Point-Biserial Correlation-Based Skill Scores for Probabilistic Forecasts

Nachiketa Acharya[1*] and Michael K Tippett[2]

[1]International Research Institute for Climate and Society, Earth Institute at Columbia University, New York, NY, USA
[2]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA

*nachiketa@iri.columbia.edu

International Research Institute
for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

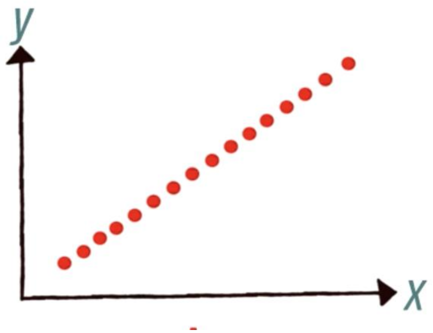
2020-IVMW-O

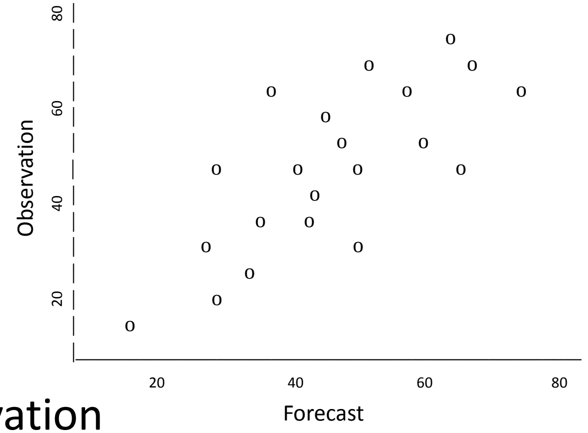
Motivation

The correlation coefficient is a popular skill score for the deterministic forecast system.

Pearson's correlation

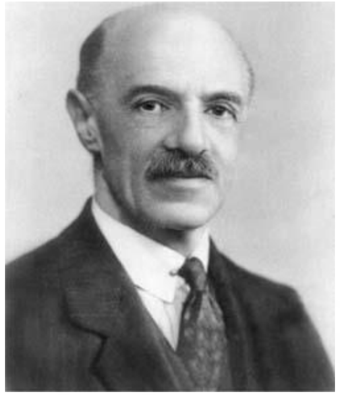
$$r = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2}\sqrt{\sum_i^n (y_i - \bar{y})^2}}$$



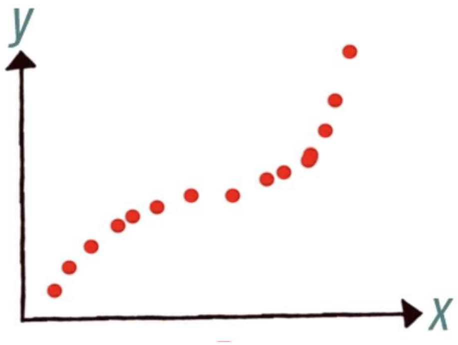


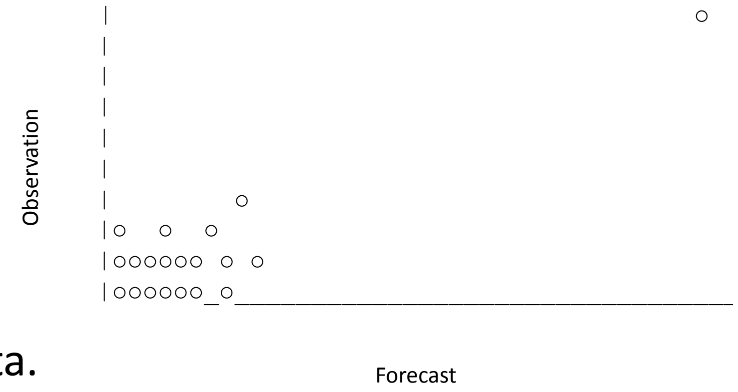
Measuring the strength of Linear relationship between forecast and observation

Spearman's correlation



$$\rho = 1 - \frac{6\sum_{i=1}^{n}\left(r_{x_i} - r_{y_i}\right)^2}{n\left(n^2 - 1\right)}$$





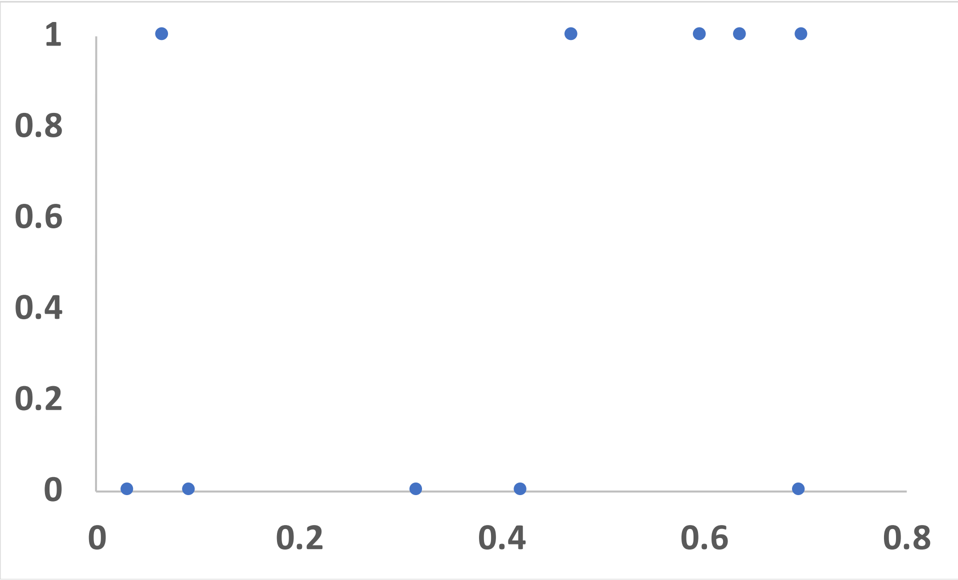
Based on the ranked values for each variable rather than the raw data.
Measuring non-Linear relationship.

It is much less sensitive to extremes!

Motivation

What if…

Forecast Probability	Rain
0.4229	No
0.0942	No
0.5985	Yes
0.4709	Yes
0.6959	No
0.6999	Yes
0.6385	Yes
0.0336	No
0.0688	Yes
0.3196	No



Can we still calculate Correlation coefficient for this case??

NOTES

CORRELATION BETWEEN A DISCRETE AND A CONTINUOUS VARIABLE. POINT-BISERIAL CORRELATION

By Robert F. Tate

University of Washington[1]

1. Introduction and Summary. A problem of some importance in statistical applications, especially in the field of psychology, is that of finding a measure of association between a discrete random variable X, which takes the values 0 and 1, and a continuous random variable Y. The ordinary product-moment correlation coefficient $\rho(X, Y)$ is used for this purpose. It has received the name point-biserial correlation coefficient because of its relation to the biserial correlation coefficient proposed by Karl Pearson for a similar problem. The usual estimator r, based on a random sample (X_i, Y_i), $i = 1, 2, \cdots, n$, is referred to as the sample point-biserial correlation coefficient.

The psychological value of ρ (and hence of r) is that it affords a measure of the degree of association between a trait and a measurable characteristic, usually an ability of some kind. For the ith individual in a random sample of n individuals, X_i has the value 1 if the trait is possessed and Y_i is a measure of the ability in question.

We shall give in Section 2 the appropriate mathematical model, based on normal theory, and the asymptotic distribution of r (Theorem 1), the derivation of which is an elementary application of a well known theorem of Cramér. An important special case of this distribution will be discussed in Section 3, namely that in which X takes the values 0 and 1 with equal probabilities. In this connection a variance-stabilizing transformation will be given (Theorem 2). Numerical work based on this transformation may be carried out with the use of existing tables. In particular, the calculation of confidence limits for ρ is immediate. Theorem 2 is especially useful in investigating the association between sex and some other characteristic, since animal populations consist of approximately half males and half females. As an illustration of the ease with which calculations may be carried out, a problem is considered in which the trait is male and the characteristic is IQ.

The small-sample distribution of r is quite easily found, although it is difficult to deal with when n is even moderately large, asymptotic methods appearing to be more desirable. This is discussed in Section 4.

Received 6/29/53, revised 1/26/54.

[1] This research was performed while the author was at the Statistical Laboratory, University of California, Berkeley, and was sponsored in part by the Office of Naval

Goal

- Most of the existing skill scores for probabilistic forecasts focusing either on the mean squared error in probabilistic space (Brier score) or degree of correspondence between issued forecast probabilities and relative observed frequencies (reliability diagrams) or the degree of correct probabilistic discrimination in a set of forecasts (ROC).

- Proposing Correlation-Based Skill Scores for Probabilistic Forecasts.

Point biserial correlation

X and Y, where Y is in interval or ratio scale with normal distribution while X is a naturally dichotomous variab

Forecast Probability	Rain
0.4229	No
0.0942	No
0.5985	Yes
0.4709	Yes
0.6959	No
0.6999	Yes
0.6385	Yes
0.0336	No
0.0688	Yes
0.3196	No

$$r_{pb} = \frac{(\bar{Y}_1 - \bar{Y}_0)}{Sd_Y}\sqrt{\frac{p_X}{q_X}}$$

$\bar{Y}_0$ = mean of Y for individuals scoring o on X.
$\bar{Y}_1$ = mean of Y for individuals scoring 1 on X.

Sdy = the standard deviation of the continuous data.
px = proportion of samples in group 0.
qx = proportion of samples in group 1.

Significance Testing

$$t = \sqrt{N-2}\frac{r_{pb}}{\sqrt{1-r_{pb}^2}}$$

If X is a artificially occurring dichotomous variable

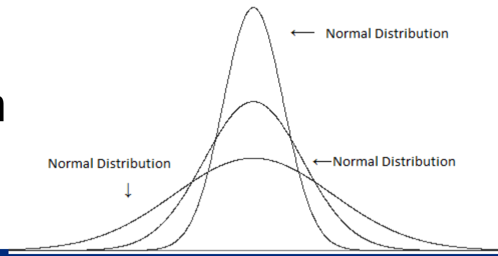
Forecast	Rain<33%le
0.4229	No
0.0942	No
0.5985	Yes
0.4709	Yes
0.6959	No
0.6999	Yes
0.6385	Yes
0.0336	No
0.0688	Yes
0.3196	No

Biserial correlation

The biserial correlation coefficient can also be computed from the point-biserial correlation coefficient

$$r_b = r_{pb}\frac{\sqrt{p_X\, q_X}}{(\lambda)}$$

λ is the ordinate (height) of standardized n distribution



Numerically, r_b obtained is always greater than r_{pb}.

Is it a very "new" in climate field?

Geofísica Internacional (2002), Vol. 41, Num. 2, pp. 203-212

Biserial correlation between vorticity field and precipitation: Rainfall diagnosis and prediction

Nora E. Ruiz
Depto. de Ciencias de la Atmósfera y los Océanos, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

Received: January 29, 1999; accepted: January 22, 2002.

RESUMEN

Este trabajo concierne al examen de una metodología de la climatología sinóptica, la técnica de correlación biserial, que permite investigar, en este caso, la interrelación entre la circulación atmosférica y la precipitación. Se analiza el significado de los campos de correlación biserial obtenidos relacionando distintas variables representativas del flujo de escala sinóptica, particularmente campos de vorticidad, con la precipitación local, con el propósito de ahondar en metodologías que sean simples, eficientes y fáciles de interpretar para ligar la circulación de gran escala con la pequeña escala o local. Se propone una interpretación basada en las configuraciones de los campos de correlación biserial entre vorticidad en 500 hPa y precipitación, que tiene en cuenta los gradientes de vorticidad anómala con los efectos de cortante y curvatura involucrados, para elucidar posibles mecanismos que favorecen la ocurrencia de lluvia. Las anomalías en la curvatura de los sistemas sinópticos son en gran medida responsables de la precipitación. Se utiliza como ejemplo la precipitación diaria de Córdoba, Argentina, para ilustrar los resultados. Se puede identificar claramente la posición de los centros anómalos de vorticidad ciclónica y anticiclónica y de la corriente en chorro en asociación con la ocurrencia de precipitación. El análisis se hace extensivo para precipitaciones más copiosas.

PALABRAS CLAVE: Precipitación, vorticidad, correlación biserial.

ABSTRACT

This work concerns the examination of a methodology of synoptic climatology, the biserial correlation technique, which allows studying the relationship between atmospheric circulation and precipitation. The physical meaning of biserial correlation fields between variables representing synoptic-scale circulation, particularly vorticity fields, and local precipitation is explored. One purpose is to examine this approach used to link the large-scale circulation and the smaller-scale surface environment, which seems to be simple, efficient and easy to interpret. An analysis based on biserial correlation configurations between 500 hPa vorticity and precipitation takes into account anomalous vorticity gradients including curvature and shear effects to describe some mechanisms favoring the occurrence of rainfall. It is shown that anomalies in the curvature of synoptic systems are largely causing precipitation. Daily precipitation at Córdoba, Argentina is used as an example to illustrate the results. The position of the cyclonic and anticyclonic anomaly centers and the position of the jet streams in association with precipitation may be clearly identified. The analysis is made extensive to heavier rainfall.

KEY WORDS: Precipitation, vorticity, biserial correlation.

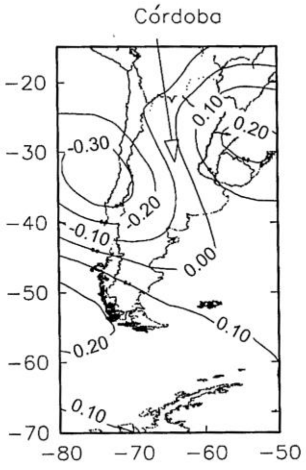


Fig. 1. Biserial correlation field between 500 hPa geopotential heights and the daily precipitation ≥0.1 mm at Córdoba (see text), during the austral summer (November to April).

Extreme floods in central Europe over the past 500 years: Role of cyclone pathway "Zugstrasse Vb"

M. Mudelsee[1]
Department of Earth Sciences, Boston University, Boston, Massachusetts, USA

M. Börngen and G. Tetzlaff
Institute of Meteorology, University of Leipzig, Leipzig, Germany

U. Grünewald
Institute of Hydrology, Technical University Cottbus, Cottbus, Germany

Received 18 May 2004; revised 24 August 2004; accepted 16 September 2004; published 2 December 2004.

[1] Anthropogenically induced climate change has been hypothesized to add to the risk of extreme river floods because a warmer atmosphere can carry more water. In the case of the central European rivers Elbe and Oder, another possibility that has been considered is a more frequent occurrence of a weather situation of the type "Zugstrasse Vb," where a low-pressure system travels from the Adriatic region northeastward, carrying moist air and bringing orographic rainfall in the mountainous catchment areas (Erzgebirge, Sudeten, and Beskids). Analysis of long, homogeneous records of past floods allows us to test such ideas. M. Mudelsee and co-workers recently presented flood records for the middle parts of the Elbe and Oder, which go continuously back to A.D. 1021 and A.D. 1269, respectively. Here we review the reconstruction and assess the data quality of the records, which are based on combining documentary data from the interval up to 1850 and measurements thereafter, finding both the Elbe and Oder records to provide reliable

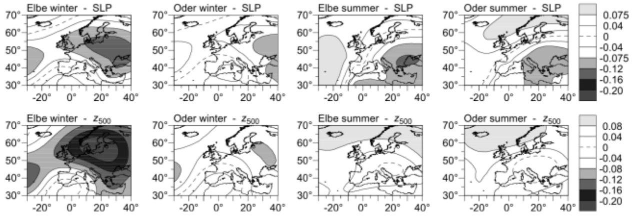


Figure 12. Contour maps of the point-wise biserial correlation coefficient between flood events (Elbe, Oder; winter, summer; classes 1–3) on the one hand and sea level pressure (SLP) or 500 hPa geopotential height (z_{500}) time series on the other; time interval, 1658–1999. Significant correlations (section 3.3) are on color scale. A negative (positive) correlation indicates a pressure below (above) the seasonal average at a geographic point during floods. Elbe and Oder catchment areas are located around 50°N, 15°E (Figure 2). See color version of this figure at back of this issue.

Exploring the Predictability of 30-Day Extreme Precipitation Occurrence Using a Global SST–SLP Correlation Network

MENGQIAN LU AND UPMANU LALL
Department of Earth and Environmental Engineering, and Columbia Water Center, Columbia University, New York, New York

JAYA KAWALE
Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

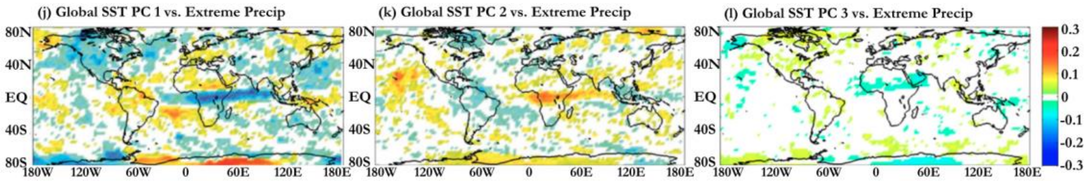
STEFAN LIESS
Department of Soil, Water, and Climate, University of Minnesota, St. Paul, Minnesota

VIPIN KUMAR
Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota

(Manuscript received 16 May 2014, in final form 2 November 2015)

ABSTRACT

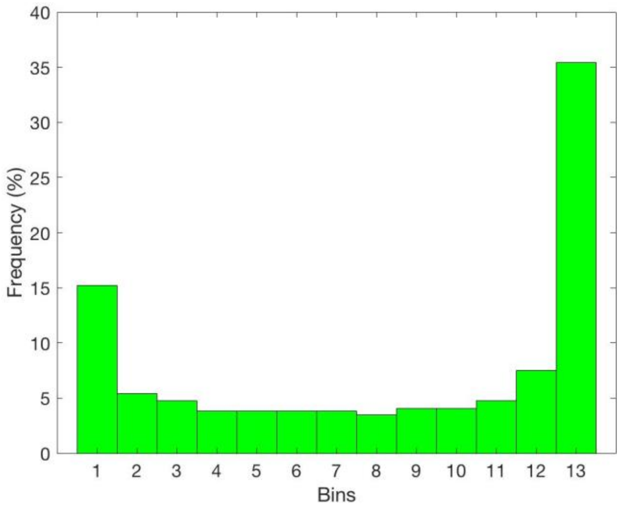
Correlation networks identified from financial, genomic, ecological, epidemiological, social, and climatic data are being used to provide useful topological insights into the structure of high-dimensional data. Strong convection over the oceans and the atmospheric moisture transport and flow convergence indicated by atmospheric pressure fields may determine where and when extreme precipitation occurs. Here, the spatiotemporal relationship among sea surface temperature (SST), sea level pressure (SLP), and extreme global precipitation is explored using a

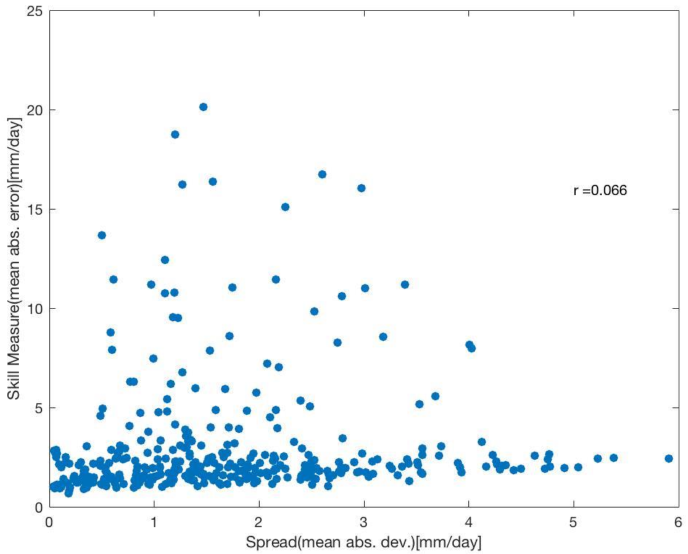


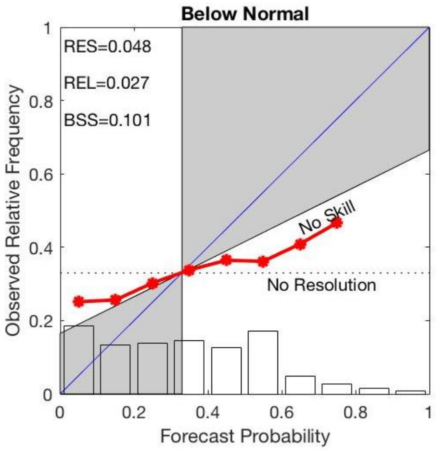
Mostly used for "teleconnection" study but not for forecast verification

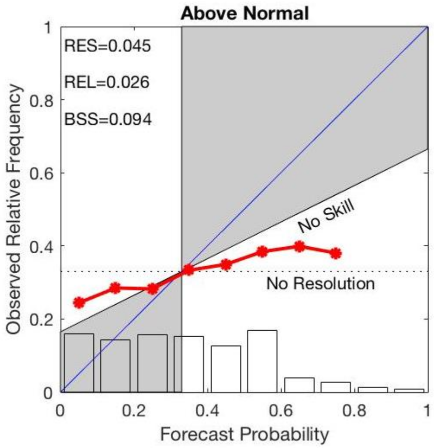
Case study: Measuring Probabilistic Seasonal forecast for Indian Monsoon

Variable: Precipitation
Season: Jun-Jul-Aug-Sep
forecast at: May
Period:1982 to 2010
GCM: GFDL-CM2p5-FLOR-B01
Method: Counting member

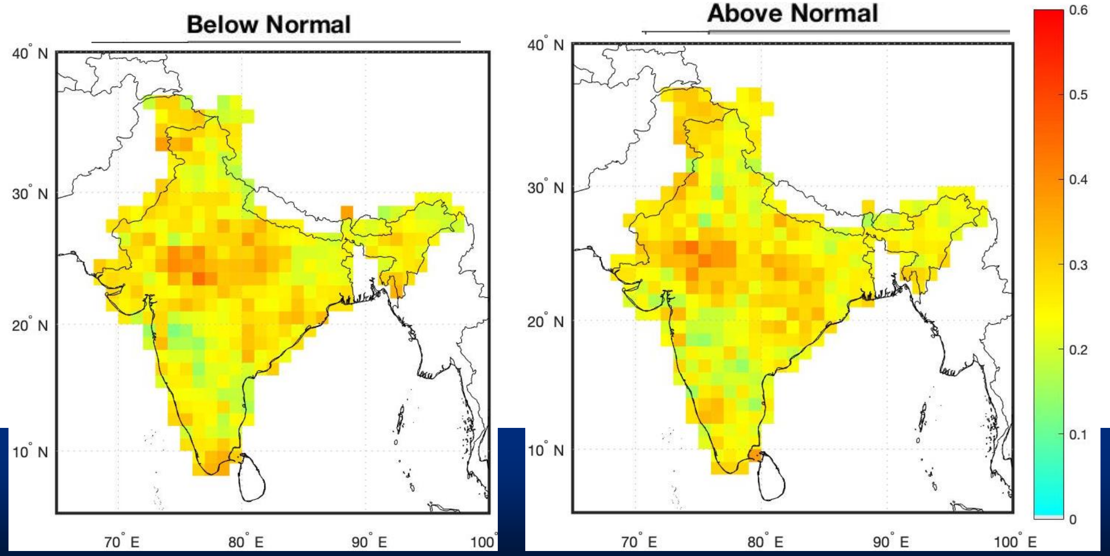




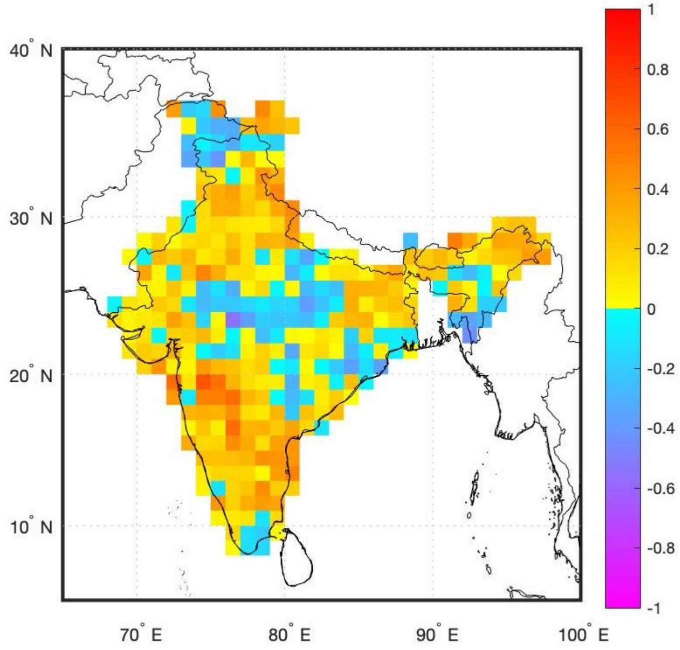




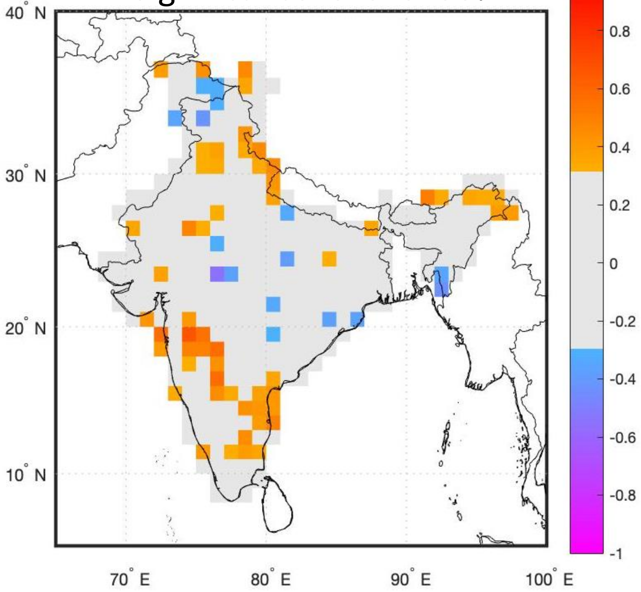
Brier Score



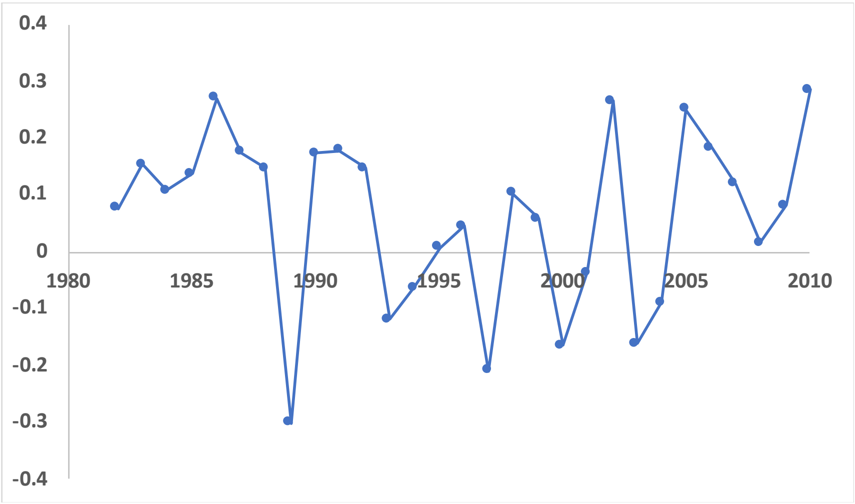
Point Biserial correlation

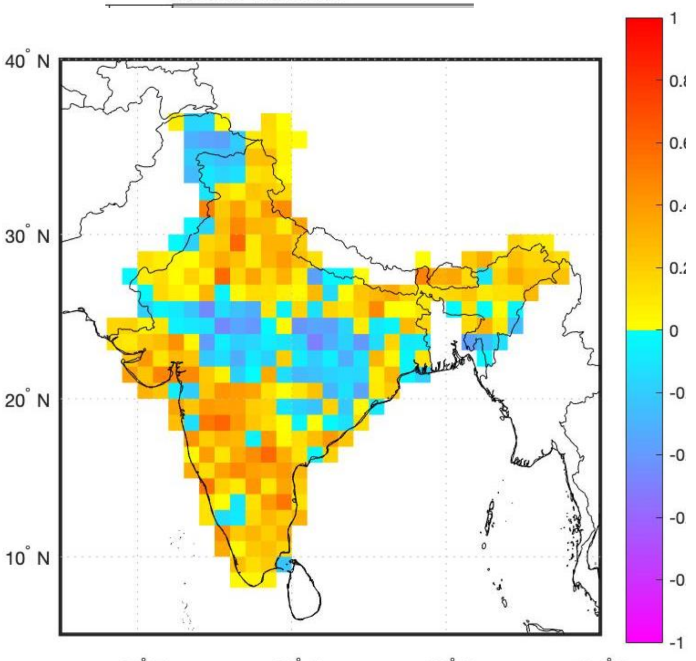


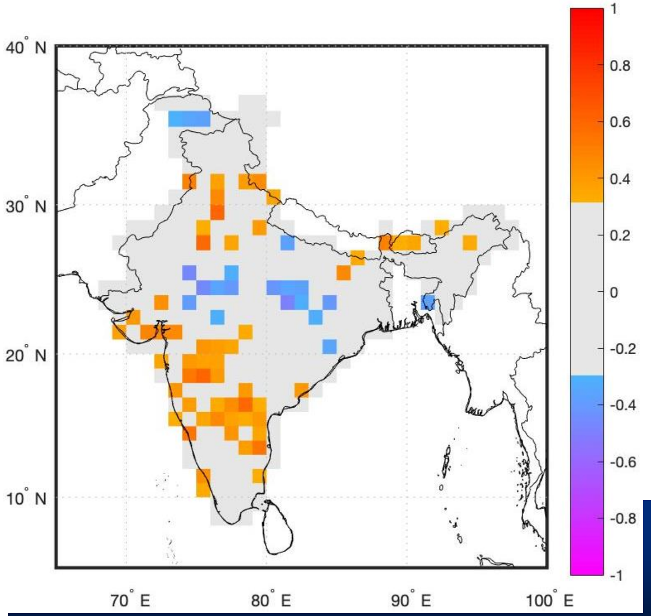
Significance level of 95%

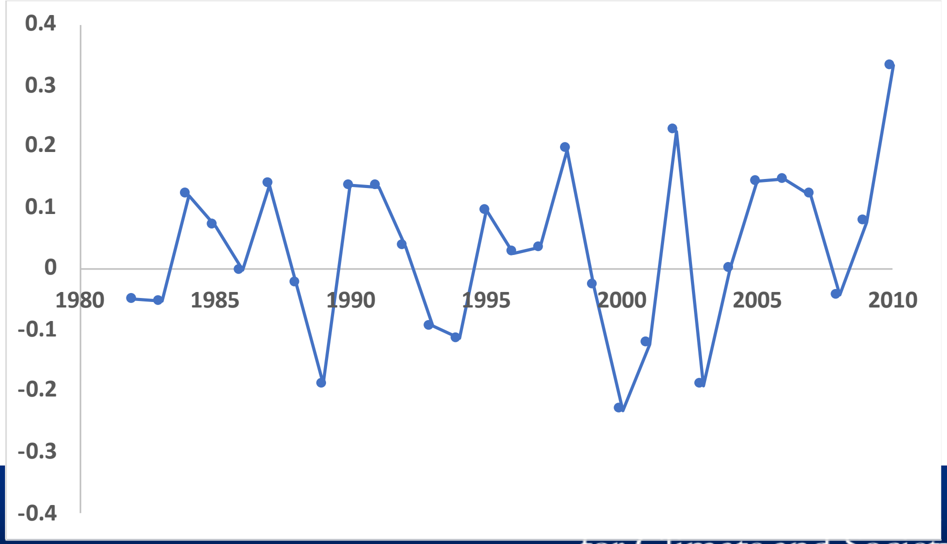


Area Average (~ACC)









Future Direction

Relationship with Brier Score/ Brier Skill Score

VOLUME 116 MONTHLY WEATHER REVIEW DECEMBER 1988

Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient

ALLAN H. MURPHY

Department of Atmospheric Sciences, Oregon State University, Corvallis, Oregon

(Manuscript received 1 February 1988, in final form 11 April 1988)

ABSTRACT

Several skill scores are defined, based on the mean-square-error measure of accuracy and alternative climatological standards of reference. Decompositions of these skill scores are formulated, each of which is shown to possess terms involving 1) the coefficient of correlation between the forecasts and observations, 2) a measure of the nonsystematic (i.e., conditional) bias in the forecasts, and 3) a measure of the systematic (i.e., unconditional) bias in the forecasts. Depending on the choice of standard of reference, a particular decomposition may also contain terms relating to the degree of association between the reference forecasts and the observations. These decompositions yield analytical relationships between the respective skill scores and the correlation coefficient, document fundamental deficiencies in the correlation coefficient as a measure of performance, and provide additional insight into basic characteristics of forecasting performance. Samples of operational precipitation probability and minimum temperature forecasts are used to investigate the typical magnitudes of the terms in the decompositions. Some implications of the results for the practice of forecast verification are discussed.

1. Introduction

Skill scores are generally defined as measures of the relative accuracy of forecasts produced by two forecasting systems, one of which is a "reference system" (e.g., see Murphy and Daan 1985). Positive skill (i.e., a favorable difference in accuracy) is usually considered to represent a minimal level of acceptable performance for a set of forecasts. To the extent that the difficulty inherent in forecasting situations is reflected in the level of accuracy of the reference forecasts, skill scores also take difficulty into account. As a result, they can be used (with appropriate caveats) to compare forecasting performance across different locations or time periods. Thus, it is not surprising that skill scores are widely used in evaluating the performance of operational and experimental forecasts (e.g., see Dagostaro et al. 1988; Murphy and Daan 1985).

In the context of forecast verification, correlation coefficients are measures of the degree of linear asso- conjunction with model verification studies (e.g., see Arpe et al. 1985; Miyakoda et al. 1972; Sanders 1987).

Despite the rather widespread use of both skill scores and correlation coefficients, the relationships between these two common types of verification measures have evidently not been explored. In addition, little if any attention has been devoted to the problem of obtaining a quantitative appreciation of the deficiencies in the correlation coefficient as a measure of forecasting performance. The primary purpose of this paper is to describe decompositions of a family of climatological skill scores that yield insight into (i) the relationships between these measures and the (product moment) correlation coefficient and (ii) the deficiencies in the latter as a performance measure.

In section 2, we define the terms "accuracy" and "skill" and identify the basic measures of these attributes—namely, the mean-square-error m curacy and the mean-square-error skill ployed in this paper. This section also de

$$\mathrm{SS}(f, \bar{x}, x) = r_{fx}^2 - [r_{fx} - (s_f/s_x)]^2 - [(\bar{f} - \bar{x})/s_x]^2. \quad (12)$$

Forecasts $f_1, \ldots, f_N$.
Observations $o_1, \ldots, o_N$.

$$\mathrm{MSESS} = 1 - \frac{\mathrm{MSE}}{\mathrm{MSE}_{\mathrm{ref}}} = 1 - \frac{\sum_{n=1}^{N}(f_n - o_n)^2}{\sum_{n=1}^{N}(o_n - \bar{o})^2} = 1 - \frac{\mathrm{MSE}}{\hat{\sigma}_o^2}$$

For this reference forecast $\bar{o}$, the decomposition of MSESS is

$$\mathrm{MSESS} = \mathrm{AC}^2 - \left(\frac{\hat{\sigma}_f}{\hat{\sigma}_o} - \mathrm{AC}\right)^2 - \frac{(\bar{f} - \bar{o})^2}{\hat{\sigma}_o^2}$$

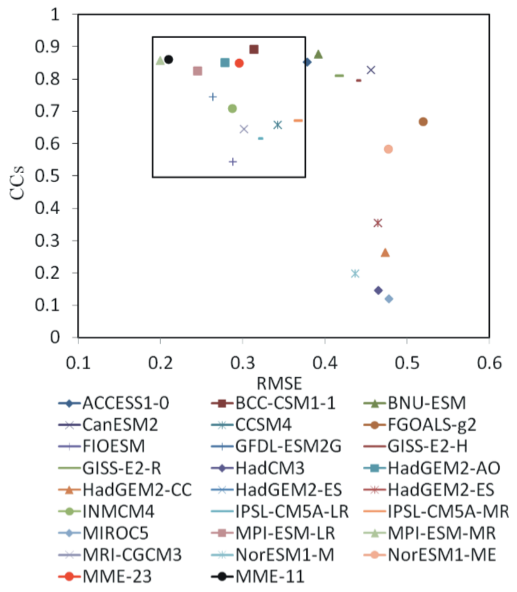
$$\mathrm{MSESS} = \text{Frac. explained var.} - \text{Cond. bias}^2 - \text{Bias}^2$$

$$BS = REL - RES + UNC$$

Each of these components can be decomposed further according to the number of p

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k(\mathbf{f_k} - \mathbf{\bar{o}_k})^2 - \frac{1}{N}\sum_{k=1}^{K} n_k(\mathbf{\bar{o}_k} - \mathbf{\bar{o}})^2 + \mathbf{\bar{o}}(1 - \mathbf{\bar{o}})$$

Application: to choose "good models"



Yajuan et al,2015

"The models with CCs greater than 0.5 and RMSEs less than 0.37 are selected to pro-duce the "best model ensemble"…"

Concluding Remark

- Correlation coefficient can be used for the probabilistic forecast.
- Point-biserial for naturally dichotomous (yes/No), Biserial for artificial dichotomous (threshold-based category).
- Statistical significance test is an advantage.
- No needs for any refence forecast (climatology) to create “skill score”.
- Simple to communicate to the user community.
- This score can use with Brier Skill Score to choose the “good model”.

Thank you!

Any Feedback?