Predictability of Seasonal Precipitation Using Joint Probabilities

M. Tugrul Yilmaz
George Mason University, Fairfax, VA

Timothy DelSole
George Mason University, Fairfax, VA and Center for Ocean-Land-Atmosphere Studies, Calverton, MD

Corresponding author address: M. Tugrul Yilmaz, 4400 University Dr., Fairfax, VA 22030
Email: myilmaz1@gmu.edu

Abstract

This paper tests whether seasonal mean precipitation is predictable using a new method that estimates and analyzes joint probabilities. The new estimation method is to partition the globe into boxes, pool all data within the box to estimate a single joint probability of precipitation for two consecutive seasons, and then apply the resulting joint probability to individual pixels in the box. Pooling data in this way allows joint probabilities to be estimated in relatively small sample sizes, but assumes that the transition probabilities of pixels in a box are homogeneous and stationary. Joint probabilities are estimated from the Global Precipitation Climatology Project data set in 21 land boxes and 5 ocean boxes during the period 1979-2008. The state of precipitation is specified by dry, wet, or normal terciles of the local climatological distribution. Predictability is quantified by mutual information, which is a fundamental measure of predictability that allows for nonlinear dependencies, and tested using bootstrap methods. Predictability was verified by constructing probabilistic and quantitative forecasts directly from the transition probabilities and showing that they have superior cross-validated skill than forecasts based on climatology, persistence, or random selection. Spring was found to be the most predictable season whereas summer was the least predictable season. Analysis of joint probabilities reveals that though the probabilities are close to climatology, the predictability of precipitation arises from a slight tendency of the state to persist from one season to the next, or if a transition occurs then it is more often from one extreme to normal than from one extreme to the other.

1. Introduction

A skillful precipitation prediction is desirable for many industrial and scientific fields i.e. agriculture, flood risk analysis, drought, fishery, etc. (Glantz 1992, chapter 15; Hansen et al., 1998). Perhaps the first serious assessment of the predictability of precipitation is that of Namias (1952, 1960), who used contingency tables estimated from historical records to show that precipitation anomalies tend to persist on seasonal time scales. Today, this persistence is assumed to be associated with slowly varying boundary conditions, such as sea surface temperatures or soil moisture (Kirtman and Schopf, 1998; Koster et al., 2000; Koster et al., 2004). Although the land surface does not have the same heat capacity as ocean, anomalies in soil moisture still can persist three months (Vinnikov, 1996). Recently, evidence has emerged that precipitation is most predictable in certain "hot spots" characterized by enhanced land-atmosphere coupling (Koster et al., 2004; Wang et al., 2007; Dirmeyer et al., 2009).

In general, investigation of precipitation predictability is challenging because precipitation rate is strongly non-Gaussian, discontinuous, and only weakly predictable (at least compared to temperature). Katz (1977, 1983) proposed a model of precipitation in which the occurrence or non-occurrence of precipitation is modeled by a two-state first order Markov chain, and the intensity of precipitation is modeled by a random process. Variants of this model have been used in a variety of contexts (e.g., Wilks 2002). These and other models could be used to draw inferences about predictability, but then the results would be model dependent, or rely on the appropriateness of the model.

In this paper, we investigate the predictability of seasonal precipitation directly from joint probabilities without assuming anything about the physics or distribution of precipitation. However, a major difference from previous studies is that we partition the globe into pre-defined boxes containing many pixels, and then pool the data over all pixels within a box to estimate a single joint probability. The resulting joint probability for the box is then assumed to apply to individual pixels. By pooling data over many pixels, we are able to estimate joint probabilities with relatively small sample sizes. This approach assumes that the system is stationary, in the sense that relations (feedback mechanisms) that exist in the past also hold into the future. However, in a changing climate system the historical transition probability may loose its relevance (Van den Dool 2007, p. 123). This approach also assumes that the statistics are homogeneous within the box. The joint probability in question is the probability that precipitation is in one of three terciles, called dry, normal, and wet, for two consecutive seasons. Thus, the joint probability can be used to predict whether next season's precipitation will be wet, dry, or normal, given the precipitation of the present season.

The predictability of precipitation is assessed in two different ways. First, we evaluate the mutual information of the joint probabilities. Mutual information is a fundamental measure of the degree of dependence between two random variables. If the precipitation between two seasons is independent, then mutual information vanishes. The statistical significance of mutual information in each study area is estimated using bootstrap methods. Second, we compare cross-validated forecasts constructed from the joint probability (or more precisely, the conditional distribution) to three benchmark forecasts, namely forecasts based on climatology, persistence, and random selection.

This paper is organized as follows. The dataset is introduced in section 2, methodology is outlined in section 3, an overview of the basic concepts of the climatology- and persistence-based forecasts is offered in section 4, results are presented in section 5, and summary and discussions are given in section 6.

2. Data set

The data set used in this study is the Global Precipitation Climatology Project (GPCP) precipitation product (Adler et al, 2003). This data set is a global monthly merged product from gauge- and satellite-based observations over 2.5° grids for the period 1979-2008. The monthly precipitation has been aggregated into boreal seasonal rainfall as: winter (December-February); spring (March-May); summer (June-August); and autumn (September-November). For ease of reading, the word "boreal" will be dropped in the remainder of the paper.

3. Methodology

a. Estimation of Joint Probabilities

In this study, the predictability of seasonal mean precipitation is investigated using joint probabilities, defined as the probability of precipitation in one season jointly with the probability in the subsequent season. The state of precipitation at each pixel is characterized by terciles of the seasonal mean precipitation. These terciles were determined by standard methods independently for each pixel; namely by sorting the observed precipitation at each pixel separately and then identifying the values below which 33% and 66% of the observations fall. Because of the non-Gaussian nature of

precipitation, the terciles are not symmetrically distributed relative to the median. Since terciles correspond to three levels, the joint probability has 9 elements. The available 29 years of data is not enough to obtain an accurate estimate of the nine elements of the joint probability *for a single pixel.*

To estimate the joint probability at a pixel, we first partition the precipitation field into the 21 land boxes proposed by Giorgi and Francisco (2000) and the 5 ocean boxes defined in table 1. All study areas are defined in table 1 and illustrated in figure 1. Each box will be called a "study area," while the grid cells within each study area will be called "pixels." Next, we estimated a single joint probability for the study area by pooling data from all pixels in the study area. This approach effectively assumes that the transitional probabilities in a study area are homogeneous and stationary—homogeneous in the sense that the joint probability at a pixel equals that of all other pixels in a study area, and stationary in the sense that the joint probability does not change in time. The homogeneity assumption does not imply that the actual states in a study area are equal, but only that the transition probabilities are equal. We emphasize that tercile/state estimations were performed separately for each pixel,—the same state may not refer to the same precipitation amount in the neighboring pixel. Some representative examples of the estimated joint probabilities are given in table 2 and discussed in the section 5.

Some comments about our proposed methodology are in order. First, we note that the boxes chosen for this study were not chosen to maximize predictability; rather, the boxes were chosen independently of this study and for a different purpose (namely, to analyze the effect of different scenarios of greenhouse gas and sulfate forcing). Second, the chosen boxes tend to encompass similar climate zones (e.g., deserts are boxed

together). We suggest that transition probabilities are more likely to be homogeneous for a box that contains a single climate zone than for a box that contains many climate zones. Third, our study period of 1979-2008 avoids crossing the year 1976, which has been suggested as a year of possible climate transition (Trenberth 1990).

Finally, if the joint probabilities in a box are neither homogeneous nor stationary, then pooling pixels within a box is not likely to yield predictability. As a specific example, suppose one area has a high probability of transitioning from wet to dry and dry to wet, and a neighboring area has high probability of transitioning from dry to dry and wet to wet. Pooling these areas tends to even out the transitions, yielding transition probabilities close to climatology. Similarly, if the process is nonstationary, then pooling data over non-stationary periods gives predictability "in between" those in individual stationary periods, thereby evening out the predictability. Thus, violation of homogeneity and/or stationarity tends to diminish predictability. In essence, rather than assuming homogeneity and stationarity, our method actually tests whether pixels are sufficiently homogeneous and stationary to detect predictability.

b. Measuring Predictability

The null hypothesis of this study is that the seasonal mean precipitation in one season is independent of the seasonal mean precipitation in the next season. A test of this hypothesis using autocorrelation methods would be limited because such methods assume linear dependence whereas precipitation is highly non-Gaussian and hence potentially nonlinear. Therefore a distribution independent method was used to quantify the precipitation dependency. Specifically, following Agresti (2002, p79), DelSole

(2004),DelSole and Tippett (2007), the statistic used to test the hypothesis is mutual information, defined as

$$M = \sum_{f}\sum_{p} P_{fp} * \log(\frac{P_{fp}}{P_{f*}P_{p}})$$

where P_{fp} is the joint probability of future and present season states, where "present" and "future" refer to two consecutive seasons, and P_f and P_p are the marginal probabilities of future and present season states respectively. Since terciles were used in this study, P_{fp} is a 3x3 matrix where each row and column sums approximately to 1/3 and all nine elements sum to 1. The marginal probabilities P_f and P_p are vectors with 3x1 dimension and each vector also sums to 1. If the present and future states are independent, then, by definition, $P_{fp}=P_f*P_p$, in which case mutual information vanishes. A standard result in information theory is that the maximum mutual information is given by the maximum entropy of the marginal distribution (Cover and Thomas 1991, p. 190), which in our case is log(3). Therefore, in the present study based on terciles, the mutual information between present and future states is between 0 and log(3), with the former value corresponding to all lack of predictability and the latter to perfect predictability.

To test the significance of the observed value of mutual information, a bootstrap method was employed in which each observed precipitation field in a given season was paired with a randomly selected precipitation field from the corresponding subsequent season in a non-matching year. This bootstrap sample is of the same size as the original data set and hence the joint probabilities and a single M-value can be computed from this sample. Repeating this procedure for 1000 times, separately for each study area and for each season, yields a sample of M-values drawn from the null distribution (i.e., drawn from a population in which the precipitation is independent from season to season). From

this sample, the upper 5% tail was estimated to determine the 5% threshold for statistical significance.

c. Prediction and Verification

If the above procedure rejects the hypothesis that the present and future precipitation states are statistically independent, then it is of interest to verify the discovered predictability with actual forecasts (or, with hindcasts at least). The prediction of seasonal precipitation can be done in two ways: probabilistically and quantitatively.

PROBABILISTIC PREDICTION AND ITS VERIFICATION

Probabilistic forecasts were constructed by using the law of conditional probabilities: $P_{f|p} = P_{fp} / P_p$, where $P_{f|p}$ is the probability of the future state f given the present state p. Specifically, given the observed seasonal precipitation state at a pixel, the transition probability $P_{f|p}$ (estimated by pooling all pixels in the study area) immediately gives the probability that precipitation in that pixel falls in the wet, dry, or normal terciles in the next season. Since all marginal probabilities are approximately 1/3, and the joint probabilities are assumed equal at all pixels in a box, the transitional probability is approximately equal at each pixel. Nevertheless, the actual probabilistic forecast differs from pixel to pixel because the initial state of precipitation differs from pixel to pixel. Construction of separate probabilistic forecasts for each pixel is particularly important for locations with major topographical change, like mountains or land-sea boundaries, which may have very different states from those of other pixels in the study area.

To avoid overfitting, cross-validation was performed by leaving one year out for verification and using the remaining years to calculate the joint probability P_{fp}. This procedure was repeated for each year withheld. The skill of the probabilistic forecast was measured using the Brier score (Brier, 1950), which is the sum squared difference between the forecast and verification probabilities over all terciles. The "verification probability" equals one for the tercile in which the precipitation fell at the verification time, and equals zero for the other two terciles. The mean Brier score for a particular season was averaged within each study area over all years.

Two reference forecasts are used as a basis of comparison, namely climatology and persistence (Wilks 1995, p. 347). The climatology forecasts are probabilistic forecasts equal to the marginal distributions, as estimated from all years excluding the verification year; the probability in each category is close to 1/3. Persistence forecasts are probabilistic forecasts such that the tercile in which the precipitation fell in the prior season is assigned a value of one at verification time, while the other terciles at the verification time are set to zero (i.e. say 1990 spring was dry for a pixel, then the transitional probability from spring to summer would be [1,0,0] representing dry, normal, and wet conditions respectively. From these probabilities, the persistence forecast for the summer 1990 would be [1,0,0] for that particular pixel and date). Both climatology and persistence forecasts were performed locally, without any information from the neighboring pixels.

QUANTITATIVE PREDICTION AND ITS VERIFICATION

A quantitative prediction from transition probabilities was constructed by computing the expected (or mean) precipitation. More specifically, the local seasonal mean values of dry, normal, and wet states were weighted according to the transition probability and summed over these states to find the quantitative forecast for each pixel. Mathematically, quantitative forecasts are calculated as

$$P_{predic} = \sum (P_{wet|p} * \mu_{wet} + P_{normal|p} * \mu_{normal} + P_{dry|p} * \mu_{dry})$$

where P_{predic} is the quantitative prediction; $P_{wet|p}$, $P_{normal|p}$, and $P_{dry|p}$ are the appropriate transitional probabilities for the forecast; and μ_{wet}, μ_{normal}, and μ_{dry} are the seasonal mean precipitations of the wet, normal, and dry terciles respectively.

In addition to climatology and persistence based quantitative forecasts, a third reference was added. This new reference, called a random quantitative forecast, was obtained by randomly selecting a precipitation field for the same study box. This field was picked from the appropriate season of a random year excluding the verification year. The total root mean square errors (RMSE) of these four quantitative precipitation forecasts were calculated over space and time;

$$RMSE = \sqrt{< (P_{predic} - P_{verif})^2 >}$$

where P_{verif} is the verification year precipitation values, and the squared-difference values are averaged over both space and time. RMSE was calculated for each study area and each season separately.

4. Discussion of Climatological and Persistence Forecasts

Although the material in this section is straightforward, it does not appear to be conveniently available in the predictability literature. Therefore, we include it for reference. As is well known, the mean square error of a climatological forecast is simply the climatological variance. The mean square error of a persistence forecast for a stationary process can be shown to be

$$< (f(t+\tau) - f(t))^2 > = 2\sigma^2{}_f (1 - \rho_\tau)$$

where <> denotes an expectation, τ is the time lag, *f(t)* is the verification at time *t*, $\sigma^2{}_f$ is the variance, and ρ_τ is the auto-correlation at time lag τ. Here, we interpret *f(t)* as the "persistence forecast" and *f(t+τ)* as the verification. It is readily seen that a persistence forecast has less MSE than a climatological forecast (i.e., MSE < $\sigma^2{}_f$) only when the autocorrelation between initial and final states exceeds 0.5. Since the autocorrelation of geophysical variables generally decreases with lead time, persistence forecasts are expected to be better at short lead times while climatology forecasts are expected to be better at long lead times. The transition between "short" and "long" lead times occurs at the time for which the autocorrelation equals 0.5.

5. Results

The joint probability for all 26 study areas and all 4 seasons (totally 104) were estimated from the 29-year GPCP data set. The estimated joint probabilities for four representative regions, as well as the joint probability for all regions and seasons together are shown in table 2. Note that if all states were independent of each other, then all table entries would be 1/9 ~ 0.11. The probabilities tend to be larger along the "diagonal",

implying a preference for persistence— that is, the probability that precipitation remains in the same state is larger than the probability to jump to a different state. However, if the state does jump, then the tables generally show a preference for "single jumps"-- that is, the transition from one extreme category to the other is less likely than from one extreme to normal. For example, for Cold Tongue area (Table 1) the probabilities for dry to dry, normal to normal, or wet to wet (diagonal) transitions are higher than any other transition (implying persistence). Also for the same region, dry to normal and wet to normal transitions are more probable than dry to wet or wet to dry transitions (implying a preference to "single jump").

The results of evaluating and testing the significance of mutual information in all study areas and all seasons are shown in fig. 2. Mutual information exceeded the 5% significance level in 67 out of 104 areas. Spring was the most predictable season (23 out of 26 study areas had predictability) whereas summer was the least predictable season (only 11 out of 26 study areas). The largest value of mutual information for any region is about 0.04, which is well below the theoretically maximum value of 1.09 (corresponding to perfect predictability). The most predictable areas are Amazon, N.W. America, E. Africa, Sahara, S.E. Africa, and Pacific (all four seasons), while Central N. America was the only region that no predictability was found for any season.

Brier scores of the probabilistic forecasts were smaller than that of either climatological probabilistic forecasts or persistence forecasts for all seasons and study areas (Table 3). Interestingly, climatological forecasts had smaller Brier scores than persistence forecasts for all seasons and study areas (Table 3).

RMSE of quantitative forecasts from the transition probabilities were smallest among all quantitative predictions, though only slightly better than climatology (Table 4). RMSE's of climatological quantitative forecasts were smaller than that of persistence forecasts. Not surprisingly, random forecast had the highest RMSE's for all seasons and study areas (Table 4).

6. Summary and Discussion

Seasonal precipitation predictability was investigated using GPCP monthly data from 1979 to 2008 over 21 land boxes defined in Giorgi and Francisco (2000) and 5 ocean boxes. The state of precipitation was defined by wet, normal, and dry terciles at each pixel. The transition probability from one season to a subsequent season was estimated by pooling all pixels within a study box. Mutual information was used to measure the degree of dependence between consecutive seasons and it was estimated separately for 26 study areas and 4 seasons. Probabilistic forecasts derived from transition probabilities were cross-validated and compared to two benchmark forecasts, namely climatology and persistence. Also, quantitative forecasts were derived by computing the expected precipitation from the transition probabilities, and compared to three benchmark forecasts, namely climatology, persistence, and random selection.

The results show that mutual information is statistically different from 0 in most study areas, implying statistical dependence between consecutive seasons (i.e., implying predictability). Although mutual information was found to be statistically significant in many study areas, the actual values of mutual information were typically less than 0.05, which is small relative to the maximum possible value of $\log(3) \sim 1.09$ (corresponding to

perfect predictability in which the transition probability is unity for one category and zero in all other categories). Presumably, larger values of mutual information can be obtained by suitable spatial filtering (e.g., area averages or canonical correlation analysis), but the optimal spatial filtering would be difficult to obtain from such a small sample.

Examination of all joint probabilities reveals that most elements are close to the climatological value of 1/9 ~ 0.11. The predictability between seasons can be attributed to a slight tendency for precipitation to persist in the same state from one season to the next, or if there is a transition to prefer "single jumps" rather than jumping from one extreme to the other (Table 2). Transitional probabilities estimated by pooling data in a study box produced better forecasts than climatology, persistence, and random selection.

Our results are relevant to understanding whether persistence forecasts are better than climatology forecasts. Specifically, the skill of forecasts based on the climatological seasonal mean precipitation was shown to be superior to persistence forecasts in all 21 land areas and 5 ocean areas considered, regardless of whether forecasts were probabilistic or quantitative. In general, climatology forecasts are superior to persistence forecasts, in a mean square error sense, when the lagged autocorrelation function is less than 0.5 (see sec. 3). Thus, the superiority of climatology-based forecasts over persistence-based forecasts is consistent with the weak autocorrelation of precipitation between seasons.

Spring was found to be the most predictable season (23 out of 26 study areas had predictability) whereas summer was the least predictable season (only 11 out of 26 study areas). The most predictable areas were Amazon, N.W. America, E. Africa, Sahara, S.E.

Africa, and Pacific (all four seasons), while Central N. America was the only region that no dependency was found for any season.

We have investigated the predictability of seasonal precipitation directly from joint probabilities without assuming anything about the physics or distribution of precipitation. Perhaps the most intriguing result of this study is the demonstration that conditional probabilities estimated by pooling seasonal precipitation data on sub-continental scales can give skillful forecasts of individual pixels. It seems likely that this result also extends to other variables such as temperature. While this study showed that seasonal mean precipitation is predictable, the degree of predictability is relatively small (but statistically significant) and the underlying mechanism for the predictability was not elucidated. It is not clear how the non-parametric methodology studied in this paper can be extended to large multivariate data sets to elucidate space-time relations between precipitation and associated boundary conditions.

Acknowledgements

We thank anonymous reviewers for constructive comments that lead to numerous clarifications in the final manuscript. This work was supported by the NSF(ATM-0830068, ATM-0830062), NOAA (NA09OAR4310058), and NASA (NESSF 09-Earth09R-80 and NNX09AN50G).

References

Adler, R. F., and Coauthors, 2003: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present). *J. Hydromet.*, **4**, 1147–1167.

Agresti, A., 2002: *Categorical Data Analysis*. John Wiley and Sons, 2nd ed, p710.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probabilities. *Mon. Wea. Rev.,* 78, 1–3.

Cover, T. and Thomas, J. 1991: Elements of Information Theory. Series in Telecommunications. John Wiley & Sons, 576 pp.

DelSole, T., 2004: Predictability and Information Theory. Part I: Measures of Predictability. *J. Atmos. Sci.*, **61**, 2425–2440.

DelSole, T., and M. K. Tippett, 2007: Predictability: Recent Insights from Information Theory. *Rev. Geophys.,* **45**, RG4002, doi:10.1029/2006RG000202.

Dirmeyer, P.A., C.A. Schlosser, and K.L. Brubaker, 2009: Precipitation, Recycling, and Land Memory: An Integrated Analysis. *J. Hydrometeor.*, 10, 278–288.

Giorgi F., and R. Francisco, 2000: Uncertainties in regional climate change predictions. A regional analysis of ensemble simulations with the HADCM2 GCM. *Climate Dyn.*, 16, 169–182.

Glantz, M. H., 1992, Climate Variability, Climate Change and Fisheries. Cambridge University Press, pp. 450.

Hansen, J.W., A.W. Hodges, and J.W. Jones, 1998: ENSO Influences on Agriculture in the Southeastern United States*. *J. Climate*, **11**, 404–411.

Katz, R.W., 1977: Precipitation as a Chain-Dependent Process. *J. Appl. Meteor.*, 16, 671–676.

Katz R. W., 1981: On some criteria for estimating the order of a Markov chain. *Technometrics*, **23**, 243–249.

Kirtman, B.P., and P.S. Schopf, 1998: Decadal Variability in ENSO Predictability and Prediction. *J. Climate*, 11, 2804–2822.

Koster, R.D., M.J. Suarez, and M. Heiser, 2000: Variance and Predictability of Precipitation at Seasonal-to-Interannual Timescales. *J. Hydrometeor.*, 1, 26–46.

Koster, Randall D., et al., 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, 305, 1138-1140.

Namias J., 1952: The annual course of month-to-month persistence in climatic anomalies. *Bull. Amer. Meteor. Soc.*, 33, 279–285.

Namias, J., 1960: Factors in the initiation, perpetuation and termination of drought. Extract of Publ. 51, International Association of Hydrological Sciences (IAHS) Commission of Surface Waters, 81–94. [Available from IAHS Press, Institute of Hydrology, Wallingford, Oxfordshire OXl0 8BB, United Kingdom.].

Trenberth, K.E., 1990: Recent Observed Interdecadal Climate Changes in the Northern Hemisphere. *Bull. Amer. Meteor. Soc.*, **71**, 988–993.

Van den Dool, Huug, 2007: *Empirical methods in short-term climate prediction*. Oxford University Press, 215 pp.

Vinnikov, K. Ya., A. Robock, N. A. Speranskaya, and C. A. Schlosser, 1996: Scales of temporal and spatial variability of midlatitude soil moisture. *J. Geophys. Res.,* 101, 7163–7174.

Wang, G., Y. Kim, and D. Wang, 2007: Quantifying the Strength of Soil Moisture–Precipitation Coupling and Its Sensitivity to Changes in Surface Water Budget. *J. Hydrometeor.*, 8, 551–570.

Wilks D. S, 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.

Wilks, D.S., 2002: Realizations of Daily Weather in Forecast Seasonal Climate. *J. Hydrometeor.*, 3, 195–207.

List of Figures

TABLE 1. Location of the 21 Giorgi boxes and 5 defined ocean boxes that are focus of this study.

Study Area	Lower Lat.	Upper Lat.	West Long.	East. Long.
Australia	-12.5°S	-45°S	110°E	155°E
Amazon Basin	12.5°N	-20°S	-82.5°W	-35°W
Southern South America	-20°S	-55°S	-75°W	-40°W
Central America	30°N	10°N	-115°W	-82.5°W
Western North America	60°N	30°N	-130°W	-102.5°W
Central North America	50°N	30°N	-102.5°W	-85°W
Eastern North America	50°N	25°N	-85°W	-60°W
Alaska	72.5°N	60°N	-170°W	-102.5°W
Greenland	85°N	50°N	-102.5	-10°W
Mediterranean	47.5°N	30°N	-10°W	40°E
Northern EU	75°N	47.5°N	-10°W	40°E
Western Africa	17.5°N	-12.5°S	-20°W	22.5°E
Eastern Africa	17.5°N	-12.5°S	22°E	52.5°E
Southern Africa	-12.5°S	-35°S	-10°W	52.5°E
Sahara Africa	30°N	17.5°N	-20°W	65°E
South-east Asia	20°N	-10°S	95°E	155°E
East Asia	50°N	20°N	100°E	145°E
South Asia	30°N	5°N	65°E	100°E
Central Asia	50°N	30°N	40°E	75°E
Tibet	50°N	30°N	75°E	100°E
North Asia	70°N	50°N	40°E	180°E
North Atlantic	60°N	40°N	-50°W	-20°W
North Subtropic Atlantic	40°N	10°N	-60°W	-20°W
South Atlantic	-10°S	-50°S	-25°W	0°E
Cold Tongue	0°N	-20°S	-120°W	-85°W
Pacific	40°N	-20°S	150°E	235°E

TABLE 2. Joint probability matrices of seasonal precipitation from 1978 to 2008 obtained from monthly GPCP data. Due to space considerations, 4 out of 104 matrices (for 26 study areas and for 4 seasons) and the global average of these 104 matrices are shown below.

		Present		
		Dry	Normal	Wet
		Winter-Autumn, Australia		
Future	Dry	0.13	0.11	0.11
	Normal	0.10	0.11	0.10
	Wet	0.12	0.10	0.10
		Winter-Autumn, SE Africa		
Future	Dry	0.17	0.11	0.07
	Normal	0.10	0.12	0.10
	Wet	0.08	0.09	0.15
		Winter-Autumn, Tibet		
Future	Dry	0.16	0.11	0.08
	Normal	0.12	0.11	0.10
	Wet	0.08	0.10	0.14
		Winter-Autumn, Cold Tongue		
Future	Dry	0.19	0.11	0.06
	Normal	0.10	0.15	0.07
	Wet	0.06	0.07	0.19
		All seasons and areas averaged		
Future	Dry	0.14	0.11	0.10
	Normal	0.11	0.10	0.10
	Wet	0.10	0.10	0.14

TABLE 3. Mean Brier scores in time (29 years) and space (study area average) for all 26 study areas and 4 seasons for transitional probability-, climatology-, and persistence-based forecasts. Since the Brier scores of the same type of forecasts had very close values for all study areas and seasons, only the minimum and the maximum values out of these 104 entries are shown.

Forecast Type	Mean Error	
	Max	Min
Trans. Prob.	0.823	0.792
Climatology	0.846	0.845
Persistence	1.173	0.997

TABLE 4. Quantitative forecasts and their Root Mean Square Errors (RMSE) for different study areas and seasons. QFE, QCE, QPE, and QRE refer to quantative forecast, climatology, persistence, random forecast errors respectively.

	SPRING-SUMMER					SUMMER-AUTUMN					AUTUMN-WINTER					WINTER-SPRING				
	RMSE				Mean	RMSE				Mean	RMSE				Mean	RMSE				Mean
	QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE	
Austrl.	44.6	45.5	56.9	61.5	127	42.7	44.5	50.8	60.0	107	107	110	137	143	232	91.0	92.9	122	136	177
Amazn.	85.2	89.6	101	119	302	81.9	85.6	98.1	115	345	118	122	145	165	452	125	129	150	158	449
S. Amer.	64.6	66.2	82.6	91.2	226	63.8	65.9	78.3	94.0	247	69.8	71.5	89.3	98.4	273	76.7	78.4	98.3	104	273
C. Amer.	129	132	165	171	460	118	122	149	161	388	63.9	65.3	80.0	93.4	128	65.2	68.1	78.9	93.9	133
WN Amr	38.0	39.0	47.6	51.6	120	48.5	49.8	60.9	64.7	133	60.0	62.7	71.8	83.8	152	41.9	43.9	50.2	61.4	122
CN Amr	75.8	77.2	97.4	108	287	72.1	73.4	97.1	96.0	225	57.5	58.6	73.0	81.6	167	77.3	79.3	95.1	102	253
EN Amr	76.1	77.9	99.7	99.1	339	88.0	90.1	116	123	347	70.1	71.5	91.6	95.3	321	71.9	74.0	87.9	99.1	304
Alaska	34.7	35.5	42.9	47.2	134	45.2	46.6	55.8	62.5	117	37.0	39.8	40.5	56.1	78	26.7	28.9	30.2	39.6	62
Greenlnd	33.9	34.9	42.6	46.2	148	41.7	43.4	51.1	56.4	178	47.1	49.1	56.3	65.0	157	31.7	33.2	37.7	47.4	111
Meditrrn.	36.1	37.0	45.3	50.6	86	61.8	63.3	81.5	87.9	170	66.5	68.0	85.4	101	187	44.2	45.4	56.6	62.3	139
N. EU	49.9	51.1	65.2	68.1	207	61.1	62.5	79.2	82.0	247	71.1	73.0	86.9	97.2	256	42.1	43.9	49.9	60.2	167
W. Afric.	77.9	79.9	98.6	98.8	286	64.6	67.9	76.9	95.4	233	51.2	52.9	61.6	67.7	125	77.9	80.3	94.3	110	241
E. Africa	52.3	54.3	63.2	70.5	173	63.1	65.4	76.4	97.1	172	64.3	67.6	76.0	93.2	181	73.7	75.9	90.7	107	234

TABLE 4. Continuation

	SPRING-SUMMER					SUMMER-AUTUMN					AUTUMN-WINTER					WINTER-SPRING				
	RMSE				Mean	RMSE				Mean	RMSE				Mean	RMSE				Mean
	QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE		QFE	QCE	QPE	QRE	
S. Africa	25.7	26.2	33.1	35.5	64	42.8	44.3	52.5	59.0	100	88.2	90.4	113	114	237	57.6	59.2	73.7	83.0	138
Sahara	19.2	19.5	24.5	24.5	19	17.7	18.0	21.9	26.5	17	22.8	23.6	27.5	37.7	26	16.9	17.6	20.3	22.6	20
SE. Afrc	157	163	191	217	582	162	170	194	218	545	148	159	168	207	486	136	144	156	195	449
E. Asia	120	123	155	170	439	88	90	119	127	253	46.9	47.8	63.4	67.6	128	78.0	80.7	95.7	104	257
S. Asia	142	146	182	187	619	112	115	140	156	362	59.9	61.7	76.9	86.2	87	77.8	80.1	94.5	106	193
C. Asia	31.9	33.4	37.9	47.1	64	35.1	36.4	42.9	51.2	78	36.0	36.7	46.4	50.8	113	40.5	42.5	47.4	56.2	112
N. Asia	40.9	42.3	50.6	56.7	125	18.8	19.3	23.5	26.3	49	12.9	13.1	17.1	18.0	28	18.8	19.5	22.6	23.8	50
Tibet	45.6	46.6	59.4	64.9	192	40.2	41.4	50.6	53.8	173	34.0	34.9	42.8	46.3	125	31.8	32.5	40.9	44.6	112
N. Atlan	49.6	50.7	64.4	70.6	268	63.9	65.5	83.4	88.6	384	73.3	74.9	94.5	97.4	475	56.1	57.3	73.6	77.1	288
N. S. At.	48.3	49.7	59.6	64.7	119	81.6	83.8	105	114	251	58.0	59.3	75.6	81.4	173	55.3	57.3	67.7	77.0	127
S. Atlan	43.0	44.1	55.2	57.7	188	44.7	46.1	57.7	65.3	161	41.6	42.5	54.2	59.7	147	49.2	50.4	62.9	69.9	206
Cld Tng	13.3	13.5	16.4	14.0	19	6.2	6.5	7.0	9.5	12	44.3	44.8	58.7	65.5	43	81.1	88.2	89.8	131.	115
Pacific	112	119	132	165	286	122	130	140	171	327	164	176	180	242	367	145	155	164	211	327

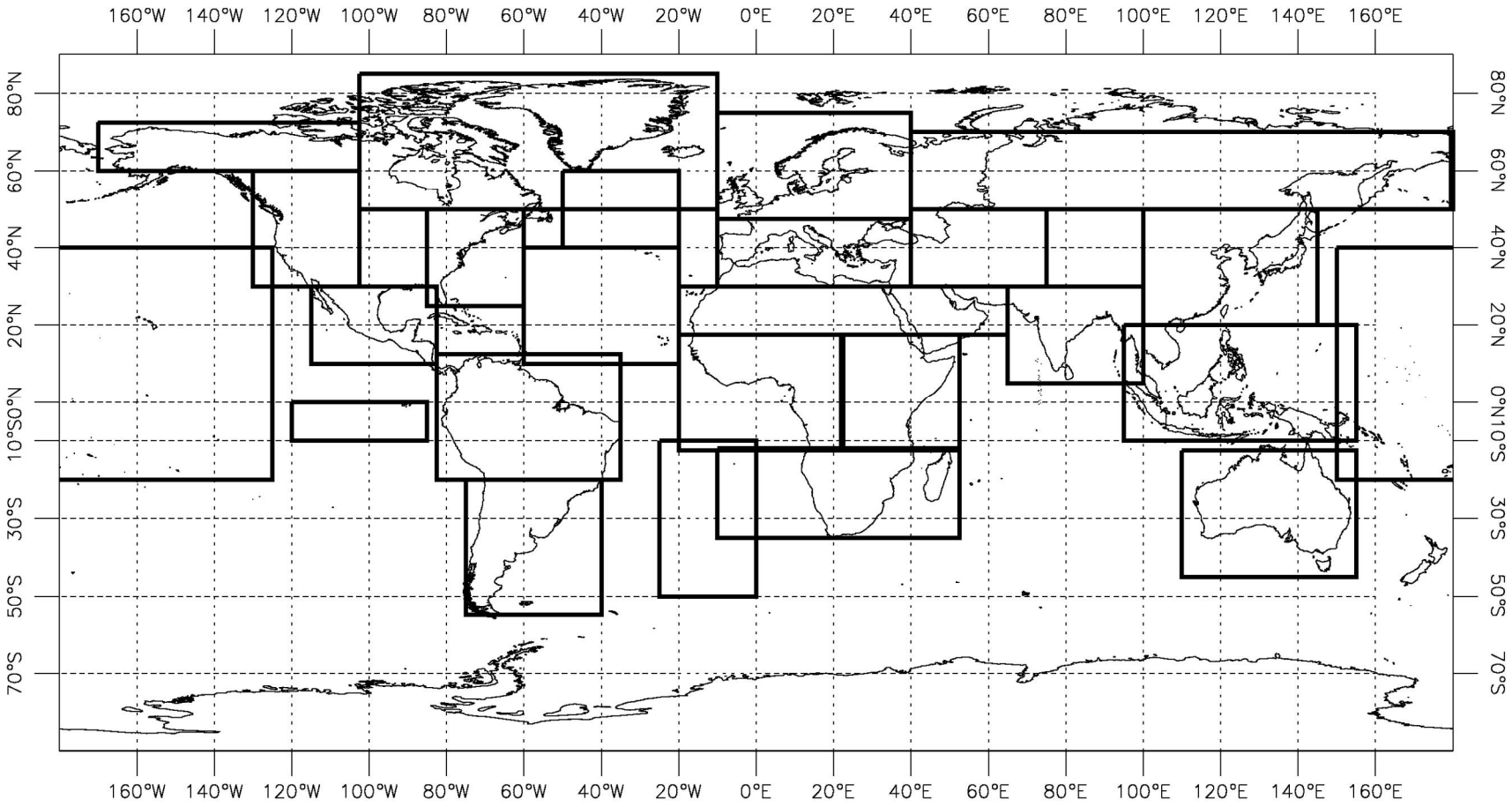

FIG. 1. Geo-locations of the 26 study areas (Latitude and Longitude coordinates are given in Table 1).

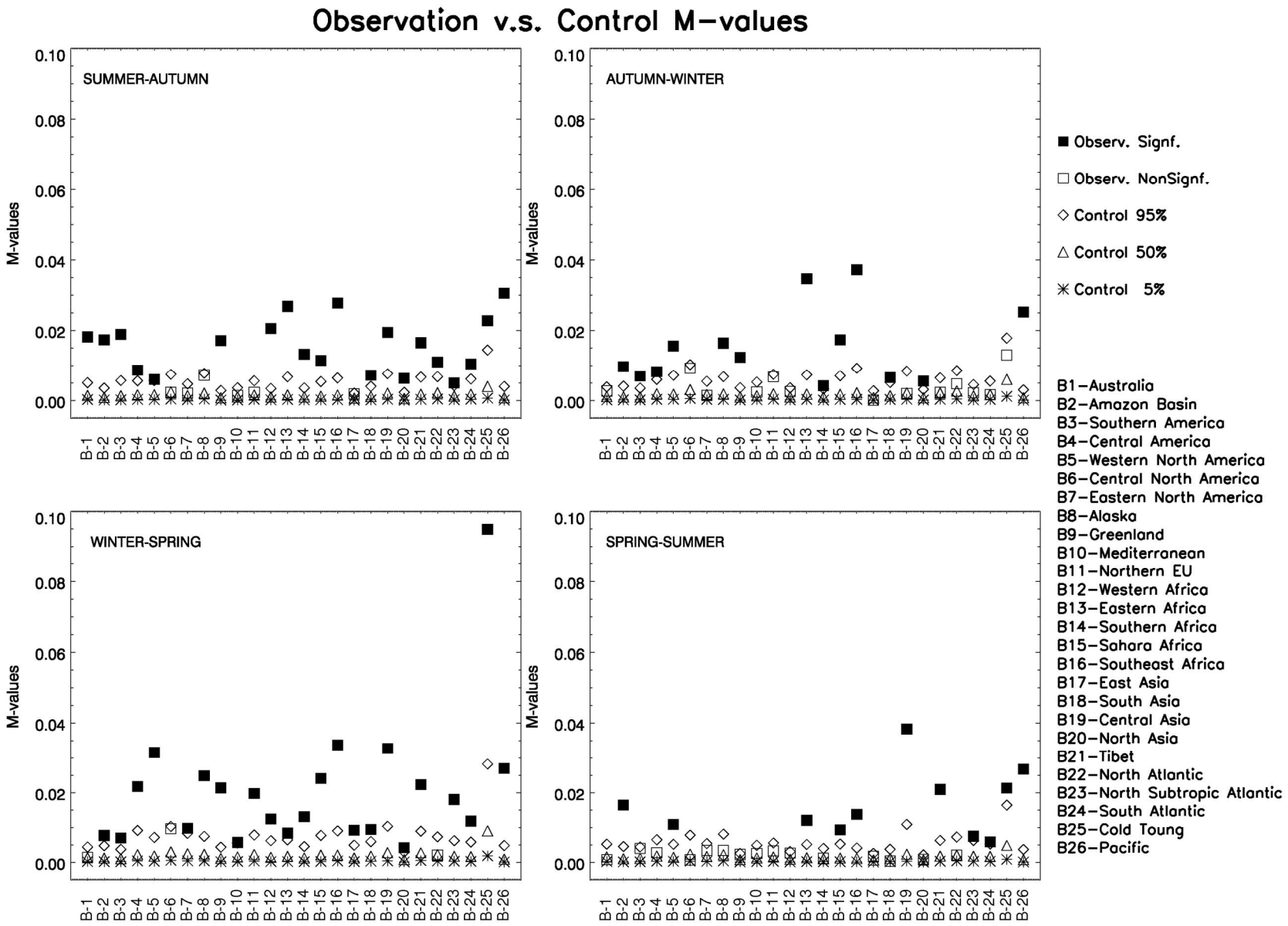

FIG. 2. Observed mutual information (M) and its null distribution for 4 seasons and 26 study areas.