text{m}^{-3}$) based on SMOS data. The black boxes in (B) to (D) delineate the African Sahel region.

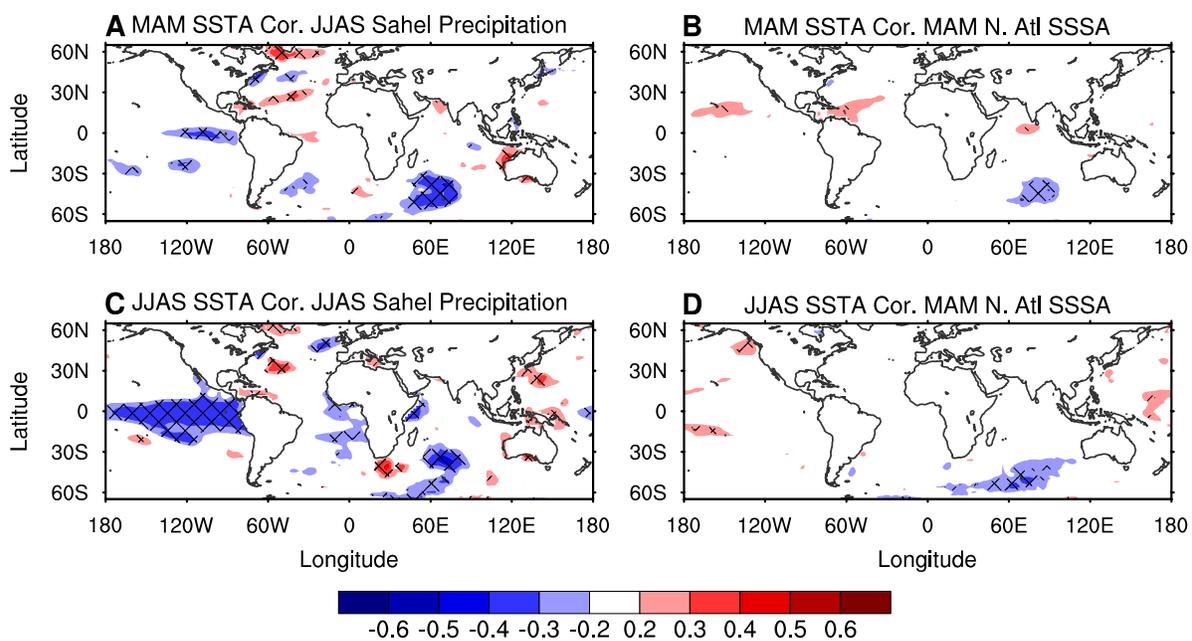


fig. S7. SSSA-rainfall relationship independent of SSTA. Correlation between global sea surface temperature anomaly (shaded) with June-September Sahel precipitation (A and C), and with springtime North Atlantic SSSA (B and D). Areas with correlation coefficients significant at the $\alpha=0.05$ level are hatched.

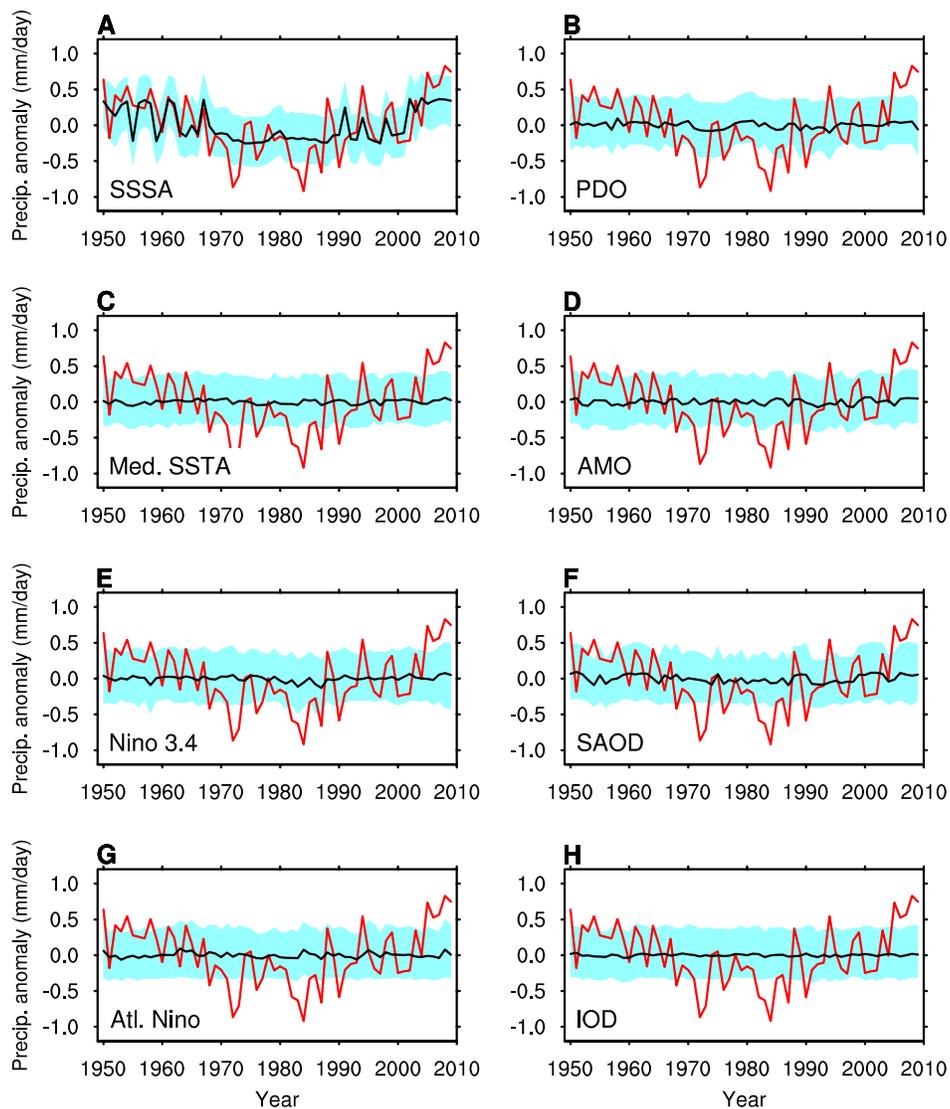


fig. S8. Random forest regression using eight predictors. The red curves are the observed precipitation anomaly over the Sahel. Sahel Precipitation anomaly (black curves, unit: mm day^{-1}) predicted by the random forest algorithm using springtime (A) North Atlantic SSSA, (B) PDO, (C) Mediterranean SSTA, (D) AMO, (E) Nino 3.4, (F) SAOD, (G) Atlantic Nino and (H) IOD. The blue shading is the 95% confidence interval derived from 1000 trails of the regression.

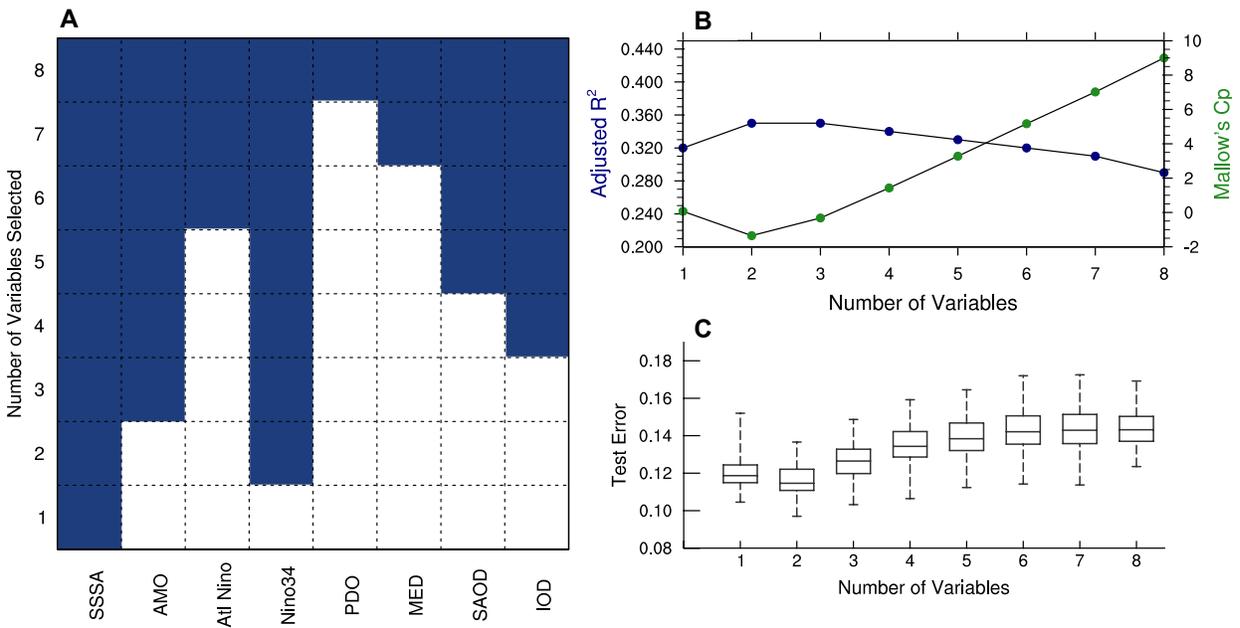


fig. S9. Linear model selection to predict the Sahel monsoon-season precipitation. (A) Variables selected in the multiple linear regression model with k variables. The blue boxes denote the selected variable, while the white boxes are the variables that are not selected. (B) Adjusted R^2 (blue dots) and Mallows's C_p value (green dots) calculated for the multiple linear regression model with k variables. (C) is the Box-and-whisker plot of test errors in the constructed linear regression model with k ($k=1, 2, \dots, 8$) variables. Each box-and-whisker plot represents the minimum, 25%, median, 75%, and maximum of the test error derived from the 1000 iterations of validations.

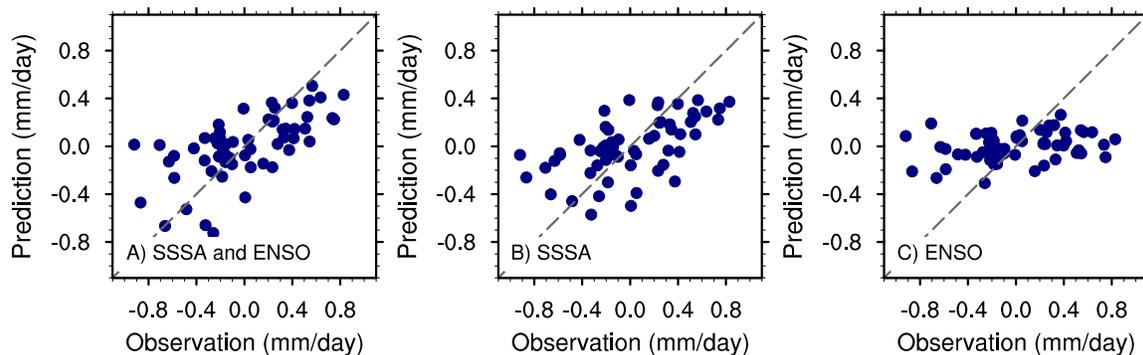


fig. S10. Linear model prediction of Sahel monsoon-season precipitation. Multi-variate linear model prediction of Sahel precipitation using the North Atlantic springtime SSSA and ENSO as predictors: (A) prediction considering both SSSA and ENSO; (B) and (C) are predictions using SSSA and ENSO, respectively. The dashed gray lines in (A)-(C) denote $y=x$.

Supplementary Tables

table S1. Four precipitation data sets used in this study

| Data sets | Spatial Resolution | Temporal Coverage | References |
|-----------|--------------------|-------------------|----------------|
| GHCN | 5°×5° | 1900/01 - 2015 | Refs. (53, 54) |
| UDEL | 0.5°×0.5° | 1900/01 - 2010/12 | Ref. (55) |
| GPCC | 0.5°×0.5° | 1901/01 - 2015 | Ref. (56) |
| CRU | 0.5°×0.5° | 1901/01-2013 | Ref. (57) |

table S2. SSTA-based predictors used to construct the random forest regression model for Sahel monsoon-season precipitation [AMO: Atlantic Multidecadal Oscillation; IOD: Indian Ocean Dipole; Med SSTA: Mediterranean SSTA; PDO: Pacific Decadal Oscillation; SAOD: South Atlantic Ocean Dipole]

| Predictor | Impact | Mechanism | References |
|---------------|--|--|------------|
| AMO | Increases in Sahel precipitation during the positive phase of AMO | The AMO leads to differential heating between Northern and Southern Hemisphere, results in meridional movement of the ITCZ | Ref. (73) |
| Atlantic Niño | Positive Atlantic Niño results in drought in the Sahel | Weakening of the sea-level pressure gradient between ocean and land, which results in weaker ITCZ movement | Ref. (74) |
| IOD | Positive IOD phase results in reduction of Sahel precipitation | Positive IOD results in an anomalous Walker-like zonal circulation with its subsidence branch located in the Sahel | Ref. (75) |
| Med SSTA | Warmer Mediterranean tends to be associated with above normal precipitation in Sahel | High SSTA leads to stronger evaporation in Mediterranean which increases moisture flux into the Sahel by the northeasterlies | Ref. (76) |
| Niño 3.4 | Positive phase (El Niño) is associated with reduced precipitation in Sahel | ENSO influence Sahel precipitation through the propagation of Kelvin waves and the displacement of tropical convection | Ref. (77) |
| PDO | Warm phase results in reduction of precipitation in Sahel | Tropospheric warming associated with the warm phase stabilizes the atmospheric column in the Sahel | Ref. (78) |
| SAOD | Positive phase results in reduction of precipitation in Sahel | The SSTA dipole change in low-level mass convergence and the location of convective activities | Ref. (79) |

table S3. Pairwise cross correlation between SSS mode, precipitation mode, North Atlantic MAM salinity, and Sahel JJAS precipitation [The coefficients in the parentheses are calculated using the time series without quadratic low-frequency variability*.]

| | SSS Mode-1 | Precip. Mode 1 | N. Atl. SSS | Sahel Precip. |
|----------------|-------------|----------------|-------------|---------------|
| SSS Mode-1 | 1.0 (1.0) | 0.84 (0.60) | 0.88 (0.76) | 0.71 (0.55) |
| Precip. Mode-1 | 0.84 (0.60) | 1.0 (1.0) | 0.59 (0.47) | 0.95 (0.93) |
| N. Atl. SSS | 0.88 (0.76) | 0.59 (0.47) | 1.0 (1.0) | 0.58 (0.39) |
| Sahel Precip. | 0.71 (0.55) | 0.95 (0.93) | 0.58 (0.39) | 1.0 (1.0) |

* Ensemble empirical mode decomposition (EEMD) method (80) is used to extract the quadratic low-frequency variability from each of the four time series.