Proceedings of 8 ICSHMO, Foz do Iguaçu, Brazil, April 24-28, 2006, INPE, p. 575-582.

BETTER STATISTICS TO ASSESS THE QUALITY OF ANALOGUE-BASED FORECAST SYSTEMS

A. H. N. Maia [1], H. Meinke [2] and W. Baethgen [3]

[1] *Embrapa Meio Ambiente, PO Box 69, Jaguariúna, SP, Brazil; ahmaia@cnpma.embrapa.br;*
[2] *Department of Primary Industries and Fisheries, PO Box 102, Toowoomba, Qld 4350, Australia, holger.meinke@dpi.qld.gov.au;*
[3] *IRI, Columbia University, PO Box 1000, Palisades, NY 10964-8000, USA, baethgen@iri.columbia.edu*

Introduction

Seasonal probabilistic forecast systems (SPFS) based on the analogue years approach (AYA) are used worldwide and provide valuable information for decision makers managing climate-sensitive systems (Sivakumar et al. 2000; Ferreyra et al. 2001; Selvaraju et al. 2004; Meinke and Stone 2005). Providing such categorisations are based on scientifically well understood mechanisms, such forecasts (or, more appropriately, scenarios) allow climate time series to be partitioned into 'year- or season-types' (analogue years) based on prevailing ocean and atmospheric conditions (i.e. Southern Oscillation Index, SOI and/or Sea Surface Temperatures SST anomalies), resulting in SOI or ENSO phases.

These time series are usually represented by their respective cumulative distribution functions (CDFs) or their complement, probability of exceeding functions (POEs): a conditional CDF_K for each class K and an unconditional CDF (CDF_{ALL}). Current oceanic and atmospheric conditions can then be assigned to a particular category K and the correspondent CDF_K is then adopted for probabilistic assessments.

To take action, decisions makers need to know: a) whether or not probabilistic forecasts provided by conditional distributions are sufficiently different from their respective from \`climatology'; b) if so, what is the magnitude of change in the prognostic variable that might lead to a change in the decision; c) is there sufficient improvement in accuracy over the 'climatology' and d) if so, what is the improvement in accuracy of this forecast over the unconditional case (Maia et al. 2006). From a methodological perspective, the assessment of questions (a) and (c) requires inferential tools such as statistical tests for the hypothesis of \`no class effect´. The assessment of questions (b) and (d) requires intuitive, descriptive statistics that are relevant for the question at hand. We propose using descriptive measures coupled with inferential methods to evaluate such SPFS. Detailed discussion about forecast quality assessments can be found in Potgieter et al. (2004).

We illustrate these approaches by quantifying signal of a SOI-based forecast system across Australia and an ENSO-based forecast system across Southeast of South America.

Data & Methods

Rainfall data

For illustration purposes, we used 3-monthly rainfall data from: a) Australia (June-August – JJA, 64 stations, all rainfall series with a record length of 103 years) and b) Southeast of the South America (October-December – OND, 60 stations, series length ranging from 57 to 87 years).

Analogue-based seasonal forecast systems

The seasonal probabilistic forecast systems evaluated were: a 3-phase forecast system based on El Niño/Southern Oscillation (ENSO) for the Southeast of South America (Ropelewski and Halpert, 1987) and a 5-phase forecast system based on the Southern Oscillation Index for Australia (Stone et al. 1996).

Descriptive quality assessments

Descriptive assessments allow exploring characteristics of the class effects over time and space. We are proposing measures that account for divergences among the entire CDFs of the decision variables (Table 1) or their percentiles such as median values (Table 2); equivalent measures could be calculated for any other statistic of interest, e.g. the 95^{th} percentile, probability of exceeding a critical value etc). Descriptive measures indicate how much the "statistic of interest" changes due to the FS class information. In other words, these measures quantify how much the 'conditional climatology' derived from the forecast system differs from the 'unconditional climatology' (i.e. the CDF derived from the entire climate record).

Table 1. Descriptive measures used to quantify divergences between conditional and unconditional probability of exceeding functions at each station.

Measure	Description
POE_k	Probability of exceeding function representing observed values of decision variable (Y) time series corresponding to class k of the forecast system adopted, also referred to as conditional POE;
POE_{ALL}	Probability of exceeding function representing observed values of decision variable (Y) time series corresponding to "climatology", also referred to as unconditional POE;
$VDist_{yk}$	Vertical distance between POE_K and POE_{ALL} at the position Y=y;
$AbsVDist_{yk}$	Absolute vertical distance between POE_K and POE_{ALL} at the position Y=y;
MaxAbsVDist	Maximum absolute vertical distance between POE_K and POE_{ALL} at the position Y=y (over all y and k);
MaxVDist	Maximum absolute vertical distance between POE_K and POE_{ALL} at the position Y=y, considering the signal (+ or -)
MaxRelAbsDiff	MaxAbsVDist expressed as percent of POE_{ALL} at the respective position Y=y (%);
MaxRelDiff	MaxVDiff expressed as percent of POE_{ALL} at the respective position Y=y (%).

Table 2. Descriptive measures used to quantify divergences between conditional and unconditional medians* at each station.

Measure	Description
$Median_k$	Median of the decision variable CDF corresponding to class K of the forecast system adopted, also referred to as conditional median (mm);
$Median_{ALL}$	Median of the the decision variable CDF corresponding to "climatology", also referred to as unconditional median (mm);
$Diff_K$	Difference between $Median_K$ and $Median_{ALL}$ (mm);
$AbsDiff_K$	Absolute difference between $Median_K$ and $Median_{ALL}$ (mm);
MaxAbsDiff	Maximum absolute difference between $Median_K$ and $Median_{ALL}$ (mm);
MaxDiff	maximum absolute difference between $Median_K$ and MedianAll, considering the signal (+ or -) (mm);
MaxRelAbsDiff	MaxAbsDiff expressed as percent of MedianAll (%);
MaxRelDiff	MaxDiff expressed as percent of MedianAll (%);

* Similar measures can be calculated for any other percentile.

Inferential quality assessments

For inferential analysis, we suggest to use distribution free statistical tests related to the descriptive measures adopted: for example, multisample Kruskal-Wallis or Median test can be used in conjunction with MaxDifMed, while the multisample Kolmogorov-Smirnov or Log-rank test can be used in conjunction with MaxVDist. Alternatively, Monte Carlo methods (e.g. randomised tests) can also be applied to develop specific inferential procedures for any descriptive measure of interest (Conover 1990; Manly, 1991).

Here we used the multisample Kruskal-Wallis test (Kruskal and Wallis, 1952) to quantify evidences against 'no class effect' on 3-monthly rainfall medians across Australia (SOI FS) and Southeast of South America (ENSO FS). Tests were applied to each location and corresponding nominal significance levels (p-values) were mapped to display spatial patterns of the signal of the underlying climate phenomena. For detailed discussion about the inferential assessments proposed here, see Maia et al. (2006).

Results & Discussion

Quantifying SOI signal over Australia

For case studies of detailed descriptive analysis, stations were selected accordingly to three different criteria: MaxDifMed, MaxRelDiff and MaxVDist. For each criteria stations corresponding to minimum (maximum negative), minimum absolute (values closest to zero), and maximum (maximum positive) were chosen, aiming to represent variability of SOI influence over the region. Depending on the criteria, different sets of stations were selected. Descriptive measures of divergence between conditional and unconditional medians and POEs at those locations are shown in Tables 3-5.

High relative differences frequently occur at seasonally dry stations (Figure 1-A2). As in the example showed below (Table 5) and in spite of the high relative values, the evidence of class effect on medians at those locations is very low (KW p-values were 0.92 and 0.32, for Ayr and Burketown, respectively). On the other hand, low relative differences can correspond to high evidence of class effect (e.g. Cape Leeuwin, KW p-value=0.0004). Spatial patterns of MaxAbsDiff, MaxRelAbsDiff, MaxAbsVDist and MaxRelAbsVDist over the region are shown in Figure 1. JJA rainfall probability of exceeding functions for stations selected accordingly MaxAbsDiff criteria (Bundaberg, Darwin and Albany) are shown in Figure 2. Vertical dashed lines highlight distances between conditional and unconditional POEs while horizontal lines highlight distances between medians.

As example of inferential quality assessment, pattern of SOI signal across Australia, as measured by KW p-values is shown in Figure 3. That map displays spatial variability of SOI classes influence on JJA rainfall medians.

Table 3. Descriptive measures of divergence between conditional and unconditional POEs at locations selected accordingly to maximum absolute vertical distance (MaxAbsVDist) criteria.

Location	POEAll At Maximum	Class POE at maximum Vertical Distance*					Maximum Vertical Distance		Signal
		1	2	3	4	5	Absolute	Relative (%)	
Albany	0.69	0.31	0.82	0.79	0.81	0.64	0.38	0.55	-
Old Halls Creek	0.25	0.19	0.14	0.36	0.23	0.36	0.12	0.46	-
Cape Leeuwin	0.45	0.25	0.23	0.14	0.81	0.56	0.36	0.81	+

*Values in red correspond to classes at which the maximum vertical distances were observed.

Table 4. Descriptive measures of divergence between conditional and unconditional JJA rainfall POEs at locations selected accordingly to maximum absolute difference (MaxAbsDiff) criteria.

Location	MedianAll (mm)	Class Median (mm)					Maximum difference between medians		Signal
		1	2	3	4	5	Absolute (mm)	Relative (%)	
Bundaberg	117.2	58.4	58.8	130.4	120.9	147.1	58.8	50.2	-
Darwin	0.7	2.0	1.3	1.0	1.5	0.0	1.3	185.7	+
Albany	395.6	342.0	66.5	395.3	461.1	399.6	65.5	16.8	+

*Values in red correspond to classes at which the maximum differences were observed.

Table 5. Descriptive measures of divergence between conditional and unconditional JJA rainfall POEs at locations selected accordingly to maximum relative absolute difference (MaxRelAbsDiff) criteria.

Location	MedianAll (mm)	Class Median (mm)					Maximum difference between medians		Signal
		1	2	3	4	5	Absolute (mm)	Relative (%)	
Ayr	11.3	3.6	15.0	14.8	9.1	6.9	7.7	68.1	-
Cape Leeuwin	511.7	472.9	475.4	504.5	532.2	546.6	38.8	7.6	-
Burketown	0.4	1.9	0.9	0.00	1.80	0.00	1.5	387.5	+

*Values in red correspond to classes at which the maximum differences were observed.

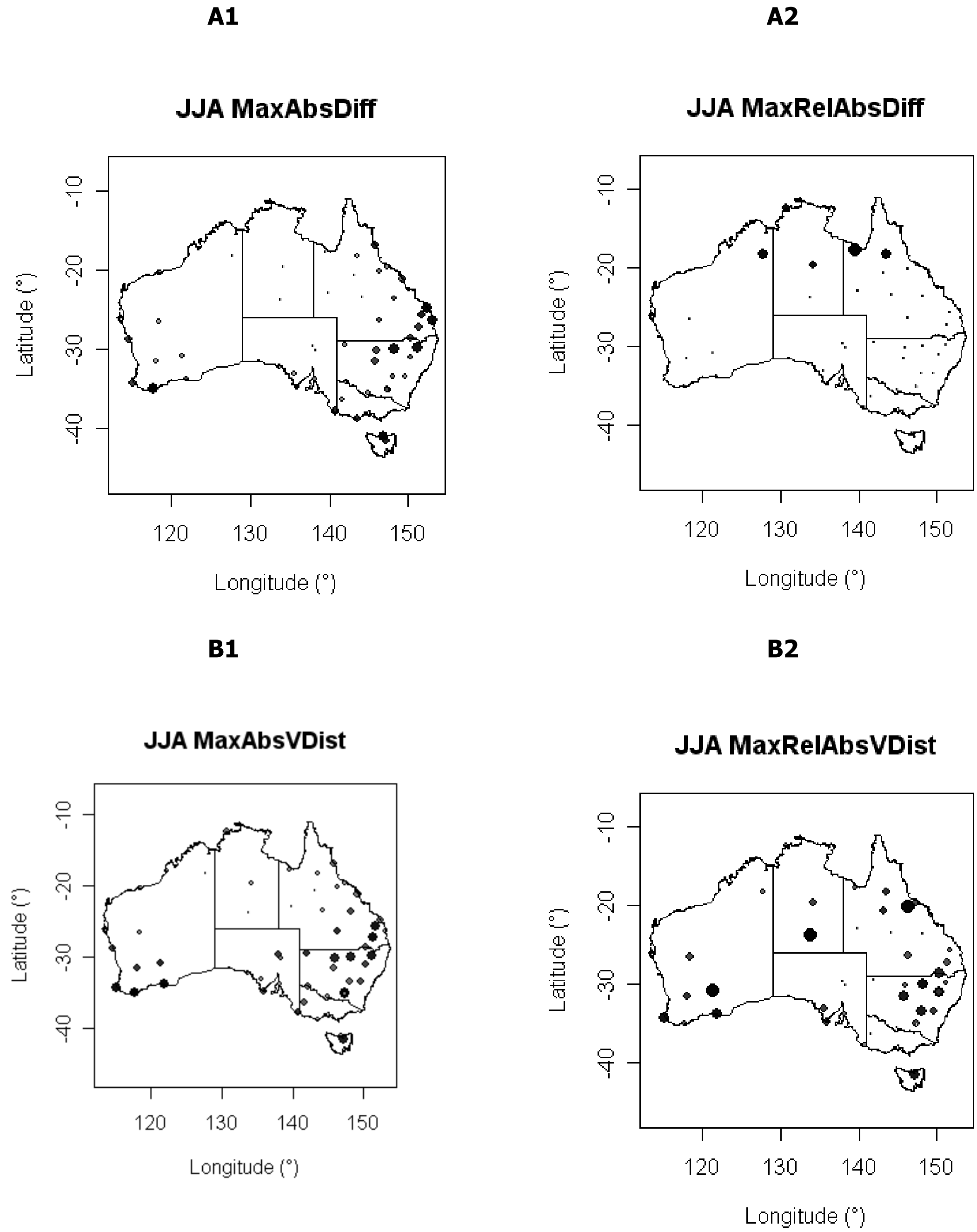

Figure 1. Spatial patterns of descriptive measures for quantifying SOI class influence on JJA rainfall POEs. A) Maximum absolute (MaxAbsDiff, A1) and maximum relative absolute difference (MaxRelAbsDiff, A2) among conditional and unconditional median and B) Maximum absolute (MaxAbsVDist, B1) and maximum relative absolute vertical distance (MaxRelAbsVDist, B2) among conditional and unconditional POEs. Point size and color intensity proportional to data values.

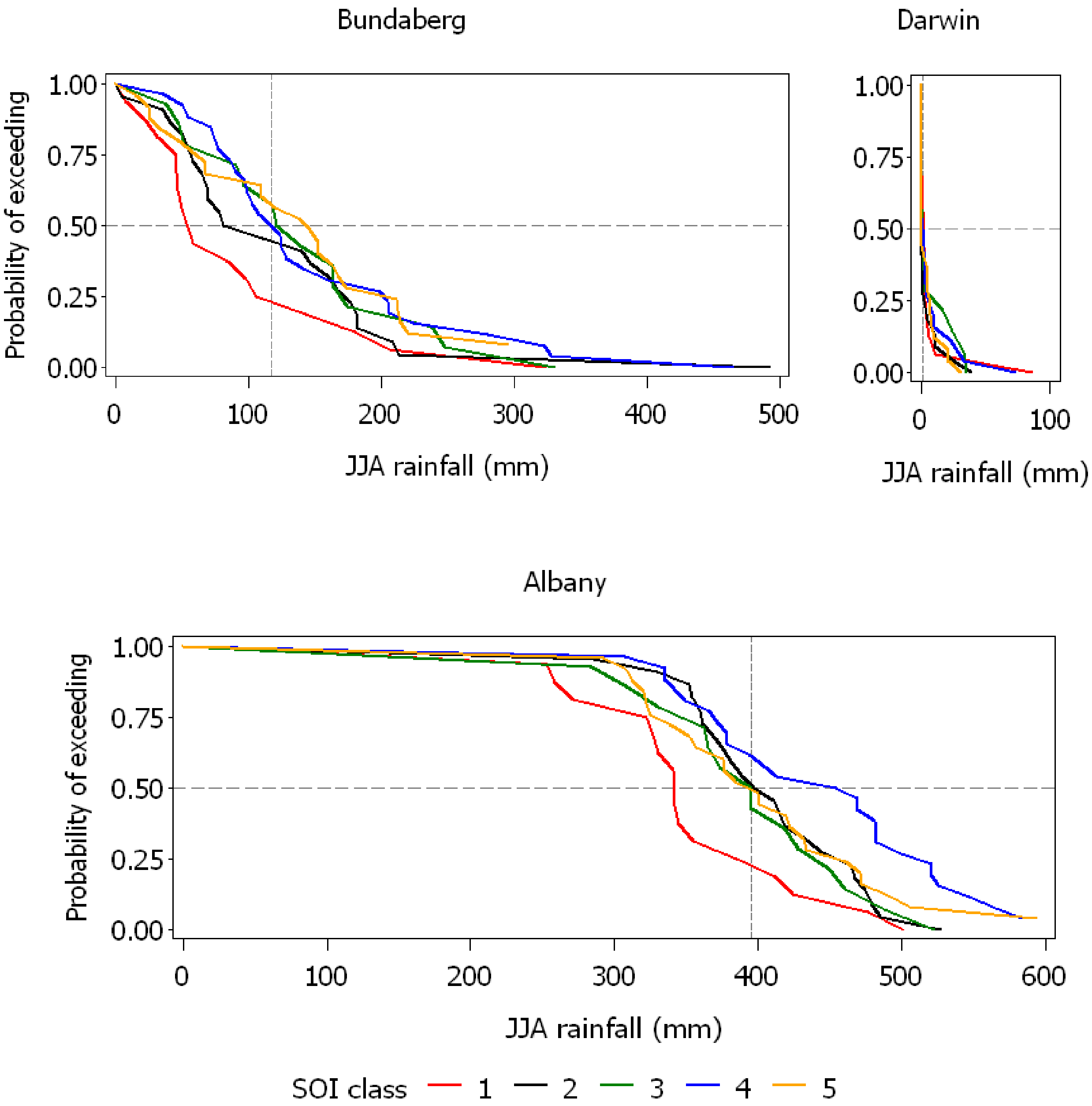

Figure 2. Probability of exceeding functions representing JJA rainfall series by SOI class at Bundaberg, Darwin and Albany. Stations correspond to the maximum negative, nearest zero and maximum positive observed MaxDiff, respectively, selected from 64 high quality stations across Australia. Vertical and horizontal dashed lines correspond to unconditional median and probability of exceeding unconditional median, respectively.

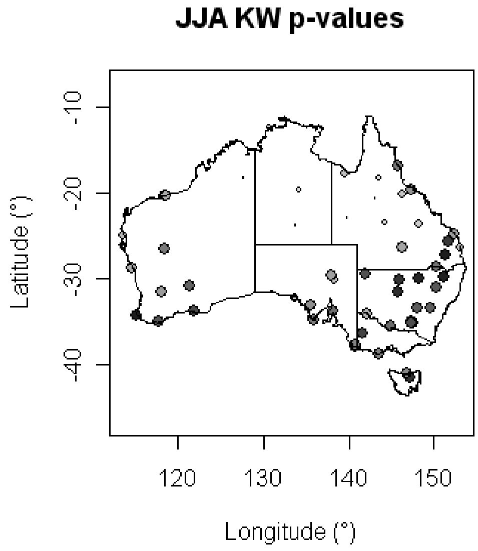

Figure 3. Example of inferential quality assessment: spatial patterns of Kruskal-Wallis nominal significance levels (p-values), used to quantify evidences of SOI classes influence on JJA rainfall medians. Data from 64 stations across Australia.

Quantifying ENSO signal across the Southeast of South America

Stations corresponding to minimum (maximum negative), minimum absolute (values closest to zero), and maximum (maximum positive) MaxDifMed, São Luiz (Brazil), Frias (Argentina) and Bella Union (Uruguay), respectively, were selected for detailed analysis (Table 6).

As expected for the ENSO FS, positive differences were observed for El Niño and negative differences at La Niña classes. In contrast to Australia (Figure1-A1 and Figure1-A2), where seasonally dry stations are frequent, spatial patterns of MaxRelDiff and MaxRelAbsDiff over the region followed similar patterns (Figure 4). Figure 5 shows OND rainfall POEs corresponding to each ENSO class at three selected locations.

As an example of an inferential approach for analogue-based FS quality assessment, we used the Kruskal-Wallis test (KW) for quantifying evidences of ENSO class effects on OND medians. KW p-values were mapped (60 stations) for displaying spatial patterns of ENSO signal for the selected period over Southeast of South America (Figure 6).

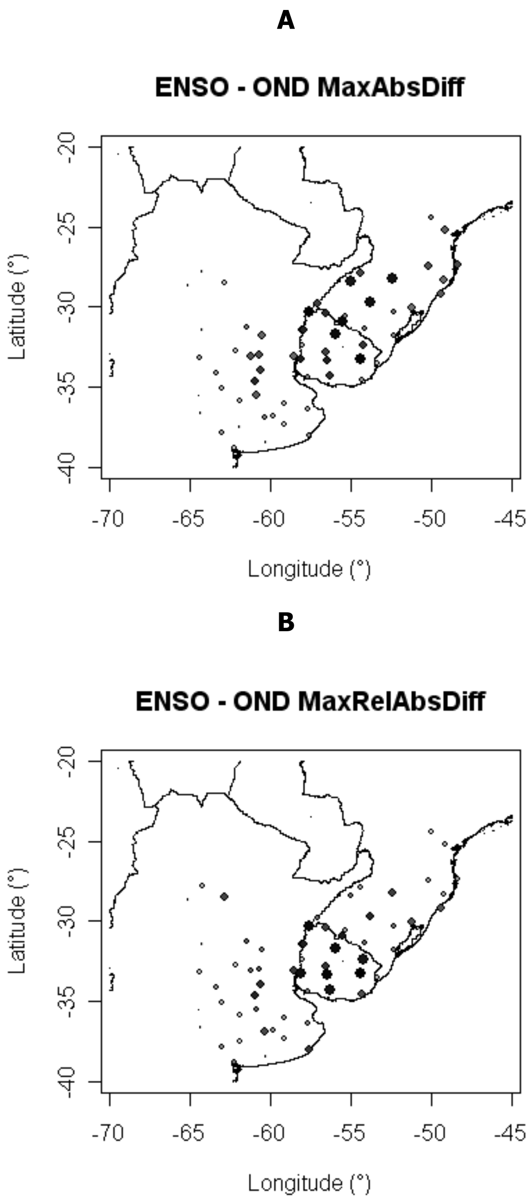

Figure 4. Spatial patterns of ENSO class influences on OND rainfall medians over Southeast of South America, as measured by (A) MaxAbsDiff and (B) MaxRelAbsDiff. (n=60 stations; point size and color intensity proportional to data values)

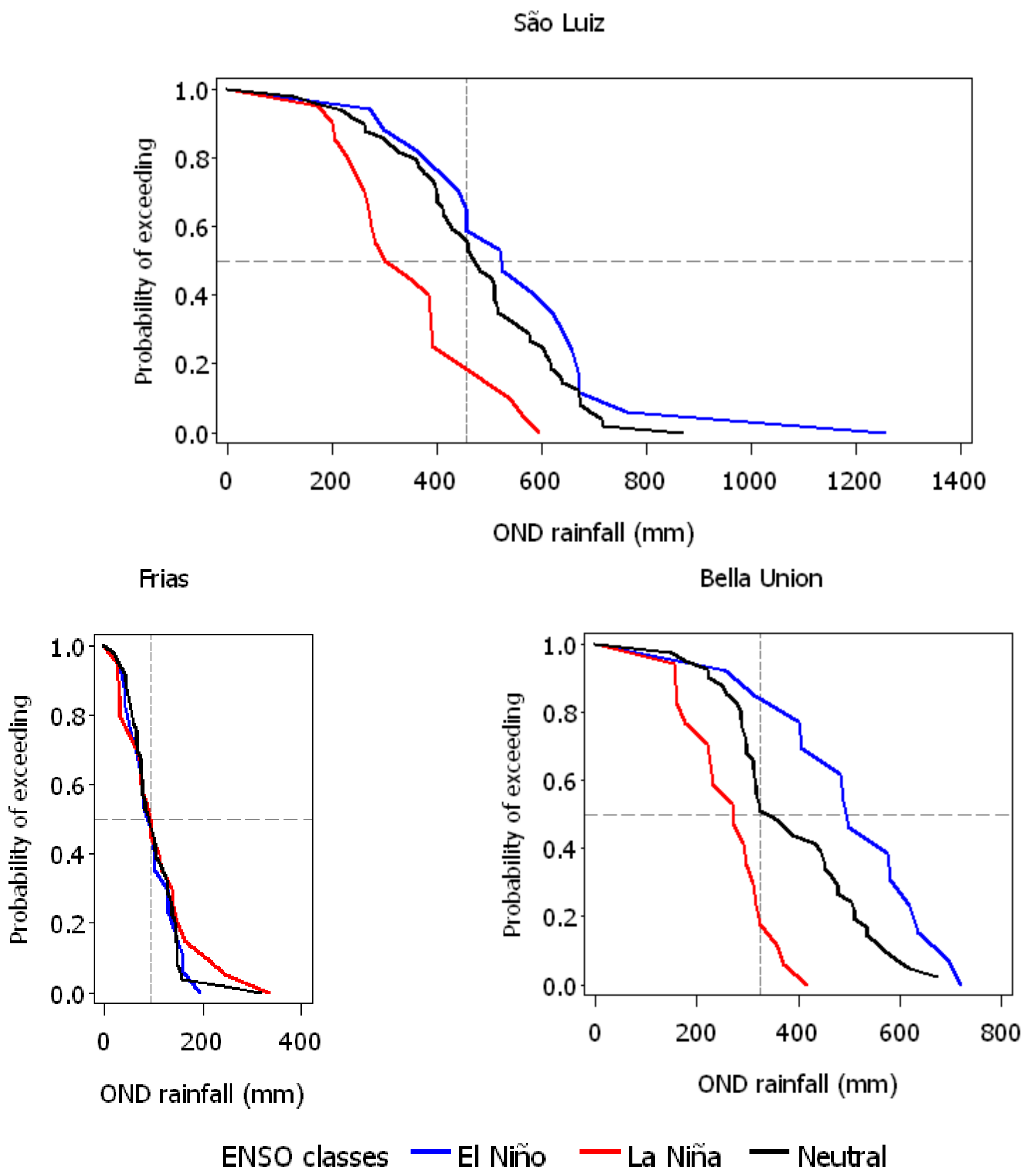

Figure 5. Probability of exceeding functions representing OND rainfall series for each ENSO class at São Luiz (Brazil), Frias (Argentina) and Bella Union (Uruguay).

Table 6. OND rainfall medians at each ENSO class and respective maximum absolute difference between conditional and unconditional medians (MaxAbsDiff), at three selected locations.

Location	MedianAll (mm)	Class Median (mm)			MaxAbsDiff (mm)	
		El Ninõ	Neutral	La Niña	value	signal*
São Luiz (n=86)	456	524	474	324	132	-
Frias (n=86)	93	93	91	94	2	-
Bella Union (n=71)	325	498	355	272	173	+

*Values in red correspond to classes at which the maximum differences were observed.

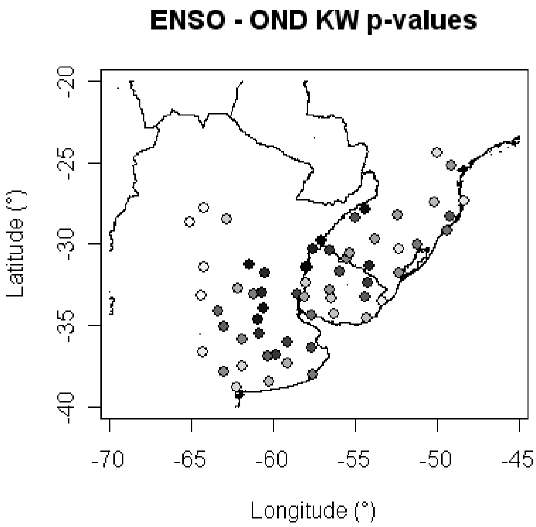

Figure 6. Kruskal-Wallis p-values applied to quantify evidences of the ENSO class influence on OND rainfall medians from 60 stations over Southeast of South America (point size and color intensity are inversely proportional to KW p-values, ranging from 0 to 1).

KW p-values indicated high ENSO signal for the region and for the 3-monthly period analysed (OND). P-values exceeded 0.10 in only three cases: Santiago del Estero (p=0.12,; MaxDiff=- 43 mm), Rio Cuarto (p=0.13; MaxDiff=-88mm) and Frias (p=0.97; MaxDiff=-2 mm).

Concluding remarks

Descriptive assessments quantify the magnitude of class effects while inferential approaches quantify the probability of observed changes in measure of interest (median, probability of exceeding median, probability of exceeding a critical threshold) arising by chance. Proper scenario planning requires both – inferential approaches to quantify the likelihood of spurious 'signals' arising by chance (eg. artificial skill) as well as thorough, quantitative assessments of signal strength for decision making.

High relative divergences between conditional and unconditional descriptive measures (MaxRelDiff or MaxRelVDist) are frequently observed for dry areas, even when evidences of class effects are negligible. Therefore, careful analyses of both relative and absolute measures, in conjunction outcomes from appropriate distribution-free tests are required in order to assess forecast quality and use such forecasts for decision making.

References

Conover, W. J., 1980: Practical Nonparametric Statistics. 3d ed. John Wiley & Sons, 584pp.

Ferreyra, R.A., G. P. Podestá, C. D. Messina, D. Letson, J. Dardanelli, E. Guevara, and S. Meira, 2001: A linked-modeling framework to estimate maize production risk associated with ENSO-related climate variability in Argentina. Agricultural and Forest Meteorology, 107, 177–192.

Kruskal, W. H. and W. A. Wallis, 1952: Use of ranks in one-criterion variance analysis. Journal of the Arerican Statistical Association, 47, 583-621.

Maia, A.H.N., Meinke, H., Lennox, S. and R. C. Stone, 2006. Inferential, non-parametric statistics to assess quality of probabilistic forecast systems. Monthly Weather Review, in review.

Meinke H. and R. C. Stone, 2005: Seasonal and inter-annual climate forecasting: the new tool for increasing preparedness to climate variability and change in agricultural planning and operations. Climatic Change, 70, 221-253.

Potgieter, A.B., Y. Everingham, and G. L. Hammer, 2003: On measuring quality of a commodity forecasting from a system that incorporates seasonal climate forecasts. International Journal of Climatology, 23, 1195-1210.

Ropelewski, C. F. and M. S. Halpert, 1987: Precipitation patterns associates with El Niño/Southern Oscillation. Monthly Weather Review, 115, 1606-1626.

Selvaraju, R., H. Meinke, and J. Hansen, 2004: Climate information contributes to better water management of irrigated cropping systems in Southern India. Proc 4th International Crop Science Congress, Brisbane, Australia.

Sivakumar, M.V.K., R. Gommes, and W. Baier, 2000: Agrometeorology and sustainable agriculture, Agricultural and Forest Meteorology, 103, 11-26.

Stone, R. C., 1992: SOI phase relationship with rainfall probabilities over Australia. Unpublished PhD Thesis, University of Queensland, Brisbane, Australia.