

# Concordance Modeling With a Gold Standard for Variables From the Three-Parameter Gamma Distribution

Márcio Paulo de Oliveira (Corresponding Author)

Federal University of Technology – Paraná (UTFPR)

Rua Cristo Rei, 19, Toledo, Paraná 85902-490, Brazil

Tel: +55 (45) 3379-6800 E-mail: [marciooliveira@utfpr.edu.br](mailto:marciooliveira@utfpr.edu.br)

Miguel Angel Uribe-Opazo

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná 85819-170, Brazil

Tel: +55 (45) 3220-3000 E-mail: [miguel.opazo@unioeste.br](mailto:miguel.opazo@unioeste.br)

Manuel Galea

Pontificia Universidad Católica de Chile

Avda. Libertador Bernardo O'Higgins, 340, Santiago, Chile

Tel: (562) 2354 4000 E-mail: [mgalea@mat.uc.cl](mailto:mgalea@mat.uc.cl)

Jerry Adriani Johann

Western Paraná State University – (UNIOESTE)

Rua Universitária, 1619, Cascavel, Paraná 85819-170, Brazil

Tel: +55 (45) 3220-3000 E-mail: [jerry.johann@hotmail.com](mailto:jerry.johann@hotmail.com)

Received: Nov. 6, 2019

Accepted: Dec. 24, 2019

Published: Jan. 3, 2020

doi:10.5296/jas.v8i2.15761

URL: <https://doi.org/10.5296/jas.v8i2.15761>

## Abstract

A way to compare two or more measurements for the same random variable can be achieved by using a negligible error reference measurement, which is called the gold standard, obtained by consolidated measurement methods. This paper presents a new methodology for comparing measurements in the presence of a gold standard with random variables from the multivariate three-parameter (shape, scale, and location) gamma distribution. The errors between gold standard measures and approximate measures have a gamma difference distribution with the same three parameters of the gamma distribution. The concordance measurements were obtained by mean of a coefficient, which measures the degree of agreement as a ratio between the variances of the gold standard and the errors. The developed methodology is illustrated with climatic data which is divided into four ranges. The measurements analyzed are rainfall forecasts of the following four national centers: Canadian Meteorological Center (CMC), European Center for Medium-Range Weather Forecasts (ECMWF), National Centers for Environmental Prediction (NCEP), and Center for Weather Forecasting and Climate Studies (CPTEC). The forecast range was 240 hours for the West mesoregion of Paraná–Brazil, and in the October 1–March 31 period of the 2010/2011–2015/2016 harvest years. The period was selected because it is related to soybean crop development in the region and because several crop estimation models use rainfall forecast data in this period. The methodology applied spatially indicated the center to be selected in each geographical location according to each rainfall range interval. The gamma model fit well with the data and is an alternative to the normal one for modelling rainfall, in particular to estimate concordances between rainfall forecasts and the gold standard, which are used to improve the selection of rainfall forecast centers.

**Keywords:** coefficient of agreement, spatial data, parameter estimates, rainfall forecast, multivariate gamma distribution

## 1. Introduction

The three-parameter (shape, scale, and location) gamma probability distribution according to Johnson et al. (1994) has several applications in stochastic modeling and hydrology. The three-parameter  $\alpha$  (shape),  $\beta$  (scale), and  $\gamma$  (location) gamma probability density function is defined by Mathal and Moschopoulos (1992) as:

$$f(x; \alpha, \beta, \gamma) = \frac{1}{\beta^\alpha \Gamma(\alpha)} (x - \gamma)^{\alpha-1} \exp - \left\{ \left( \frac{x - \gamma}{\beta} \right) \right\}, \quad (1)$$

in which  $x > 0$ ,  $\alpha > 0$ ,  $\beta > 0$ ,  $\gamma \in \mathbb{R}$  and  $\Gamma(\alpha)$  is the gamma function. The function (1) is central in defining the various forms to the multivariate gamma distribution whose marginal are gamma distributions. The literature presents several particular application cases for the multivariate gamma distribution that include bivariate cases whose discussion can be found in McKay (1934), Cherian (1941), Jensen (1970), Royen (1991), Mathal and

Moschopoulos (1992), and references cited therein.

For evaluation of the degree of agreement (concordance) for measurements of a random variable with gamma distribution which were obtained by  $p$  approximation methods, one can use the standard model (Lord & Novick, 1968; Donner, 1986; Fleiss, 1999; Galea, 2013) of the reproducibility for measures (agreement) with respect to a reference measure, called gold standard,

$$y_{ij} = x_i + \varepsilon_{ij}, \quad (2)$$

in which,  $\varepsilon_{ij}$  is the random observation measurement error for the unity  $i = 1, \dots, n$  of the method  $j = 1, \dots, p$ ,  $y_{ij}$  is the measurement performed via the  $j$ -th method on the  $i$ -th unity,  $x_i$  is the gold-standard-based measurement on the  $i$ -th unity, with a mean of  $E[x_i] = \mu$  and variance of  $Var[x_i] = \phi$ ; considering  $\varepsilon_{ij}$  independent of  $x_i$ , with a mean of  $E[\varepsilon_{ij}] = 0$  and variance of  $Var[\varepsilon_{ij}] = \sigma_j^2$ .

The model (2) can be written in matrix notation (Laurent, 1998) as

$$\mathbf{Y}_i = x_i \mathbf{1}_p + \boldsymbol{\varepsilon}_i, \quad (3)$$

in which  $\mathbf{Y}_i = (y_{i1}, y_{i2}, \dots, y_{ip})^T$ ,  $\mathbf{1}_p$  is a vector of ones and  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{ip})^T$ . Let  $\mathbf{Z}_i = (x_i, \mathbf{Y}_i^T)^T$ , a vector  $q \times 1$ , with  $q = p + 1$ , of the measurements performed via gold standard and the approximation methods on the  $i$ -th unity.

## 2. Gamma Model Specification

Suppose that the random variables in the vector  $\mathbf{Z}_i$  are independent and identically distributed (iid) with gamma distribution, i.e.,  $x \sim Gama(\alpha_1, \beta, \gamma_1)$  and  $y_j \sim Gama(\alpha_j, \beta, \gamma_j)$ ,  $j = 2, \dots, q$ . Let  $R_1 = x$ ,  $R_2 = x + y_1$ , ... ,  $R_q = x + y_1 + \dots + y_p$ , with  $q = p + 1$ . The joint distribution of  $\mathbf{R} = (R_1, R_2, \dots, R_q)$  is a  $q$ -variate gamma

distribution Mathal and Moschopoulos (1992) defined by the density function given by

$$f_{\mathbf{R}}(r_1, r_2, \dots, r_q) = \frac{(r_1 - \gamma_1)^{\alpha_1 - 1} (r_2 - r_1 - \gamma_2)^{\alpha_2 - 1} \dots (r_q - r_{q-1} - \gamma_q)^{\alpha_q - 1}}{\beta^{\alpha_q^*} \prod_{i=1}^q \Gamma(\alpha_i)} \cdot \exp - \left\{ \left( \frac{r_q - \sum_{i=1}^q \gamma_i}{\beta} \right) \right\}, \quad (4)$$

in which  $r_i > 0$ ,  $\alpha_i > 0$ ,  $\beta > 0$ ,  $\gamma_i \in \mathbb{R}$ ,  $r_{i-1} + \gamma_i < r_i$ ,  $\gamma_i < r_i$ ,  $\alpha_q^* = \alpha_1 + \alpha_2 + \dots + \alpha_q$ ,

$r_0 = 0$ , for  $l = 1, \dots, q$  and zero elsewhere.

The multivariate gamma distribution given in (4) has several important properties (Mathal & Moschopoulos, 1992) some of them are:

i) The marginal distribution of  $R_l$  are three-parameter gamma with density function given in

(1), i.e.,  $R_l \sim \text{Gama}(\alpha_l^*, \beta, \gamma_l^*)$ , which  $\alpha_l^* = \alpha_1 + \dots + \alpha_l$  and  $\gamma_l^* = \gamma_1 + \dots + \gamma_l$

for  $l = 1, \dots, q$ .

ii) The mean and variance of  $R_l$  are, respectively, given by

$$E(R_l) = \beta \alpha_l^* + \gamma_l^*, \quad (5)$$

$$\text{Var}(R_l) = \beta^2 \alpha_l^*. \quad (6)$$

iii) The correlation matrix  $\mathbf{P}_{\mathbf{R}}$  of  $\mathbf{R}$  has a positive correlation between  $R_l$  and  $R_m$  and is given by

$$\text{Corr}(R_l, R_m) = \sqrt{\frac{\alpha_l^*}{\alpha_m^*}}, \quad (7)$$

in which  $l, m = 1, \dots, q$ .

iv) The covariance of  $R_l$  and  $R_m$ , for  $l < m$  is

$$\text{Cov}(R_l, R_m) = \text{Cov}(R_l, R_l + V_{l+1} + \dots + V_m) = \text{Var}(R_l) = \beta^2 \alpha_l^*, \quad (8)$$

v) The covariance matrix of the vector  $\mathbf{R}$  denoted by  $\Sigma_{\mathbf{R}}$ , is given by

$$\Sigma_{\mathbf{R}} = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 & \dots & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 & \dots & \sigma_1^2 + \sigma_2^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 & \dots & \sigma_1^2 + \dots + \sigma_q^2 \end{pmatrix}, \quad (9)$$

in which  $\sigma_l^2 = \alpha_l \beta^2$  and  $|\Sigma_{\mathbf{R}}| = \prod_{l=1}^q \alpha_l^2$ ,  $l = 1, \dots, q$ .

Suppose  $r_1 = x$ ,  $r_2 = x + y_1$ ,  $r_q = x + y_1 + \dots + y_p$ , then

$$f_{\mathbf{Z}}(x, y_1, \dots, y_p) = f_{\mathbf{R}}(r_1, r_2, \dots, r_q) |\mathbf{J}|, \quad (10)$$

in which  $\mathbf{J}$  is the Jacobean of the transformation of order  $q \times q$  given by

$$\mathbf{J} = \frac{\partial(r_1, r_2, \dots, r_q)}{\partial(x, y_1, \dots, y_p)} = \begin{pmatrix} \frac{\partial r_1}{\partial x} & \frac{\partial r_2}{\partial x} & \dots & \frac{\partial r_q}{\partial x} \\ \frac{\partial r_1}{\partial y_1} & \frac{\partial r_2}{\partial y_1} & \dots & \frac{\partial r_q}{\partial y_1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_1}{\partial y_p} & \frac{\partial r_2}{\partial y_p} & \dots & \frac{\partial r_q}{\partial y_p} \end{pmatrix},$$

then,

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

note that the  $|\mathbf{J}| = 1$ , thus one has

$$f_{\mathbf{Z}}(x, y_1, \dots, y_p) = f_{\mathbf{R}}(x, x + y_1, \dots, x + y_1 + \dots + y_p). \quad (11)$$

### 2.1 Parameter Estimation

Let  $f_{\mathbf{Z}}$  be the probability density function of  $\mathbf{Z}$ , which is given from (11) as

$$f_{\mathbf{Z}}(x, y_1, \dots, y_p) = \frac{(x - \gamma_1)^{\alpha_1 - 1} \exp - \left\{ \left( \frac{x - \gamma_1}{\beta} \right) \right\}}{\beta^{\alpha_1} \Gamma(\alpha_1)} \dots \frac{(y_p - \gamma_q)^{\alpha_q - 1} \exp - \left\{ \left( \frac{y_p - \gamma_q}{\beta} \right) \right\}}{\beta^{\alpha_q} \Gamma(\alpha_q)}. \quad (12)$$

Thus, the estimation of  $(2q + 1)$  parameters of the  $q$ -varied gamma distribution given in (12) can be performed by maximum product spacings method (Cheng & Amin, 1983).

Considering the ordered observations of the components of the vector  $\mathbf{Z}_i = (x_i, \mathbf{Y}_i^T)^T$ , i.e.,

$\mathbf{Z}_{i-1} < \mathbf{Z}_i$ , for  $i = 1, \dots, n + 1$ , in which  $\mathbf{Z}_0 = \mathbf{0}$ ,  $\mathbf{Z}_{n+1} = \infty$ . Let

$D_{li}(\theta_l) = F(z_{li}; \theta_l) - F(z_{li-1}; \theta_l)$ ,  $\theta_l = (\alpha_l, \beta, \gamma_l)$   $l = 1, \dots, q$ , where  $z_{1i} = x_i, \dots,$

$z_{qi} = y_{ip}$ , then

$$D_{li}(\theta_l) = \int_{z_{li-1}}^{z_{li}} \frac{(z_l - \gamma_l)^{\alpha_l - 1} \exp - \left\{ \left( \frac{z_l - \gamma_l}{\beta} \right) \right\}}{\beta^{\alpha_l} \Gamma(\alpha_l)}, \quad (13)$$

in which  $D_{11}(\theta_1) = F(z_{11}; \theta_1)$  and  $D_{ln+1}(\theta_l) = 1 - F(z_{ln}; \theta_l)$ , with  $\sum_{i=1}^{n+1} D_{li} = 1$ .

The maximum space product method requires maximization of the geometric mean of the spaces given by

$$G_l = \left( \prod_{i=1}^{n+1} D_{li} \right)^{1/(n+1)}, \quad (14)$$

or, equivalently, its logarithm  $H_l = \log G_l$ , in which

$$H_l = \log G_l = \frac{1}{n+1} \left( \sum_{i=1}^{n+1} \log D_{li} \right), \quad (15)$$

for  $l = 1, \dots, q$ . The maximum product of spacings estimator for  $\theta_l$  is the one which maximizes the logarithm of the geometric mean of sample spacings, i.e.,

$\hat{\theta}_l = \arg \max_{\theta_l \in \Theta} S_n(\theta_l)$  in which

$$S_{ln}(\theta_l) = \log \sqrt[n+1]{\prod_{i=1}^{n+1} D_{li}} = \log(F(z_{1n}; \theta_l)) + \sum_{i=2}^n \log(D_{li}(\theta_l)) + \log(1 - F(z_{1n}; \theta_l)). \quad (16)$$

The estimates of  $(2q + 1)$  parameters were obtained by maximizing the function given in (16), which is performed using the software R (R Core Team, 2019) via function *optim* from package *stats*, and with the method BFGS (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970).

## 2.2 Gamma Difference Model

Suppose that reference measurement  $x_i$  and approximate measurements  $\mathbf{Y}_i^T$  are independent, with gamma distributions, i.e.,  $x \sim \text{Gama}(\alpha_1, \beta, \gamma_1)$  and  $y_j \sim \text{Gama}(\alpha_j, \beta, \gamma_j)$ , in which  $j = 2, \dots, q$ . The errors  $\varepsilon_j$  are obtained via model (2)

and given by  $\varepsilon_j = y_j - x$ . The density function of  $\varepsilon_j$  (Mathal, 1993)  $g(\varepsilon_j)$  is given by

$$g(\varepsilon_{ij}) = \begin{cases} c_1 (\varepsilon_{ij} - (\gamma_1 + \gamma_j))^{(\alpha_1 + \alpha_j)/2 - 1} e^{-((\varepsilon_{ij} - (\gamma_1 + \gamma_j))/2)(1/\beta - 1/\beta)} \\ \quad \cdot W_{(\alpha_1 - \alpha_j)/2, (1 - \alpha_1 - \alpha_j)/2} (\beta_0 (\varepsilon_{ij} - (\gamma_1 + \gamma_j))) \\ \quad \text{for } (\varepsilon_{ij} - (\gamma_1 + \gamma_j)) > 0 \\ c_2 \left( -(\varepsilon_{ij} - (\gamma_1 + \gamma_j)) \right)^{(\alpha_1 + \alpha_j)/2 - 1} e^{-((\varepsilon_{ij} - (\gamma_1 + \gamma_j))/2)(1/\beta - 1/\beta)} \\ \quad \cdot W_{(\alpha_j - \alpha_1)/2, (1 - \alpha_1 - \alpha_j)/2} \left( -\beta_0 (\varepsilon_{ij} - (\gamma_1 + \gamma_j)) \right) \\ \quad \text{for } (\varepsilon_{ij} - (\gamma_1 + \gamma_j)) \leq 0, \end{cases} \quad (17)$$

in which  $\beta_0 = 2/\beta$ ,  $c_1$  and  $c_2$  are defined by

$$c_1^{-1} = \Gamma(\alpha_1)\beta^{(\alpha_1-\alpha_j)/2}\beta^{(\alpha_j-\alpha_1)/2}(\beta + \beta)^{(\alpha_1+\alpha_j)/2} \quad \text{and}$$

$c_2^{-1} = \Gamma(\alpha_2)\beta^{(\alpha_1-\alpha_j)/2}\beta^{(\alpha_j-\alpha_1)/2}(\beta + \beta)^{(\alpha_1+\alpha_j)/2}$ . The mean and variance of  $\varepsilon_j$  are, respectively, given by

$$E(\varepsilon_j) = \gamma_1 - \gamma_j + (\alpha_1 - \alpha_j)\beta, \quad (17)$$

$$\begin{aligned} \text{Var}(\varepsilon_j) &= (\gamma_1 - \gamma_j)^2 + (1 + \alpha_1)\alpha_1\beta^2 + (1 + \alpha_j)\alpha_j\beta^2 \\ &\quad + 2\alpha_1\beta(\gamma_1 - \gamma_j) - 2\alpha_j\beta(\gamma_1 - \gamma_j) - 2\alpha_1\alpha_j\beta^2 \\ &\quad - \gamma_1^2 - \gamma_j^2 + 2\gamma_1\gamma_j - 2\gamma_1(\alpha_1 - \alpha_j)\beta \\ &\quad + 2\gamma_j(\alpha_1 - \alpha_j)\beta - (\alpha_1 - \alpha_j)^2\beta^2. \end{aligned} \quad (18)$$

### 3. Coefficient of Agreement

The *ad hoc* coefficient (Lin, 1989; Laurent, 1998) used to evaluate the measure of agreement (concordance) between  $p$  approximate methods and the gold standard, which is widely used in the sciences (Galea, 2013) for evaluating the reproducibility of the measurements is given by

$$\rho_j = \phi / (\phi + \sigma_{jj}). \quad (19)$$

The estimator of (20) is

$$\hat{\rho}_j = \hat{\phi} / (\hat{\phi} + \hat{\sigma}_{jj}), \quad (20)$$

in which  $\hat{\phi} = \widehat{\text{Var}}(R_1)$  and  $\hat{\sigma}_{jj} = \widehat{\text{Var}}(\varepsilon_j)$ , for  $j = 1, \dots, p$ , according to the equations (6) and (18), respectively.



### 3.1 Confidence Interval

The confidence interval for  $\rho_j$  can be obtained via the bootstrapping pairs method according to Chernick and LaBudde (2011), using the following steps:

1. Resample with the replacement pairs of observations,  $b$  times, yielding vector

$$\mathbf{Z}_{bi} = (x_{bi}, \mathbf{Y}_{bi}^T)^T, \text{ in which } \mathbf{Y}_{bi} = (y_{bi1}, y_{bi2}, \dots, y_{bip})^T;$$

2. Generate the variables  $R_{b1}, R_{b2}, \dots, R_{bq}$  using convolutions;

3. Estimate the  $(2q + 1)_b$  parameters of the  $q$ -varied gamma distribution for  $\mathbf{Z}_{bi}$ ;

4. Compute estimates for the concordances  $\hat{\rho}_{bj}$ ;

5. Build up the vector  $\hat{\boldsymbol{\rho}}_{bj} = (\hat{\rho}_{b1}, \hat{\rho}_{b2}, \dots, \hat{\rho}_{bp})^T$ ;

6. Estimate the confidence intervals with  $\alpha_{sig}\%$  for  $\rho_j$ , i.e.,  $IC[\hat{\rho}_{j(\alpha_{sig}/2)}, \hat{\rho}_{j(1-\alpha_{sig}/2)}]$ , in

which  $b = 1, \dots, 1000$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$  and  $\hat{\rho}_{j(\alpha_{sig}/2)}$  is the empirical percentile

calculated with the bootstrap replicates from step (6).

## 4. Climate Data Application

For the agreement analysis with the multivariate gamma model, the dataset from TIGGE base, which is used in several studies (Aminyavari et al., 2018; Jha et al., 2018), were spatially stratified by selection of geographic coordinates (55 °W, 53 °W, 27 °S, 23 °S), corresponding to the rectangle containing the state of Paraná – Brazil (Figure 1). The pixels (P) have a dimension of 0.5 °x 0.5 °, with 84 pixels in total. Values from each pixel were obtained by database interpolation.

This study considered 75 meteorological stations from National Water Agency – Brazil (ANA) and 13 virtual stations indicated by the centroids of pixels in the mesoregion West of Paraná–Brazil (Figure 1). The correspondence between pixels (virtual stations) and ANA stations were obtained considering to each pixel a group of ANA stations with distance less than or equal to 0.36 °(about 40 km) from the centroid of pixels.

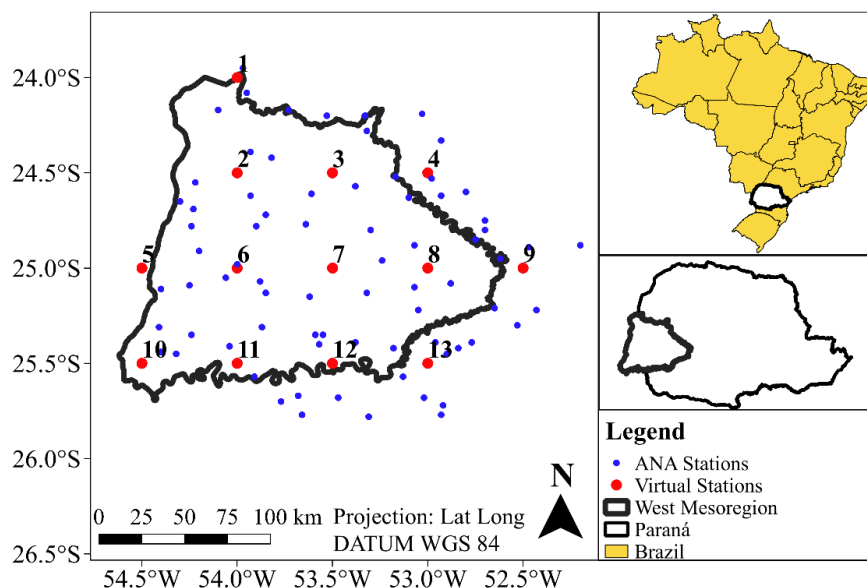


Figure 1. Location map of the West mesoregion of Paraná containing the ANA physical meteorological stations, virtual stations corresponding to CMC, ECMWF, NCEP and CPTEC

A temporal stratification was performed, selecting October 1–March 31 of the 2010/2011–2015/2016 harvest years as the temporal range. The range was selected because the state of Paraná is one of the largest soybean producers in Brazil, and the Agricultural Defense Agency of Paraná–Brazil (ADAPAR, 2018) establishes the proper period for sowing soybean crops of each agricultural year. Thus, the period was selected because it is related to soybean crop development in the region and because several crop estimation models use rainfall forecast data in this period. Agrometeorological variables directly influence crop yield estimation models (Battisti et al., 2018). In soybean culture, water availability is important, especially, during two development stages: germination-emergence and flowering-grain filling. (Rodrigues et al., 2017).

The rainfall forecast models of the TIGGE base used in this research were Canadian CMC, European ECMWF, North American NCEP and Brazilian CPTEC. The reference datasets, gold standard, were obtained of daily precipitation from ANA meteorological stations. Missing data were disregarded in the correspondences.

Descriptive statistics are presented in Table 1, for the data grouped in ten-day periods, corresponding to the 13 pixels in West mesoregion of Paraná–Brazil, and in the October 1–March 31 period of the 2010/2011–2015/2016 harvest years. The period includes soybean sowing in Paraná (Meotti et al., 2012; Bornhofen et al., 2015).

Several studies reveal the relation between spatial variability of the rainfall and the crop yield (Bezabih & Di Falco, 2012; Moraes et al., 2014; Jajoria et al., 2015). The crop yield estimation models are sensitive to variable precipitation (Cera et al., 2017). Missing ten-day periods were not found in ANA stations. The number of missing ten-day periods of centers are: 2 in CMC, 1 in ECMWF, 10 in NCEP, and 79 in CPTEC.

Table 1. Descriptive statistics of the 977 ten-day periods in the 13 pixels in West mesoregion of Paraná– Brazil, for the measurements of ANA stations (A), ten-day rainfall forecasts of centers CMC (B), ECMWF (C), NCEP (D) and CPTEC (E)

P	N	O	Min	Q1	Q2	Med	Q3	Max	P	N	O	Min	Q1	Q2	Med	Q3	Max
1	4	A	0.00	17.62	44.43	49.84	71.62	170.65	8	9	A	0.00	17.10	40.67	52.39	75.37	210.44
		B	0.00	28.59	52.75	58.59	80.89	213.31			B	0.00	29.28	53.78	61.32	84.38	261.62
		C	0.00	34.92	62.27	67.30	93.00	266.00			C	0.29	36.24	62.89	69.49	97.04	251.93
		D	0.00	34.67	59.96	67.12	91.86	240.50			D	0.00	29.21	55.94	62.86	89.04	257.42
		E	0.00	20.03	35.68	39.18	52.17	187.56			E	0.37	23.03	38.60	44.00	58.04	169.63
2	9	A	0.74	23.51	50.02	56.91	79.86	189.48	9	11	A	0.00	20.46	43.23	51.18	73.14	208.91
		B	0.00	28.56	57.12	63.25	88.00	229.75			B	0.00	28.50	51.75	60.03	83.44	229.62
		C	0.01	37.38	62.38	66.57	91.50	256.41			C	0.17	30.09	54.36	60.53	85.06	231.01
		D	0.00	31.72	55.68	64.70	88.64	291.92			D	0.00	29.62	55.02	64.37	90.66	270.89
		E	0.27	21.90	37.50	42.97	57.74	168.25			E	0.26	20.84	36.32	41.37	55.63	179.49
3	8	A	0.00	16.05	42.34	50.90	75.09	235.72	10	4	A	0.00	16.55	38.48	50.63	72.05	212.15
		B	0.00	29.26	57.00	62.00	83.88	252.88			B	0.12	31.00	56.03	59.50	82.09	241.50
		C	0.00	34.47	58.35	64.14	88.27	247.09			C	0.02	32.38	56.62	63.04	86.72	232.10
		D	0.00	32.45	56.26	64.17	87.94	264.53			D	0.24	33.59	57.03	64.93	87.84	231.06
		E	0.11	19.69	34.29	39.49	52.93	178.03			E	0.32	31.46	52.57	57.02	75.84	236.85
4	7	A	0.00	19.04	41.77	49.54	74.46	230.11	11	6	A	0.00	13.55	37.50	52.45	80.43	237.17
		B	0.00	28.25	52.75	60.74	83.19	276.00			B	0.12	32.50	57.88	61.65	84.56	265.25
		C	0.02	36.51	62.04	69.37	95.81	275.99			C	0.07	30.96	54.22	61.49	85.47	217.85
		D	0.00	31.34	55.6	63.63	87.33	272.42			D	0.28	34.14	58.88	66.41	88.11	222.40
		E	0.04	19.24	33.16	38.47	52.07	187.50			E	0.58	33.07	53.33	57.90	76.89	221.84
5	5	A	0.00	19.58	40.38	46.52	67.18	181.12	12	10	A	0.00	18.32	43.17	54.22	80.97	244.36
		B	0.00	29.68	54.88	58.74	81.67	240.94			B	0.12	31.64	56.62	59.15	81.22	230.38
		C	0.00	41.01	66.24	74.73	104.72	272.02			C	0.14	30.98	54.17	61.55	84.78	251.97
		D	0.00	31.60	56.15	65.77	89.54	255.41			D	0.28	34.41	58.27	66.68	89.37	253.69
		E	0.50	28.75	47.95	53.13	70.35	197.00			E	0.99	34.26	55.31	59.88	79.70	212.90
6	10	A	0.00	16.14	40.24	49.44	73.26	203.34	13	10	A	0.00	16.90	44.81	52.73	81.11	225.55
		B	0.00	28.75	50.75	56.23	78.75	206.00			B	0.12	32.09	56.38	59.55	81.09	223.75
		C	0.03	32.32	55.35	62.72	87.63	239.00			C	0.16	31.62	56.36	63.97	88.67	223.41
		D	0.00	30.68	55.28	66.04	92.69	298.62			D	0.40	39.93	65.24	74.81	98.55	315.34
		E	1.31	30.97	50.91	55.59	74.03	187.01			E	1.30	34.08	55.25	58.32	76.16	191.19
7	6	A	0.00	17.03	43.98	52.75	79.92	223.98			A	0.00	17.03	43.98	52.75	79.92	223.98
		B	0.00	30.38	53.44	59.57	84.66	277.75			B	0.00	30.38	53.44	59.57	84.66	277.75
		C	0.11	32.38	52.17	58.67	79.57	204.00			C	0.11	32.38	52.17	58.67	79.57	204.00
		D	0.00	30.26	58.34	64.99	90.48	280.92			D	0.00	30.26	58.34	64.99	90.48	280.92
		E	0.62	22.83	38.51	43.72	58.30	160.17			E	0.62	22.83	38.51	43.72	58.30	160.17

Note: P: Pixel number, N: number of ANA stations, O: Origin center, Min: Minimum value, Q1: First quartile, Q2: Second quartile, Med: Mean, Q3: Third quartile, Max: maximum value.

The value of 100 mm was used as an indicator of extreme events (Zandonadi et al., 2016). The following percentages of ten-day periods above 100 mm, according to the gold standard for each pixel, were computed (Figure 1). For pixel 1, 10,5%; pixel 2: 16%; pixel 3: 12,5%; pixel 4: 11%; pixel 5: 6,2%; pixel 6: 12,4%; pixel 7: 14,4%; pixel 8: 16,3%; pixel 9: 13,7%; pixel 10: 14,6%; pixel 11: 16%; pixel 12: 17,6%; and in pixel 13: 14,3%.

The highest incidence of extreme precipitation events was identified in pixel 12, and the pixel 1 was the least affected by these events. The coefficients of variation (CV) of the gold standard with the lowest value of 36.80% were obtained in pixel 7, and the highest value of

45.90% in the pixel 4. The CV indicated heterogeneity of values of rainfall in the stations. For the forecast centers, the CV with the lowest value of 55.60% was obtained for CPTEC in the pixel 13 and the highest value of 73.30% for NCEP in the pixel 6. The values of the CV indicated heterogeneity of the data from the ANA stations and the TIGGE base.

The water stress is the main cause for losses in the soybean culture (Confalone et al., 2010; Nunes et al., 2016; Souza et al., 2016). For non-irrigated areas, the water deficit caused, mainly, during drought periods can increase losses in the agricultural crops (Nunes et al., 2016; Pugh et al., 2019). The spatial variability of the soybean culture and the several agricultural crops occur according to the water availability (Iglesias et al., 2012; Vivan et al., 2013; Zanon et al., 2016). Thus, crop yield estimation models to increase the reliability should consider a climatic center, which rainfall forecasts are closer to the gold standard measures in a study region. The main source of water for an agricultural system comes from rainfall, which can be modeled using a gamma distribution (Sadiq, 2014; Cristaldo, 2017; Hasan et al., 2019).

The three-parameter multivariate gamma distribution given in (4) can be used for modelling a group of variables with gamma distribution. In order to compute the degree of agreement of  $j$ -th approximate measure with the gold standard,  $\hat{\rho}_j$  coefficient given by equation (21), it is required to obtain the variance of the difference of random variables with gamma distribution according to the model (Mathal, 1993) given in (17).

Descriptive statistics are presented in Table 2, for the agreements  $\hat{\rho}_j$ , in which  $j = 1, \dots, 4$  (centers) in the 13 pixels for each range of rainfall with data grouped in ten-day periods. The ranges were defined as follows: range 1 (0,00 |– 61,09 mm), range 2 (61,09 |– 122,18 mm), range 3 (122,18 |– 183,27 mm), and range 4 (183,27 |– 244,36 mm).

Table 2. Descriptive statistics of the agreements  $\hat{\rho}$  for the ranges 1-4 in the  $P = 13$  pixels in each center

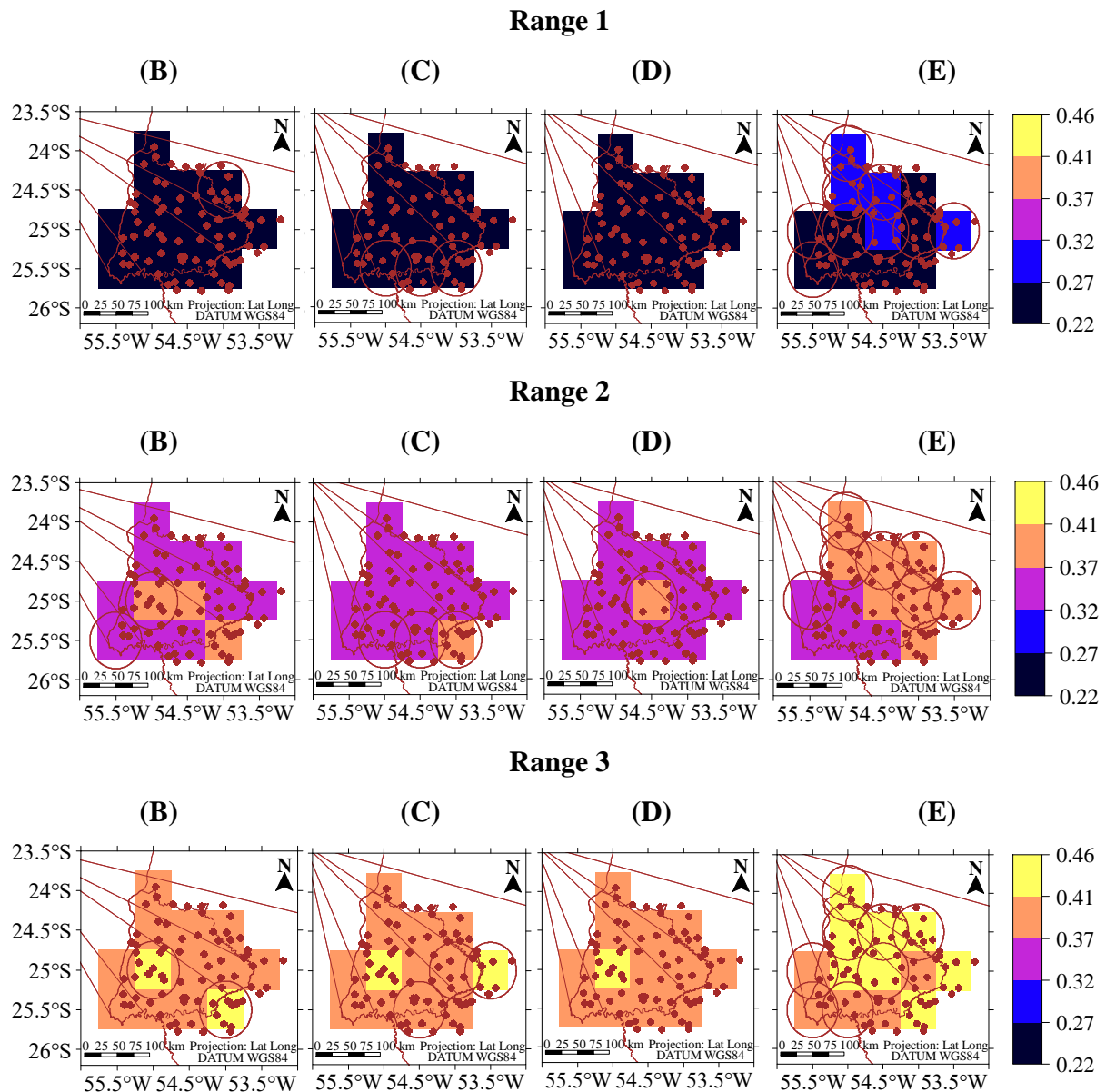
Ranges	Center	Min	Q1	Q2	Med	Q3	Max	SD	CV(%)
Range 1 0 + 61.09 mm	CMC (B)	0.233	0.246	0.250	0.250	0.256	0.265	0.009	3.571
	ECMWF (C)	0.224	0.244	0.248	0.247	0.253	0.264	0.011	4.294
	NCEP (D)	0.225	0.234	0.239	0.242	0.252	0.257	0.011	4.509
	CPTEC (E)	0.240	0.252	0.270	0.269	0.285	0.298	0.019	7.184
	CMC (B)	0.346	0.354	0.360	0.361	0.364	0.377	0.009	2.484
Range 2 61.09 + 122.18 mm	ECMWF (C)	0.333	0.353	0.356	0.356	0.360	0.382	0.012	3.308
	NCEP (D)	0.339	0.348	0.355	0.355	0.360	0.384	0.012	3.266
	CPTEC (E)	0.335	0.359	0.383	0.372	0.385	0.398	0.020	5.354
	CMC (B)	0.390	0.399	0.400	0.404	0.407	0.426	0.009	2.272
	ECMWF (C)	0.378	0.393	0.403	0.402	0.412	0.430	0.015	3.781
Range 3 122.18 + 183.27 mm	NCEP (D)	0.385	0.389	0.392	0.395	0.396	0.422	0.010	2.438
	CPTEC (E)	0.392	0.407	0.423	0.416	0.426	0.432	0.013	3.102
	CMC (B)	0.410	0.426	0.430	0.431	0.432	0.455	0.011	2.609
	ECMWF (C)	0.406	0.412	0.427	0.424	0.430	0.438	0.011	2.603
	NCEP (D)	0.407	0.422	0.430	0.428	0.434	0.445	0.012	2.696
Range 4 183.27 + 244.36 mm	CPTEC (E)	0.417	0.426	0.437	0.436	0.444	0.463	0.013	2.981

Note: Min: Minimum value, Q1: First quartile, Q2: Second quartile, Med: Mean, Q3: Third

quartile, Max: maximum value, SD: standard deviation, CV: coefficient of variation.

The spatial variability of the concordances, in the study area, is indicated for the ranges 1, 2, 3, and 4 in Figure 2. The difference of concordances and the distinct spatial variation for each range in the pixels were evident.

The comparison method with a gold standard, which aims to evaluate the ratio between the variances of the gold standard and the errors Lin (1989), Feng et al. (2015), and Chabert et al. (2019), was used for the ranges in each pixel.



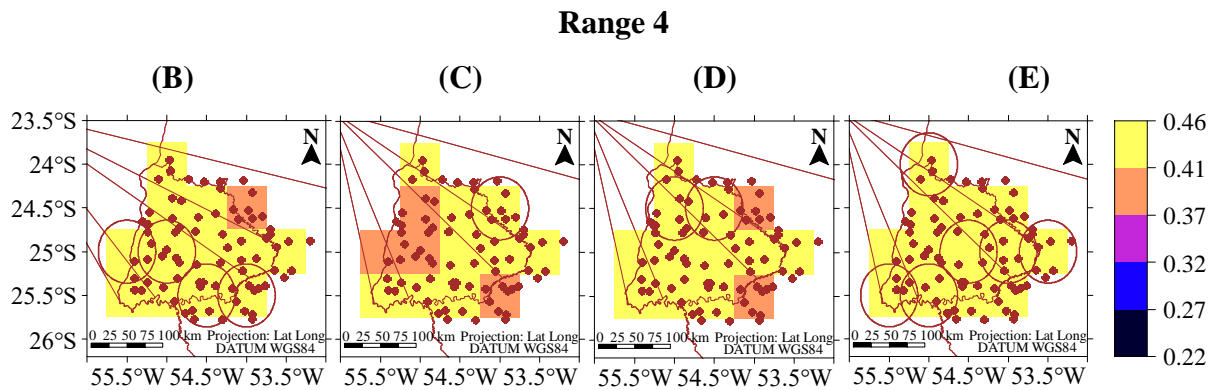


Figure 2. Concordances for the ranges 1-4 between gold standard (ANA) measures and the forecasts of the centers CMC (A), ECMWF (B), NCEP (C) and CPTEC (D) in each pixel of the West mesoregion of Paraná-Brazil. The circles indicate the centers with the highest agreement

The results presented in Figure 2 suggest that a calibration procedure must be applied before a precipitation forecast is used. According to Li et al. (2008), calibration procedures are required to remove bias and increase accuracy of spatial data. The selection of a forecast model should consider the highest value of concordance with gold standard (Harris et al., 2001; Barnhart et al., 2007).

The percentage of selected centers, which were indicated by circles in Figure 2, for each range were as follows: range 1: 7,7% of CMC, 23,1% of ECMWF, and 69,2% of the CPTEC; range 2: 15,4% of CMC, 23,1% of ECMWF, 7,7% of NCEP, and 53,8% of the CPTEC; range 3: 15,4% of CMC, 23,1% of ECMWF, and 61,5% of the CPTEC; and in range 4, 30,8% of CMC, 7,7% of ECMWF, 15,4% of NCEP, and 46,1% of the CPTEC.

For the concordances in each range, which were indicated by circles in Figure 2 95% confidence limits for  $\rho$  were estimated, using the bootstrapping pairs method according to Chernick and LaBudde (2011). The values of the 95% lower confidence limits (LCL) and 95% upper confidence limits (UCL) are presented in Table 3. Therefore, the selection of rainfall forecasts from the four CMC, ECMWF, NCEP, and CPTEC centers for use in yield estimation models should follow the Table 3 for crops such as soybean grown in the study period.

The qq-plots are presented in Figure 3 for the three-parameter gamma distribution, using the gold standard data grouped in ten-day periods and the corresponding 240 h range of the centers CMC, ECMWF, NCEP, and CPTEC. The three-parameter gamma distribution presented better data fit when compared to the normal distribution.

Table 3. Estimated concordances with lower and upper 95% confidence limits for the selected centers, which were indicated by circles in Figure 2, considering each range in the 13 pixels of the West mesoregion of Paraná-Brazil

Range 1									
P	SEL	$\hat{\rho}$	CL		P	SEL	$\hat{\rho}$	CL	
			LCL	UCL				LCL	UCL
1	CPTEC	0.298	0.238	0.391	8	CPTEC	0.270	0.219	0.361
2	CPTEC	0.289	0.233	0.371	9	CPTEC	0.293	0.235	0.388
3	CPTEC	0.282	0.225	0.375	10	CPTEC	0.251	0.209	0.317
4	CMC	0.259	0.212	0.336	11	ECMWF	0.255	0.202	0.336
5	CPTEC	0.272	0.220	0.353	12	ECMWF	0.253	0.208	0.313
6	CPTEC	0.267	0.221	0.332	13	ECMWF	0.252	0.202	0.315
7	CPTEC	0.285	0.235	0.359					

Range 2									
P	SEL	$\hat{\rho}$	CL		P	SEL	$\hat{\rho}$	CL	
			LCL	UCL				LCL	UCL
1	CPTEC	0.385	0.382	0.471	8	CPTEC	0.383	0.378	0.455
2	CPTEC	0.387	0.381	0.444	9	CPTEC	0.385	0.365	0.448
3	CPTEC	0.393	0.375	0.451	10	CMC	0.346	0.339	0.436
4	CPTEC	0.398	0.377	0.459	11	ECMWF	0.358	0.341	0.440
5	CPTEC	0.359	0.348	0.452	12	ECMWF	0.356	0.352	0.432
6	CMC	0.372	0.359	0.452	13	ECMWF	0.382	0.374	0.408
7	NCEP	0.384	0.379	0.440					

Range 3									
P	SEL	$\hat{\rho}$	CL		P	SEL	$\hat{\rho}$	CL	
			LCL	UCL				LCL	UCL
1	CPTEC	0.424	0.365	0.451	8	ECMWF	0.413	0.409	0.417
2	CPTEC	0.432	0.399	0.410	9	ECMWF	0.430	0.350	0.464
3	CPTEC	0.428	0.385	0.450	10	CPTEC	0.392	0.387	0.392
4	CPTEC	0.426	0.368	0.454	11	CPTEC	0.400	0.391	0.400
5	CPTEC	0.407	0.373	0.428	12	ECMWF	0.403	0.390	0.404
6	CMC	0.426	0.415	0.430	13	CMC	0.416	0.379	0.435
7	CPTEC	0.423	0.378	0.444					

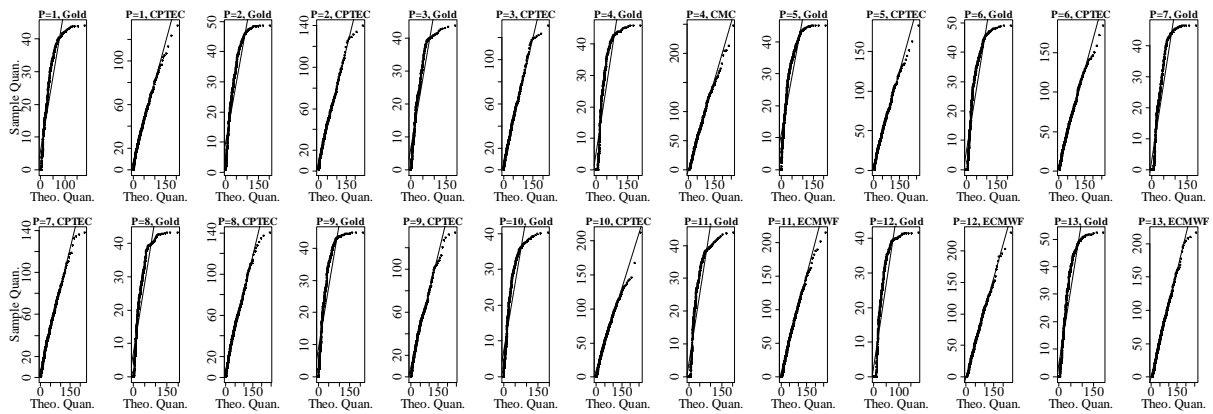
  

Range 4									
P	SEL	$\hat{\rho}$	CL		P	SEL	$\hat{\rho}$	CL	
			LCL	UCL				LCL	UCL
1	CPTEC	0.463	0.405	0.491	8	CPTEC	0.448	0.390	0.471
2	NCEP	0.445	0.371	0.446	9	CPTEC	0.444	0.399	0.464
3	NCEP	0.445	0.357	0.460	10	CPTEC	0.437	0.400	0.458
4	ECMWF	0.420	0.381	0.468	11	CPTEC	0.437	0.405	0.451
5	CMC	0.455	0.445	0.466	12	CMC	0.430	0.399	0.443
6	CMC	0.443	0.415	0.454	13	CMC	0.418	0.404	0.428
7	CPTEC	0.448	0.393	0.465					

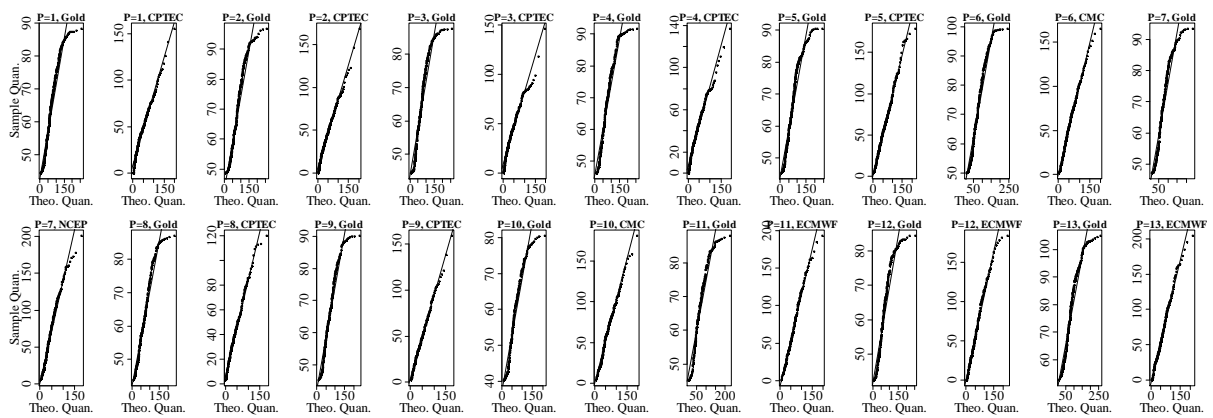
Note: P: Pixel number, SEL: Selected center,  $\hat{\rho}$ : Estimated concordance index, CL: 95% confidence limits for populational  $\rho$ , LCL: 95% lower confidence limit, UCL: 95% upper confidence limit.



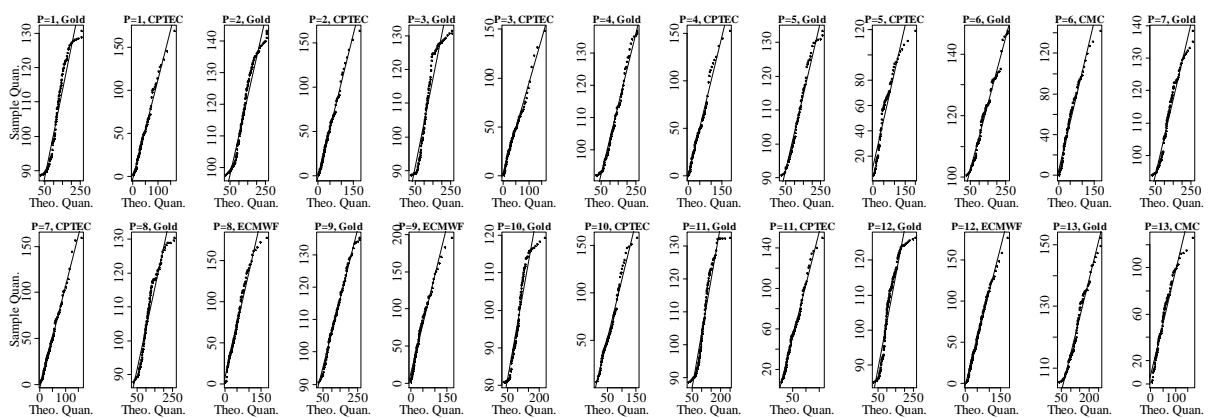
**Range 1**



**Range 2**



**Range 3**



**Range 4**

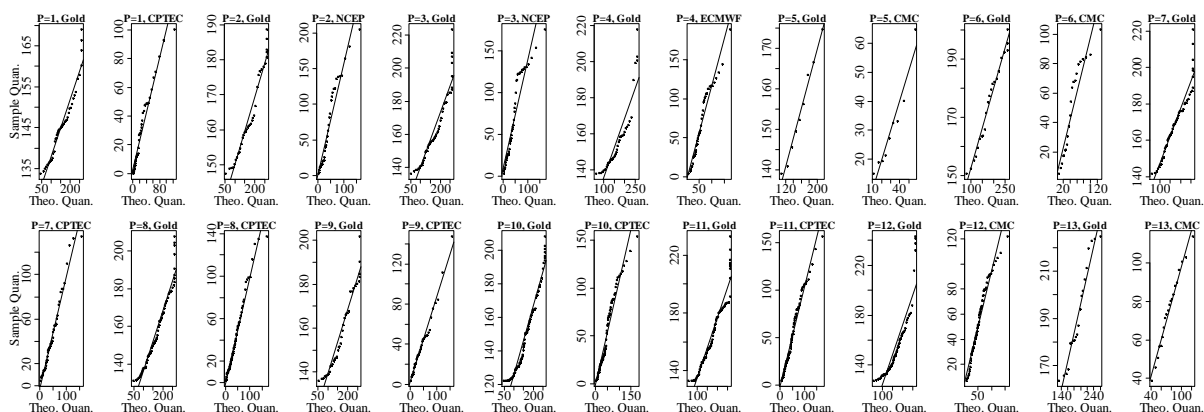


Figure 3. QQ-plots for the three-parameter gamma distribution, considering the gold standard and the selected centers for each range in the 13 pixels of the West mesoregion of Paraná

## 5. Conclusions

The estimated population variances for the gold standard measures and the errors of measurements, which were obtained with their respective distribution, allowed detecting spatial variability of the concordances in the study area. The detected variability was independent of forecast centers CMC, ECMWF, NCEP, and CPTEC. The geographical location and the range of precipitation should be considered when choosing a forecast center.

The estimated concordances using multivariate gamma distribution suggest that a calibration procedure, which aims to increase accuracy of spatial data, must be applied to the forecast data, before a precipitation forecast is used in a crop yield estimated model.

The crop yield estimates for soybean should use predictions from selected centers at locations within that pixel. For rainfall forecasts to be used in a yield estimation model of other crop cultures, a concordance analysis is required and should be applied according to crop development cycle to select the center.

The correspondence between the reference measurements, which are obtained from meteorological stations, and the climate model data from centers CMC, ECMWF, NCEP, and CPTEC requires that the geographical location and the precipitation range are respectively matched. A spatial correspondence of reference measures with climate model data can be obtained using the precipitation average value of the meteorological stations with distance from the centroid of pixel and which covers the pixel area. For the correspondence of precipitation range, the ten-day grouped period can be applied.

The confidence intervals for the concordances of selected centers CMC, ECMWF, NCEP or CPTEC indicated small variability for each precipitation range. The ranges 3 and 4 of precipitation with values between 122.18 and 244.36 mm, in general, presented the highest agreements with gold standard measures. The highest concordances in ranges 3 and 4 suggest that the forecast models used by centers CMC, ECMWF, NCEP, and CPTEC are more suitable for detection of extreme precipitation events, above 100 mm.

The multivariate gamma and the gamma difference distributions were used as an alternative to the normal distribution. The fitted gamma distribution for precipitation data from gold

standard and centers of forecast, and the fitted gamma difference distribution for the errors were more suitable as compared to normal distribution.

### Acknowledgements

The authors are grateful for the partial financial support from UTFPR, UNIOESTE / PGEAGRI, Coordination for the Improvement of Higher Level Personnel – Brazil (CAPES) – Finance Code 001, National Technological and Scientific Development (CNPq), and FONDECYT Chile (Project No. 1150325).

### References

- ADAPAR. Agência de Defesa Agropecuária do Paraná (2018). *Estabelece o período do vazão sanitário, as datas limites para semeadura e colheita da soja, e outras medidas para o controle da ferrugem asiática (Phakopsora pachyrhizi) no Estado do Paraná Portaria n. 264 de 17 de setembro de 2018*. Curitiba, Paraná Paraná
- Aminyavari, S., Saghafian, B., & Delavar, M. (2018). Evaluation of TIGGE Ensemble Forecasts of Precipitation in Distinct Climate Regions in Iran. *Advances in Atmospheric Sciences*, 35(4), 457-468. <https://doi.org/10.1007/s00376-017-7082-6>
- Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An Overview on Assessing Agreement with Continuous Measurements. *Journal of Biopharmaceutical Statistics*, 17(4), 529-569. <https://doi.org/10.1080/10543400701376480>
- Battisti, R., Sentelhas, P. C., & Boote, K. J. (2018). Sensitivity and requirement of improvements of four soybean crop simulation models for climate change studies in Southern Brazil. *International Journal of Biometeorology*, 62(5), 823-832. <https://doi.org/10.1007/s00484-017-1483-1>
- Bezabih, M., & Di Falco, S. (2012). Rainfall variability and food crop portfolio choice: evidence from Ethiopia. *Food Security*, 4(4), 557-567. <https://doi.org/10.1007/s12571-012-0219-7>
- Bornhofen, E., Benin, G., Galvan, D., & Flores, M. F. (2015). Épocas de semeadura e desempenho qualitativo de sementes de soja. *Pesquisa Agropecuária Tropical*, 45, 46-55. <https://doi.org/10.1590/1983-40632015v4529143>
- Broyden, C. G. (1970). The Convergence of a Class of Double-Rank Minimization Algorithms 1. General Considerations. *IMA Journal of Applied Mathematics*, 6(1), 76-90. <https://doi.org/10.1093/imamat/6.1.76>
- Cera, J. C., Streck, N. A., Fensterseifer, C. A. J., Ferraz, S. E. T., Bexaira, K. P., Silveira, W. B., & Cardoso, Â. P. (2017). Soybean yield in future climate scenarios for the state of Rio Grande do Sul, Brazil. *Pesquisa Agropecuária Brasileira*, 52, 380-392. <https://doi.org/10.1590/s0100-204x2017000600002>

- Chabert, A., Amossé A., & Sarthou, J. P. (2019). Assessing landscape composition using visual assessment: accuracy of rapid description compared to digital mapping. *Landscape Research*, 44(1), 6-18. <https://doi.org/10.1080/01426397.2017.1392493>
- Cheng, R. C. H., & Amin, N. A. K. (1983). Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *Journal of the Royal Statistical Society. Series B (Methodological)*, 45(3), 394-403. <https://doi.org/10.1111/j.2517-6161.1983.tb01268.x>
- Cherian, K. C. (1941). A Bi-Variate Correlated Gamma-Type Distribution Function. *Journal of the Indian Mathematical Society*, 5, 133-144.
- Chernick, M. R., & LaBudde, R. A. (2011). *An Introduction to Bootstrap Methods with Applications to R*. New Jersey: John Wiley & Sons, Inc.
- Confalone, A. E., Bernardes, M. S., Costa, L. C., Righi, C. A., Dourado Neto, D., Martin, T. N., & Pereira, C. R. (2010). Expolinear model on soybean growth in Argentina and Brazil. *Ciência Rural*, 40, 1009-1016. <https://doi.org/10.1590/S0103-84782010000500002>
- Cristaldo, M. F., Souza, C. C. de, Jesus, L. de, Padovani, C. R., Oliveira, P. T. S. de, & Viganó, H. H. da G. (2017). Analysis and Distribution of the Rainfall Monitoring Network in a Brazilian Pantanal Region. *Revista Brasileira de Meteorologia*, 32, 199-205. <https://doi.org/10.1590/0102-77863220007>
- Donner, A. (1986). A Review of Inference Procedures for the Intraclass Correlation Coefficient in the One-Way Random Effects Model. *International Statistical Review / Revue Internationale de Statistique*, 54(1), 67-82. <https://doi.org/10.2307/1403259>
- Feng, D., Baumgartner, R., & Svetnik, V. (2015). A Robust Bayesian Estimate of the Concordance Correlation Coefficient. *Journal of Biopharmaceutical Statistics*, 25(3), 490-507. <https://doi.org/10.1080/10543406.2014.920342>
- Fleiss, J. L. (1999). *The Design and Analysis of Clinical Experiments*. New York: Wiley. <https://doi.org/10.1002/9781118032923>
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3), 317-322. <https://doi.org/10.1093/comjnl/13.3.317>
- Galea, M. (2013). Comparación de métodos de medición en presencia de un gold standar. *Estadística*, 65(185), 1-17.
- Goldfarb, D. (1970). A Family of Variable-Metric Methods Derived by Variational Means. *Mathematics of Computation*, 24(109), 23-26. <https://doi.org/10.2307/2004873>
- Harris, I. R., Burch, B. D., & Laurent, R. T. St. (2001). A blended estimator for a measure of agreement with a gold standard. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(3), 326. <https://doi.org/10.1198/108571101317096541>
- Hasan, M. M., Croke, F. B., & Karim, F. (2019). Spatial and Seasonal Variations and Inter-Relationship in Fitted Model Parameters for Rainfall Totals across Australia at Various Timescales. *Climate*, 7(1). <https://doi.org/10.3390/cli7010004>

- Iglesias, A., Quiroga, S., Moneo, M., & Garrote, L. (2012). From climate change impacts to the development of adaptation strategies: Challenges for agriculture in Europe. *Climatic Change*, *112*(1), 143-168. <https://doi.org/10.1007/s10584-011-0344-x>
- Jajoria, D. K., Sharma, S. K., Narolia, G. P., & Dotaniya, M. L. (2015). Rainfall Variability: A Tool for Crop Planning of Udaipur Region of India. *National Academy Science Letters*, *38*(2), 95-98. <https://doi.org/10.1007/s40009-014-0305-9>
- Jensen, D. R. (1970). The Joint Distribution of Quadratic Forms and Related Distributions. *Australian Journal of Statistics*, *12*(1), 13-22. <https://doi.org/10.1111/j.1467-842X.1970.tb00108.x>
- Jha, S. K., Shrestha, D. L., Stadnyk, T. A., & Coulibaly, P. (2018). Evaluation of ensemble precipitation forecasts generated through post-processing in a Canadian catchment. *Hydrol. Earth Syst. Sci.*, *22*(3), 1957-1969. <https://doi.org/10.5194/hess-22-1957-2018>
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous Univariate Distributions*. New York: Wiley.
- Laurent, R. T. St. (1998). Evaluating Agreement with a Gold Standard in Method Comparison Studies. *Biometrics*, *54*(2), 537-545. <https://doi.org/10.2307/3109761>
- Li, B., Eriksson, M., Srinivasan, R., & Sherman, M. (2008). A geostatistical method for Texas NexRad data calibration. *Environmetrics*, *19*(1), 1-19. <https://doi.org/10.1002/env.848>
- Lin, L. I. K. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics*, *45*(1), 255-268. <https://doi.org/10.2307/2532051>
- Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Massachusetts: Addison-Wesley.
- Mathal, A. M. (1993). On Noncentral Generalized Laplacianess of Quadratic Forms in Normal Variables. *Journal of Multivariate Analysis*, *45*(2), 239-246. <https://doi.org/10.1006/jmva.1993.1036>
- Mathal, A. M., & Moschopoulos, P. G. (1992). A form of multivariate gamma distribution. *Annals of the Institute of Statistical Mathematics*, *44*(1), 97-106. <https://doi.org/10.1007/BF00048672>
- McKay, A. T. (1934). Sampling from Batches. *Supplement to the Journal of the Royal Statistical Society*, *1*(2), 207-216. <https://doi.org/10.2307/2983603>
- Meotti, G. V., Benin, G., Silva, R. R., Beche, E., & Munaro, L. B. (2012). Épocas de semeadura e desempenho agrônômico de cultivares de soja. *Pesquisa Agropecuária Brasileira*, *47*, 14-21. <https://doi.org/10.1590/S0100-204X2012000100003>
- Moraes, R. A., Rocha, J. V., & Lamparelli, R. A. C. (2014). Determination of total accumulated rainfall, global radiation, evapotranspiration and degree-days originated from the ECMWF model to sugar cane crop. *Engenharia Agrícola*, *34*, 322-331. <https://doi.org/10.1590/S0100-69162014000200013>

Nunes, M. S., Robaina, A. D., Peiter, M. X., Braga, F. de V. A., Pereira, T. dos S., & Buske, T. C. (2016). Resposta da produção de soja à variabilidade espacial sob pivô central. *Brazilian Journal of Irrigation and Drainage*, 19-27. <https://doi.org/10.15809/irriga.2016v1n1p19-27>

Pugh, S., Heaton, M. J., Svedin, J., & Hansen, N. (2019). Spatiotemporal Lagged Models for Variable Rate Irrigation in Agriculture. *Journal of Agricultural, Biological and Environmental Statistics*, 1-17. <https://doi.org/10.1007/s13253-019-00365-3>

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <https://www.r-project.org/>

Rodrigues, T. R., Casaroli, D., Evangelista, A. W. P., & Alves Júnior, J. (2017). Water availability to soybean crop as a function of the least limiting water range and evapotranspiration. *Pesquisa Agropecuária Tropical*, 47, 161-167. <https://doi.org/10.1590/1983-40632016v4743746>

Royen, T. (1991). Expansions for the multivariate chi-square distribution. *Journal of Multivariate Analysis*, 38(2), 213-232. [https://doi.org/10.1016/0047-259X\(91\)90041-Y](https://doi.org/10.1016/0047-259X(91)90041-Y)

Sadiq, N. (2014). Stochastic Modelling of the Daily Rainfall Frequency and Amount. *Arabian Journal for Science and Engineering*, 39(7), 5691-5702. <https://doi.org/10.1007/s13369-014-1132-5>

Shanno, D. (1970). Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24(111), 647-656. <https://doi.org/10.2307/2004840>

Souza, P. J. de O. P., Ortega-Farias, S., Rocha, E. J. P. da, Sousa, A. M. L. de, & Souza, E. B. de. (2016). Consumo hídrico da soja no nordeste paraense. *Brazilian Journal of Irrigation and Drainage*, xx, 218-231. <https://doi.org/10.15809/irriga.2016v1n01p218-231>

Vivan, G. A., Peiter, M. X., Robaina, A. D., Barboza, F. da S., & Buske, T. (2013). Rendimento relativo da cultura da soja em função da lâmina de irrigação. *Brazilian Journal of Irrigation and Drainage*, 18(2), 282-292. <https://doi.org/10.15809/irriga.2013v18n2p282>

Zandonadi, L., Acquattrota, F., Fratianni, S., & Zavattini, J. A. (2016). Changes in precipitation extremes in Brazil (Paraná River Basin). *Theoretical and Applied Climatology*, 123(3), 741-756. <https://doi.org/10.1007/s00704-015-1391-4>

Zanon, A. J., Streck, N. A., & Grassini, P. (2016). Climate and Management Factors Influence Soybean Yield Potential in a Subtropical Environment. *Agronomy Journal*, 108, 1447-1454. <https://doi.org/10.2134/agronj2015.0535>

### **Copyright Disclaimer**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).