

Construction of a daily precipitation grid for southeastern South America for the period 1961–2000

P. D. Jones,^{a*} D. H. Lister,^a C. Harpham,^a M. Rusticucci^b and O. Penalba^b

^a Climatic Research Unit, University of East Anglia, Norwich, UK

^b Departamento de Ciencias de la Atmósfera y los Océanos, FCEN Universidad de Buenos Aires, Buenos Aires, Argentina

ABSTRACT: Daily station precipitation totals are used to develop a gridded dataset for the region (14°–40°S, 45°–70°W) on a 0.5° × 0.5° latitude/longitude grid, primarily for comparison with regional climate model (RCM) simulations. The gridded dataset covers the period 1961–2000. Much of the paper discusses the quality control of the basic station precipitation series. Although the primary aim of the development has been RCM validation, we have assessed trends in seasonal precipitation totals as well as trends in two measures of precipitation extremes (R95p, the daily precipitation amount exceeded only 5% of the time and Rx5day, the maximum 5-d precipitation total during each season). Relatively few regions across the large domain have statistically significant trends, but those that do tend to be located in the eastern two thirds of the grid, particularly over southeastern Brazil and Uruguay. Significant trends are also more evident in the DJF and MAM seasons. There is good spatial agreement between the trends in seasonal totals and trends in the extreme indices. Copyright © 2012 Royal Meteorological Society

KEY WORDS precipitation; South America; gridding

Received 12 July 2012; Revised 11 September 2012; Accepted 15 September 2012

1. Introduction

To validate climate models, it is generally essential to use some form of gridded observational data. Much assessment of this kind is undertaken, especially, using the output from one of the various Reanalysis products that are now available. There will always be some element of circularity in such comparisons as Reanalysis products are model derived, albeit through an assimilation of various forms of observational data. Reanalysis products are also often considered to be relatively poor when it comes to precipitation data, especially at the daily timescale (e.g. Dulière *et al.*, 2011). The purpose of this article is to develop a daily gridded precipitation dataset for 1961–2000 for the southeastern part of South America encompassing the catchment area of the La Plata Basin (LPB) for comparison with regional climate model (RCM) simulations. Other papers resulting from the European Union funded project (CLARIS-LPB) will discuss the RCM comparisons (the project is discussed in detail by Boulanger *et al.*, 2010). This paper will assess the quality of the daily observational precipitation data, the accuracy of the gridding and conclude with some discussion of precipitation extremes in the gridded product, in comparison with extremes in the original daily station precipitation series.

There are a number of gridded precipitation products available, some of which additionally include satellite estimates and/or Reanalysis output. Here we will concentrate on those that are solely based on daily precipitation totals measured at observational stations. The earliest study of this approach, with the principal aim of RCM assessment, was for the Alpine region developed by Frei and Schär (1998). Interpolation to a grid was also a means of providing access to a form of the data – the gridded product, as the National Meteorological Services (NMSs) in the region were (and still are) not able to make all the daily station precipitation data freely available. Daily station data access is also an issue in the Asian grids developed recently by Yatagai *et al.* (2009) and earlier for India by Rajeevan *et al.* (2006). The easily accessible station data for the contiguous United States has enabled a number of different techniques for interpolation to be intercompared with Di Luzio *et al.* (2008) being a recent review of a number of approaches. This paper and Ensor and Robeson (2008) go further and address comparisons (particularly for extremes) of the original station data with the gridded products (see also Haylock *et al.* 2008). In data-dense regions, the latter produce complete datasets overcoming the problems of missing point observations, but what is the cost of possible spatial smoothing when looking at extremes in the grid and also at station point locations? As it will likely be the extremes that could be emphasized in any RCM intercomparisons, we will return to this aspect towards the end of this study.

* Correspondence to: P. D. Jones, Climatic Research Unit, University of East Anglia, Norwich, UK. E-mail: p.jones@uea.ac.uk

For our study region, the most relevant earlier analysis is that from Liebmann and Allured (2005), who developed relatively coarse-resolution daily precipitation grids across much of South America east of the Andes for the period 1940–2003. Although not going into much detail, this paper touches on all of the problems of data quality in this region that we will return to in the next section. An extensive set of daily precipitation grids has been produced for Europe (Haylock *et al.*, 2008) and these are updated routinely through the European Climate Assessment & Dataset (ECA&D, <http://eca.knmi.nl/>). The interpolation procedure used in this analysis has already been used in the same CLARIS-LPB project for the LPB for daily maximum and minimum temperature by Tencer *et al.* (2011). In this paper, we will use the same interpolation technique with the daily station precipitation data collected by the CLARIS-LPB project. The next section (Section 2) discusses the quality control (QC) of the daily precipitation dataset. Section 3 briefly summarizes the gridding approach of Haylock *et al.* (2008). Section 4 validates the developed grid using short station records not used in the gridding. Section 5 analyses the grid and some of the station series for extremes and Section 6 concludes.

2. Precipitation data and its QC assessment

The CLARIS-LPB project has been collecting daily station precipitation, temperature and other climate variables for the LPB study region. A full list of the data sources can be found at <http://www.claris-eu.org/> and then go to collaborators. Tencer *et al.* (2011) have used daily station temperature (maximum and minimum) data to develop a $0.5^\circ \times 0.5^\circ$ latitude/longitude gridded database across the region (20° – 40° S, 45° – 70° W). The number of daily temperature series is markedly fewer than those collected for daily precipitation. The much greater density of the precipitation series means that there is greater potential to assess the quality of the basic data, but it must be remembered that correlation decay lengths for precipitation are markedly shorter than for temperature. So a denser network does not necessarily mean that the quality can be better assessed. The data are not evenly distributed, however, over the study region, so it will only really be possible over parts of the LPB, principally southeastern Brazil. This section will briefly allude to the problems encountered when attempting to assess the quality of the daily precipitation series. The initial problem is that the station data have been collected from a variety of sources, particularly across Brazil. Station names are not included in the data base for many stations, but this is not a problem as the stations all have a latitude and longitude location. They do not, however, have an elevation. Elevation is an important variable, especially when observed series will be interpolated over geographically diverse terrain. For this reason, it was necessary to extract the elevation values from a high-resolution global elevation database. We used the facility provided by CGIAR (SRTM 90 m – see:

<http://www.cgiar-csi.org/data/elevation/item/45-srtm-90m-digital-elevation-database-v41>). To verify this process, we estimated the elevation for all stations within the latitude/longitude box we use for the precipitation grid (14° – 40° S, 45° – 70° W) within the World Meteorological Organization (WMO) system, where the station elevation is known. The elevations are very well estimated ($r = 0.994$ based on 301 locations).

The total number of observed daily precipitation series available was 8110, but some of these were either very short or contained no data at all. Some form of QC of the series requires there to be a reasonable length of data, so the first aspect was to develop series of monthly totals and to remove very short series. During the reformatting and conversion to monthly series, filters were applied that removed short-duration series. The first stage of filtering ignored any series having less than 1500 daily, non-missing values. This is the equivalent to 4–5 years of data. In the daily-to-monthly calculation, a minimum of 28 values in any 1 month were required – otherwise, the monthly total precipitation was coded to the missing value. In addition, there were a number of series which had leading and/or trailing blocks of missing values. This is largely because of differences in the availability of data between temperature and precipitation values which were stored together. Blocks of leading and/or trailing missing values were removed because they serve no purpose when working with a single variable. All of the detection of QC issues has been conducted on monthly total precipitation series, derived from the daily database. That is, if a monthly value is in error, all daily values for that station-month are replaced with the missing code. If a monthly value was calculated to be a missing value, any daily values would not be subject to QC.

The number of monthly series emerging from the exercise was 7065. However, the number of potential series that can enter the daily gridding process will likely be much lower than this as it is necessary for each station to have a minimum length of series to derive average values for a base period. After checking the geographical and temporal density of observed series across the LPB, it was clear that the optimum period for the gridding operation was 1961–2000. Series with less than 15 but more than 5 years of data were retained in a separate file, as they will be useful for independent assessment of the quality of the gridded product. These series (referred to as Group 2) were only assessed for duplicates (see later). Series with at least 15 years are referred to as Group 1.

The monthly series from Group 1 were initially assessed for cases of extreme-high and -low outliers, and also for cases of anomalous sequences of zero-precipitation months. The latter phenomenon usually occurs if zeroes have been used in a database instead of missing value codes. This was additionally seen at the start and sometimes at the ends of the monthly or daily series (see earlier). The decision trees which judged whether or not extreme outliers are real phenomena rely on monthly deviations calculated from long-term means and standard deviations. For this to work it depends on

the series having sufficient monthly values to satisfy the minimum criterion set for the generation of means and standard deviations (in this case – 15 years). For sequences of zero precipitation months, we flagged the series with more than 3 months in a row and compared with neighbours. Apparently erroneous sequences were set to missing, except in regions with very low seasonal precipitation totals.

After undertaking a preliminary gridding operation using the Group 1 daily series, a series of validation tests were run to see how grid-box series compare with nearby observed series from Group 2. The basis of these tests was correlation over their common period. To gauge the degree of correlation that should be expected in the LPB region, a separate correlation exercise was performed (not shown) where all series (used in the gridding) were correlated with their nearest neighbours. As a result of these preliminary exercises it became apparent there was an additional problem with the daily dataset. A significant number of neighbouring series were showing perfect correlation over their common period. This means that precipitation sequences must be exactly identical with other series having different ID codes. The series were in close proximity to each other, but the location coordinates were not identical. Even two rain gauges at the same site do not give perfect correlations when their time series are compared. The implication is that the same data have entered the database from different sources or been assumed to be different when later data were received from the same source at some later time. A list of pairs of stations with common data subsets was produced (552 cases, with nearly all in southern Brazil). It was noted that the pairs of stations often have a different period of coverage (with a common period). This permitted a merger of these pairs of series which produced a single series that maximizes the temporal coverage within the period 1961–2000 (the period chosen for the gridding operation). In addition, it removes the ‘duplicate’ series. Group 2 stations were also assessed for duplicates with exactly the same approach and also against Group 1 data.

In addition to the cases of perfect correlation between neighbouring station series, it was also noted that there were some cases of very poor/negligible correlation. Where distances were short, and there was corroboration of poor correlation from the other near neighbours, this was taken to mean that there was some error with the station location or its data. There were 13 station series that showed this problem. These were removed from the gridding process.

In a further simple QC test, all precipitation series used in the gridding were assessed for homogeneity by the plotting of time series of annual precipitation totals. This requires the manual inspection of a large number of time series. The purpose of this test was to ensure that the series with spurious/erroneous zero precipitation values had been removed and to check that all the series were in the same units. A few stations showed evidence of possible homogeneity issues. Figure 1 shows an example of a Brazilian station where annual precipitation totals

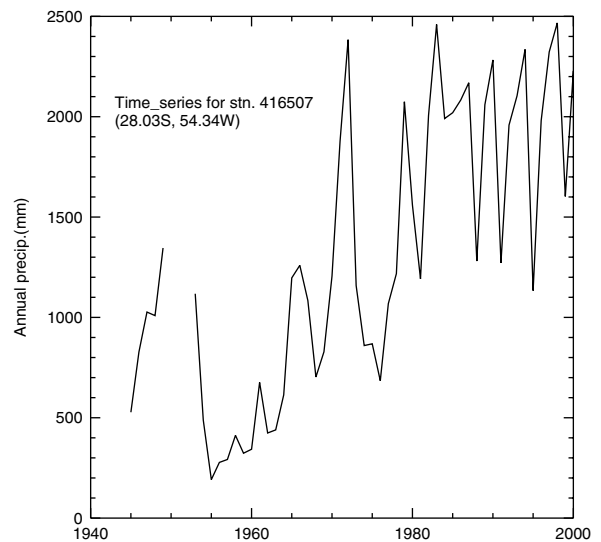


Figure 1. Example of a probable inhomogeneity in a long annual precipitation record from southern Brazil.

after 1970 are roughly twice the totals for years before. We are unsure of the reason for this, but where this was clear and unambiguous the entire record for the station was removed. Altogether 13 stations were removed and a further eight had short sections removed because of homogeneity issues. All were from the data-dense regions in Brazil. However, another phenomenon was noted – that of very low annual rainfall in 1987 – at a number of Brazilian series. For this region in southeastern Brazil, 1987 was an El Niño year, so rainfall would be expected to be average or above average and not the driest on record. This was most common in series that ended in 1987. Series showing the 1987 problem (87 in total) had their daily values for 1987 coded to missing. Overall, the number of removed stations or parts of their data set to missing is relatively small compared with the overall count of stations used. All removals were from data-dense regions, so efforts to correct them were not attempted.

Following the QC and other processes described above, a total of 3945 daily precipitation series were available for the daily gridding process. Figure 2 shows a map of the locations of these sites. The greater density of sites across Brazil is clearly apparent, together with the even greater density over São Paulo state. The overall number of Group 2 stations is 1288, but the number available in an individual year varies from 70 to 680. Figure 3 shows the locations of the Group 2 stations (for four example years, 1966, 1974, 1983 and 1996) that will be used as validation of the gridded product. Stations within 1° latitude and longitude of the bounding box were included in both figures, but stations beyond this were excluded.

3. The gridding process

As used by Tencer *et al.* (2011) for the development of daily temperature grids for the CLARIS-LPB region, we also use the gridding software developed by Haylock *et*

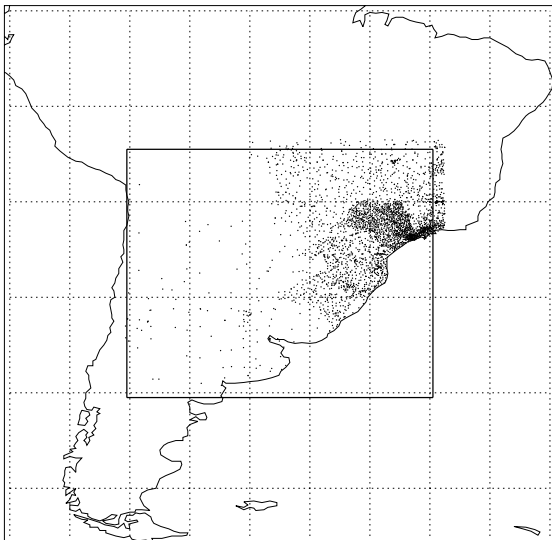


Figure 2. Locations of the 3945 rain gauges with at least 15 years of data during the 1961–2000 period. The box bounds the grid that will be produced from 14° to 40° S \times 45° to 70° W. The total includes stations which are beyond the bounding box by 1° of latitude and longitude.

al. (2008). This software is used operationally by the European Climate Assessment and Dataset (ECA&D) to update gridded products across the European domain (<http://eca.knmi.nl/>). For precipitation, the gridding is a two-stage process with first monthly totals being interpolated to the $0.5^{\circ} \times 0.5^{\circ}$ latitude/longitude grid for each month for the period 1961–2000. The resulting product is an areal average for the grid box. This is achieved by interpolating to each of the $0.1^{\circ} \times 0.1^{\circ}$ boxes within each grid box and then averaging these 25 smaller boxes into a grid-box average. The approach achieves an areal average, more akin to that produced by an RCM. This first stage interpolation uses elevation (at each of the smaller boxes) as a covariate and is undertaken using splines. Haylock *et al.* (2008) found that for the base period splines were better than other interpolation techniques such as kriging. The search radius for the splines was 1200 km.

The second stage uses kriging for the daily amounts. In order to ensure compatibility with the monthly grids, we interpolated the rainfall deviation from the monthly total. As a daily precipitation series can be considered binary, due to the occurrence nature of the process, the gridding transforms the rainfall to a binary distribution dependent on whether the value was above or below the threshold for a rainy day (0.5 mm). This is similar to the process introduced by Barancourt and Creutin (1992). The final component of producing the grid involves the simple combination of the daily anomaly grids with the monthly to develop daily grids in millimetre units.

It is important to assess the uncertainty involved in the gridding process. There are many factors to be incorporated into any uncertainty estimate. Here, the estimates derived are from the interpolation, but not from the measurement uncertainty. The latter is very difficult to estimate, but likely to be much smaller

than the interpolation uncertainty (Hofstra *et al.*, 2008). Interpolation uncertainty is composed of a combination of the two stages of interpolation: the monthly spline-based and the daily kriging-based. Full details of the approach used are given in the study by Haylock *et al.* (2008). Another traditional method of assessing uncertainty is to leave some data out and assess how well the interpolated values compare with the omitted data. During the course of the assessment of the precipitation data quality, we have omitted data with relatively few years of data (<15 but >5), so we can use these data as a guide to the quality of the interpolated fields. This aspect is discussed in the next section.

The ECA&D datasets have been assessed for quality in Haylock *et al.* (2008) and by Hofstra *et al.* (2009). Hofstra *et al.* (2008) have additionally compared a number of different interpolation techniques and shown that the two-stage interpolation process used here works well with the density of data available across Europe. For the CLARIS-LPB region the density of station data is markedly poorer (Tencer *et al.*, 2011), but from Figure 2 the density is greater across parts of southeastern Brazil. Because the density is sparser overall than in Europe, the coarser of the two resolutions used across Europe was chosen ($0.5^{\circ} \times 0.5^{\circ}$ latitude/longitude as opposed to $0.25^{\circ} \times 0.25^{\circ}$). In the CLARIS-LPB project, RCM simulations have already been transformed from their native grids to this latitude/longitude grid.

In Section 1, we stated that the primary aim of this interpolation exercise is not to look at changes in precipitation patterns and distributions across the study region, but to develop a dataset that can principally be used to compare with RCM simulations for the CLARIS-LPB region. The gridded fields can, however, additionally be used to look at changes in climate, but the grids are not being updated to the present as in Europe and they only cover the period from 1961 to 2000. More importantly, Haylock *et al.* (2008) noted that the gridded fields involve, as expected, some degree of smoothing. The two-stage gridding process attempts to alleviate this to some extent, but smoothing is still evident in the final gridded fields. This smoothing is most evident in the extreme precipitation totals, particularly values occurring only a few times in each station dataset. It is likely that these extremes will be reduced in a gridded product (especially one that is developing grid-box averages as opposed to point values), in an analogous way to ‘areal reduction factors’ when looking at extremes in point and areal rainfall totals (e.g. Svensson and Jones, 2010). However, the reduction that takes place will be very dependent on the density of the network and the type of precipitation (frontal or convective) that caused the extreme.

Figure 4 shows the completeness of the grid. The northwest corner of the area is always missing, and coverage is incomplete in this region and in the very southwest of the grid. These regions are well outside the LPB though. For the basin itself the only area which is incomplete is parts of São Paulo state for a couple of

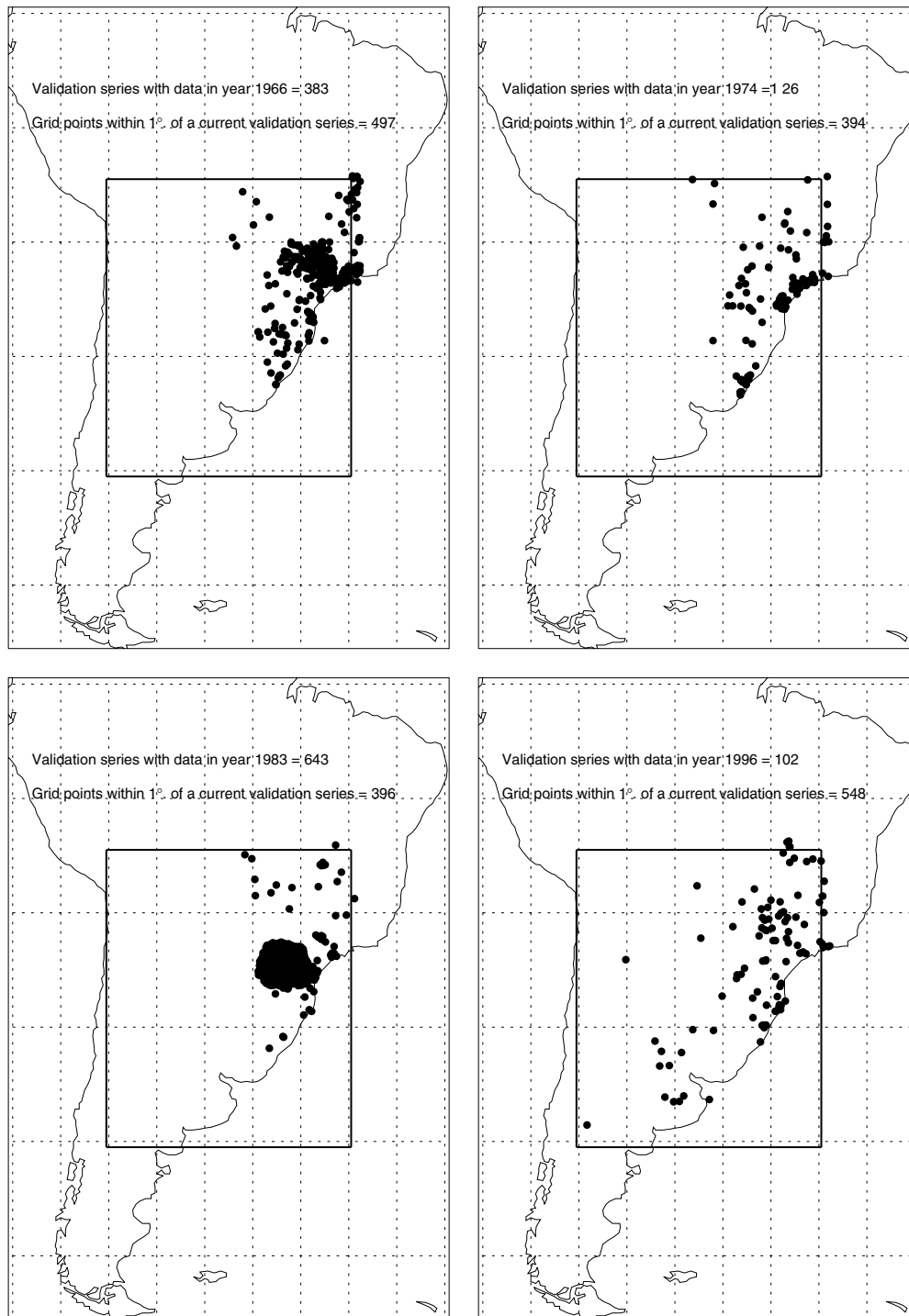


Figure 3. Locations of the 1288 stations in the validation set (i.e. the Group 2 stations that have between 5 and 15 years of daily precipitation data). The number of Group 2 stations and the number of grid boxes validated is also given for each of the 4 years.

recent years. This is despite the density of the network in this region in Figures 2 and 3.

4. Validation

A total of 1288 Group 2 stations with between 5 and 15 years of data will be used in this assessment. These stations are independent of those that were used in the development of the grid. The coverage of these stations

has been shown earlier in Figure 3 for four example years. Figure 5 shows the correlations between these validation series and the nearest grid-box series in the gridded product. Most of the correlations between the Group 2 data and the gridded product are above 0.6, but there are 162 stations below 0.5. These validation stations have relatively short records and the correlations can cover any period during the 1961–2000 period (see caption to Figure 5). From the earlier QC work, it became obvious

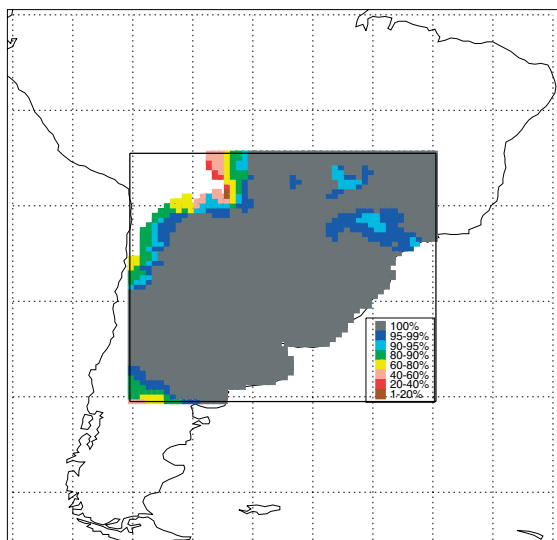


Figure 4. Completeness of the grid over the period 1961–2000. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

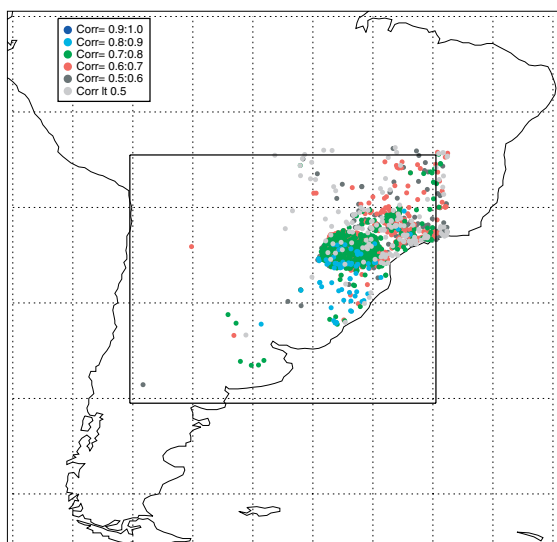


Figure 5. Validation of the gridded product against the stations with between 5 and 15 years of daily precipitation data that were not used in the gridding (restricted to stations within 1° of latitude/longitude of the bounding box). Correlations are based on all available daily data in the period 1961–2000, so we use between 60 and 180 of the total of 480 months of daily data in the gridded product. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

that the majority of these data will come from the 1970s and 1980s as opposed to the 1960s and 1990s. Also, most of the validation data comes from Brazil, particularly São Paulo state.

In Figure 6, we try and illustrate the variations in the temporal availability of the grid boxes that can be validated. We do this by showing the percentage of grid boxes with at least one daily rainfall station available in each of the 40 years for validation purposes. The time series are somewhat surprising looking at the distribution of sites shown in Figures 2 and 3. More grid-box series

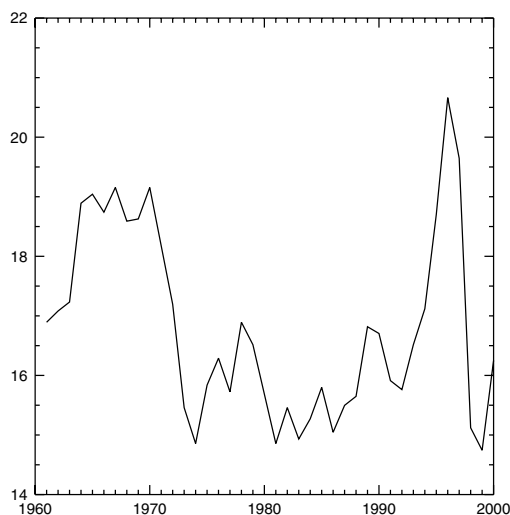


Figure 6. Time series of the percentage of the grid boxes having at least one validation station within 1° of latitude and longitude of the bounding box (for the year in question).

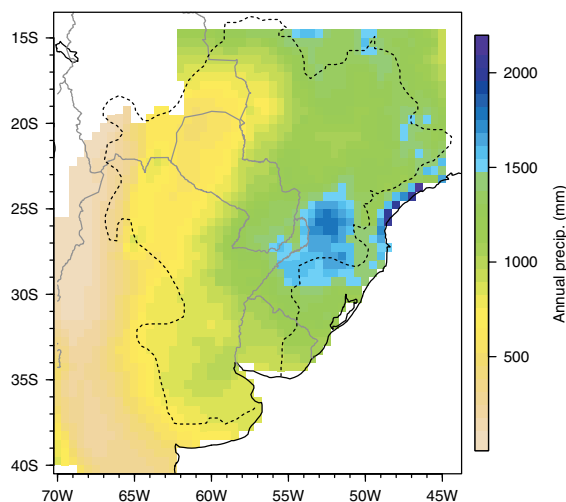


Figure 7. Average annual precipitation (mm) for the base period 1961–2000 from the gridded product. Country outlines are shown in grey and the La Plata Basin outline in black dashed on this and subsequent plots. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

can be validated during 1964–1970 and again in the mid-1990s than at other times. Figure 3 explains this by showing the number and location of the validation series available for four different years during the 1961–2000 period (1964, 1972, 1983 and 1996). There were a great number of sites in parts of Brazil for the period 1976–1987. Many of these rain-gauge series are so close together that they duplicate validation correlations against the same grid-box series as other Group 2 gauges. So despite the great number of potential validation sites coming from São Paulo state, many of these only have data for 12 of the 40 years that have been developed in the gridded product. It is also somewhat ironic that the most densely covered part of the region is not complete enough over the full 40 years to develop complete gridded series (Figure 4).

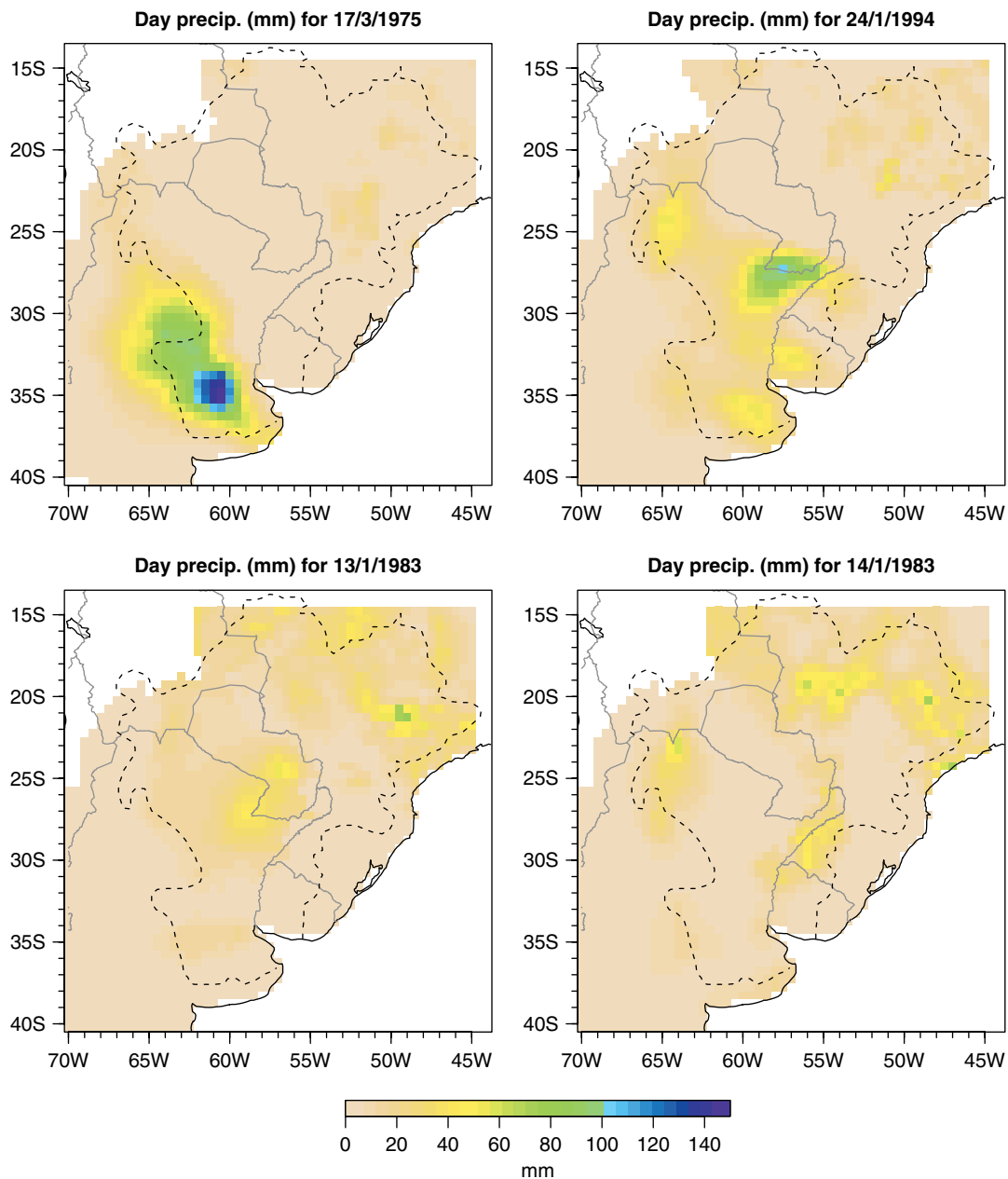


Figure 8. Examples of daily precipitation (mm) for four of the wettest days during the 1961–2000 period. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

5. Analyses of the gridded product

The principal aim of this study has been to derive a gridded daily precipitation dataset for the assessment of the growing number of RCM simulations that are now available for South America. In addition to this, and in this section, we will undertake some analyses of extreme precipitation trends across the LPB. The Expert Team on Climate Change Detection and Indices (ETCCDI, <http://www.clivar.org/organization/etccdi>), a joint working group between the World Meteorological Organization's Commission for Climatology, CLIVAR and the Joint WMO-International Oceanographic Commission (IOC) Technical Commission for Oceanography and Marine Meteorology (WMO/CCL, CLIVAR, JCOMM),

has developed a number of indices of extremes, particularly for temperature and precipitation. In this exercise we will consider two: the maximum amount of precipitation over a 5-d period (Rx5day) and the daily precipitation amount exceeded only 5% of the time (R95p). The formulae used to estimate the values for each year (or season) are given on the ETCCDI website and we used the F-CLIMDEX version of the available software. These two indices are defined on this web page (http://ccma.seos.uvic.ca/ETCCDI/list_27_indices.shtml).

Both these indices will be calculated for each season and calendar year for each of the 40 years of the gridded product, but we will just show the trends in the four seasonal totals. The results of this exercise will also

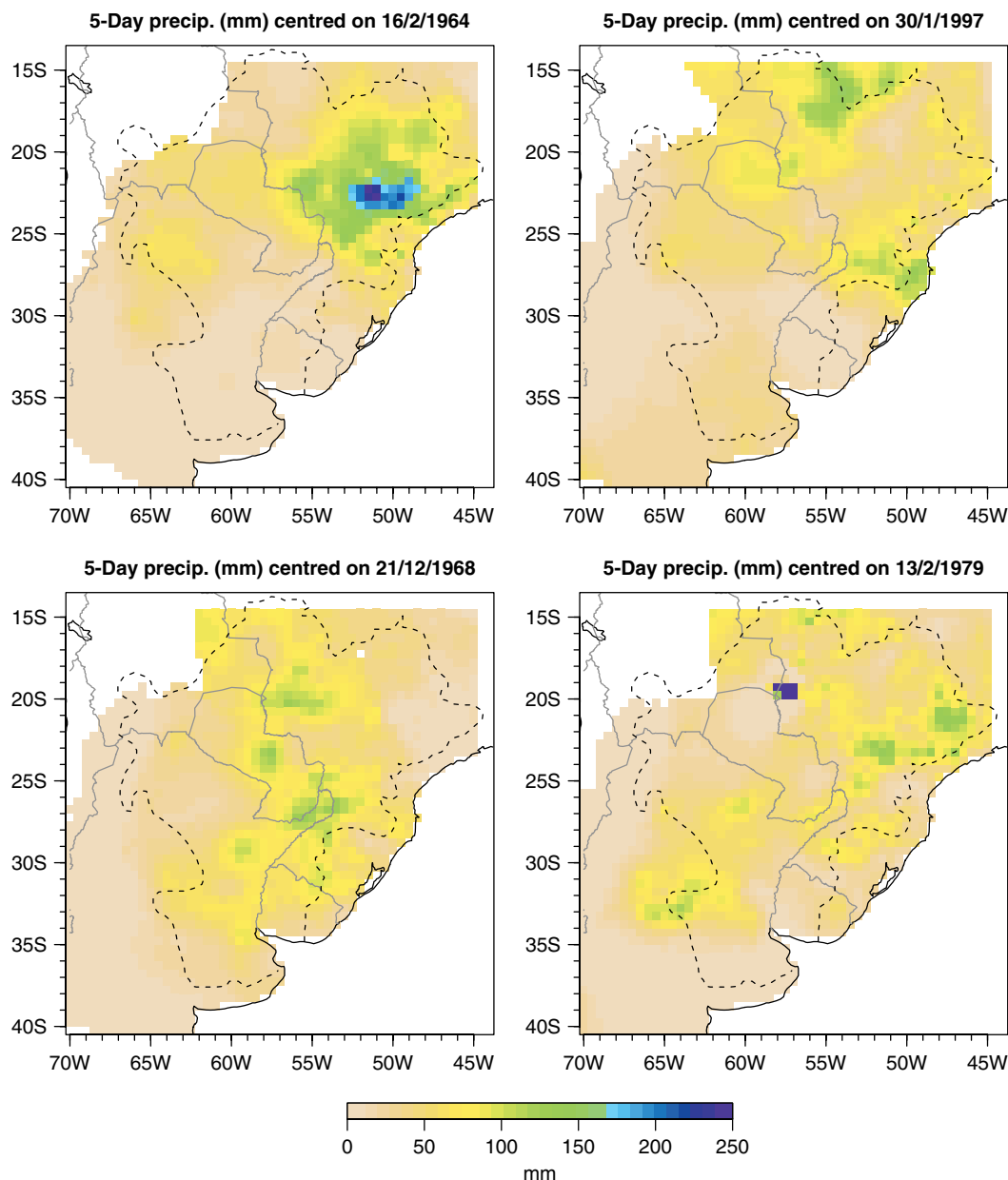


Figure 9. Examples of 5-d precipitation totals (mm) for four of the wettest 5-d sequences during the 1961–2000 period. The day given is the central day of the sequence. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

be compared with Haylock *et al.* (2006) who looked at precipitation extremes across much of South America for approximately the same period (1961–2000) and Re and Barros (2009) who looked at extreme heavy precipitation indicators across southeastern South America for the years 1959–2002.

In order to compare changes in these two extremes, we additionally need to show some more basic precipitation maps with which the changes in extremes can be put into context. Figure 7 shows the average annual precipitation totals for the 1961–2000 period. Annual totals are low in the lee of the Andes, with the wettest amounts over southeastern Brazil. Seasonally (using the standard 3-month seasons, but not shown) indicate that over southeastern Brazil the wettest season is summer

(December to February, DJF). Figure 8 shows some example days of extreme precipitation totals for individual days across the entire gridded region. To obtain these dates, we produced an average daily total across the region and ranked the average values. The 1982–1983 DJF season was exceptionally wet and of the four wettest days two are adjacent days (13 and 14 January 1983). Figure 9 shows further extreme precipitation totals, this time for 5-d periods. The Rx5day index introduced above is the maximum value of this measure in any year or season (depending on the period analysed) for each grid box. The examples chosen in Figure 9 are 5-d periods among the wettest values in the 1961–1990 period, but not those that include the days used in Figure 8. With the days shown in Figure 9, the region highlighted is always

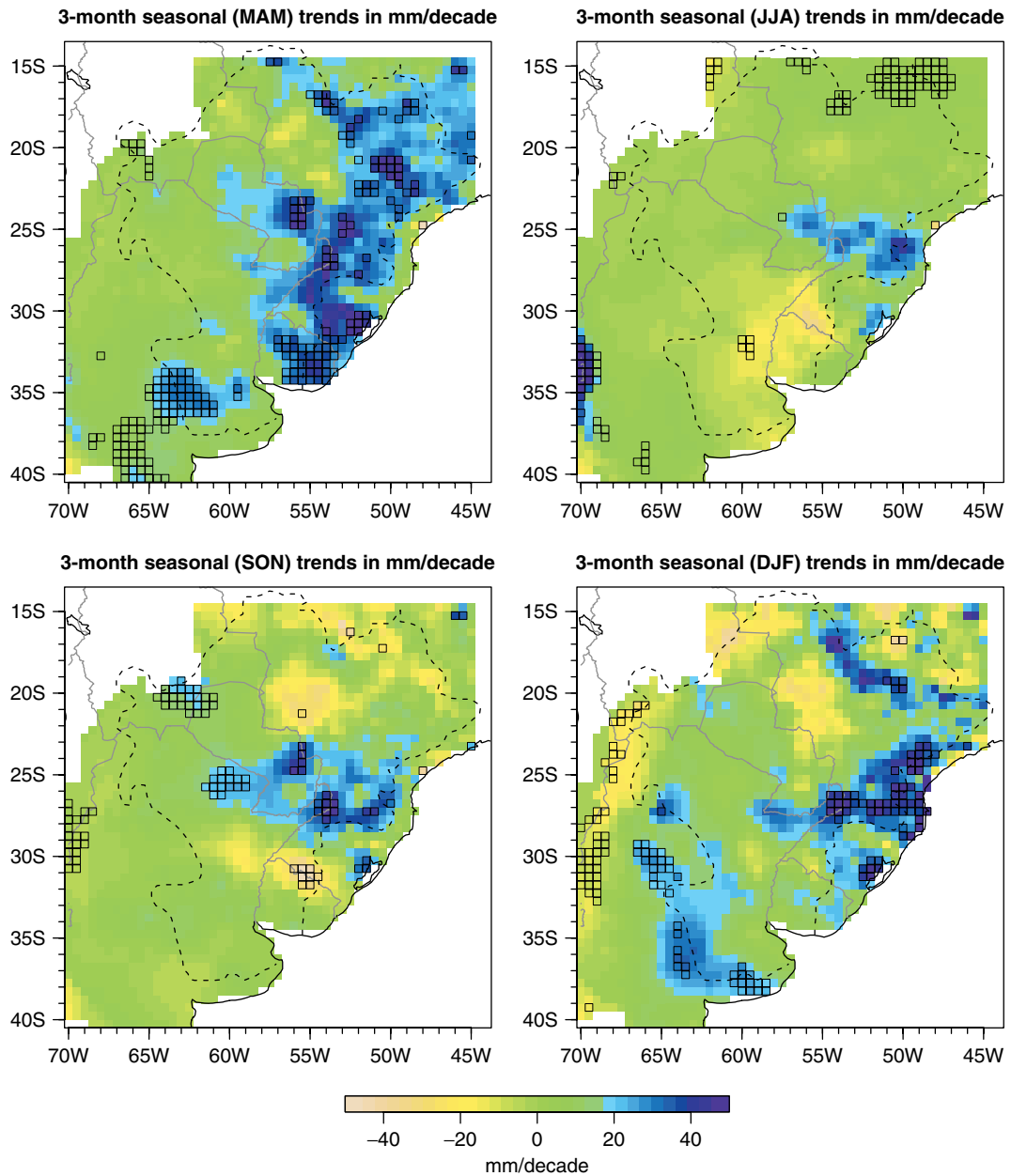


Figure 10. Trends of seasonal precipitation totals (mm/decade) for the 1961–2000 period. Statistical significance (5/95%) is indicated by outlining the square. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

southeastern Brazil, the wettest area within the domain. On the daily scale southeastern Brazil is emphasized, but two of the days have high precipitation totals further south encompassing northern or northeastern Argentina.

Figure 10 shows the trend of precipitation totals for each of the four standard seasons. Linear trends have been calculated, but with the number of degrees of freedom reduced according to any autocorrelation in the series, using the formula modified from Quenouille (1952). Here the effective number of degrees of freedom is reduced by $(1 - r_1)/(1 + r_1)$. Few regions exhibit statistically significant trends, particularly for the winter and spring (JJA and SON) seasons. The areas with statistically significant trends are greater in the other two seasons, particularly for autumn (MAM). Most of the

regions with significant trends are positive and located in the eastern two thirds of the domain. Areas with significant decreases are few, mainly in DJF in the lee of the Andes, but in this season this is the rainy season. These trends are also observed in the frequency of daily rainfall (>0.1 mm) (Penalba and Robledo, 2010).

Figures 11 and 12 show the linear trend maps for R95p and Rx5day, respectively, for the same four standard seasons. The effect of autocorrelation on the trends has been similarly reduced in a similar fashion to Figure 10. For R95p, the regions with significant trends occur more frequently in the DJF and MAM seasons and they generally indicate increases in this measure. Patterns for Rx5day are somewhat similar, although slightly more frequent, and again indicate more significant increases particularly

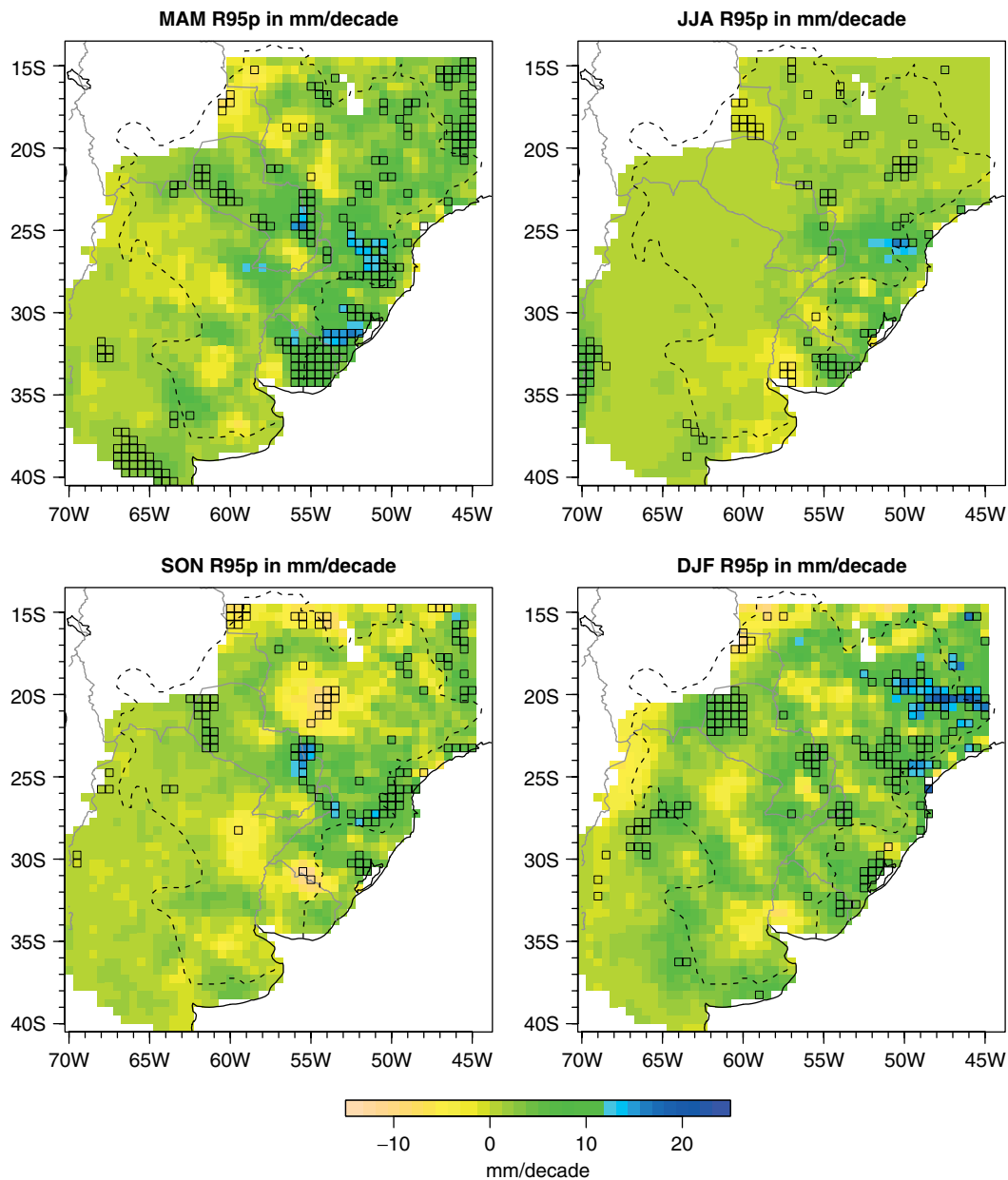


Figure 11. Seasonal trends of R95p in mm/decade for the period 1961–2000. Statistical significance (5/95%) is indicated by outlining the square. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

in the DJF and MAM compared with the other two seasons. Haylock *et al.* (2006) show annual trends in these two indices (along with trends in many more) for the period 1960–2000 using 54 series across the whole of South America. If we calculate similar annual trends for the 1961–1990 (not shown) the significant increases are also principally found in southeastern Brazil, Uruguay and northeastern Argentina (see also Re and Barros, 2009), more so for R95p compared to Rx5day.

Finally in Figure 13, we extract an example grid-box time series for the extreme index R95p that shows a strong upward trend. We also extract the single long daily precipitation series nearest to this grid box. The software is then used to calculate the indices from the station series and it is compared with that of the grid-box series.

6. Conclusions

This paper has developed a gridded daily precipitation dataset for the region (14°–40°S, 45°–70°W) on a 0.5° × 0.5° latitude/longitude grid. Considerable discussion in the paper has been expended on determining the quality of the daily precipitation data that will be used in the production of the grid. Several problems in the development of the basic daily station dataset have been noted, particularly the problem of duplicate series mainly across southern Brazil. Over half of the potential 8000 station series were not used because they were either too short, were duplicates or contained important errors. The latter number was particularly small, with only 26 series being discarded. Five hundred and fifty-two series were duplicates and were combined in order to develop

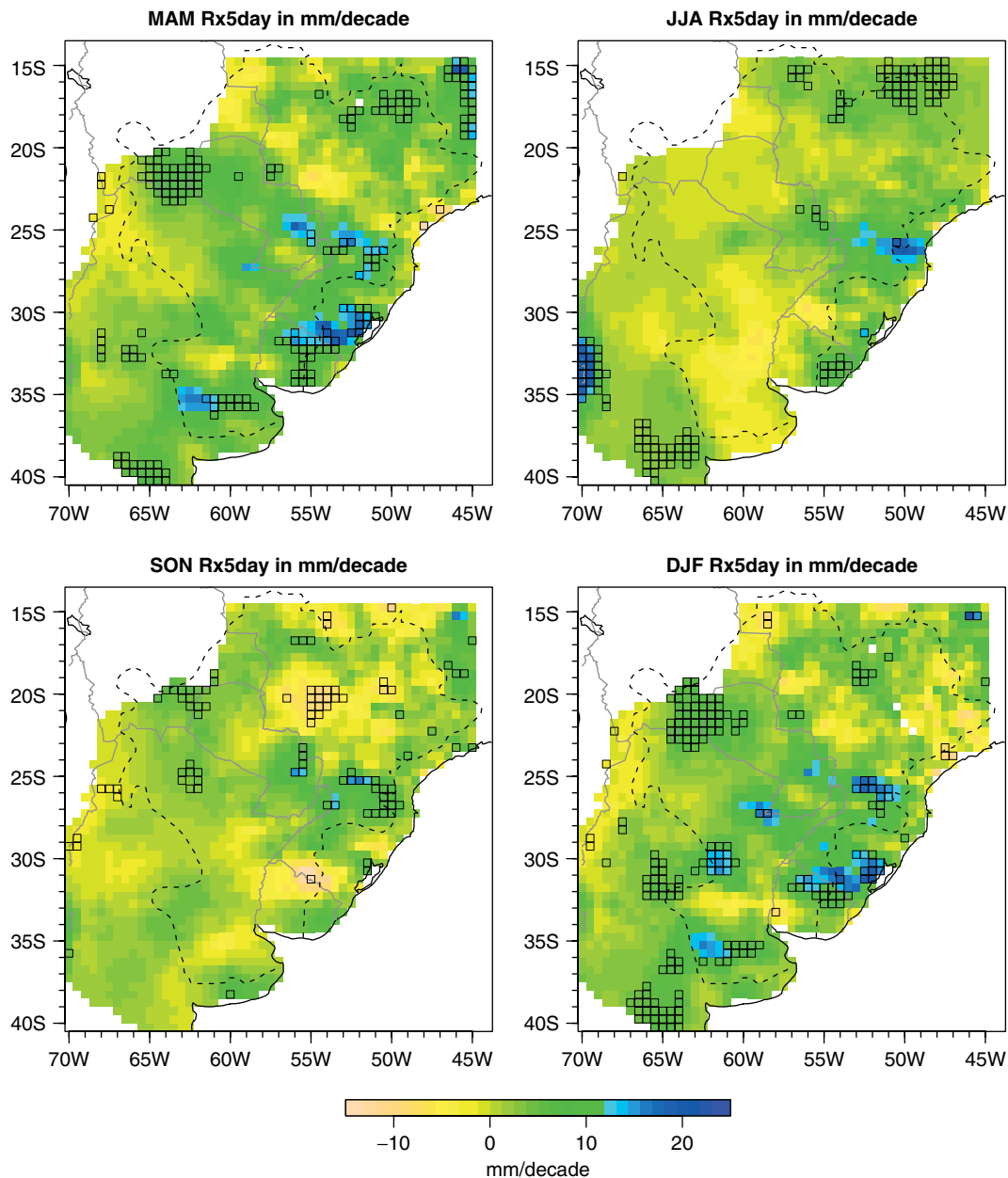


Figure 12. Seasonal trends of Rx5day in mm/decade for the period 1961–2000. Statistical significance (5/95%) is indicated by outlining the square. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

slightly longer series. The final number of usable stations (at least 15 years of daily data from the 1961–2000 period) was 3945, but a further 1288 with between 5 and 15 years of daily data were used in an independent assessment of the quality of the resulting grid. The gridding technique used was the same as that employed by Haylock *et al.* (2008) for Europe and also for a slightly smaller domain for the LPB region for maximum and minimum temperature by Tencer *et al.* (2011). The gridding first develops a monthly grid using thin-plate splines and then uses kriging of the daily precipitation totals expressed as percentages of their monthly totals. The two sets of grids are then recombined to millimetre, producing daily grids that sum to the monthly gridded totals.

The main aim of the development of the grid is for the comparison with RCM simulated daily precipitation totals by other groups in the CLARIS-LPB project. Although there is no intention to update the grid to more recent years, we have assessed trends in seasonal values of precipitation totals along with two extreme measures (R95p and Rx5day) recommended by the ETCCDI for the full period of analysis from 1961 to 2000. Relatively few regions across the large domain have statistically significant trends, but those that do are located in the eastern two thirds of the grid, particularly over southeastern Brazil and Uruguay. Significant trends are also more evident in the DJF and MAM seasons, and the locations of these agree well with earlier work by Haylock *et al.* (2006).

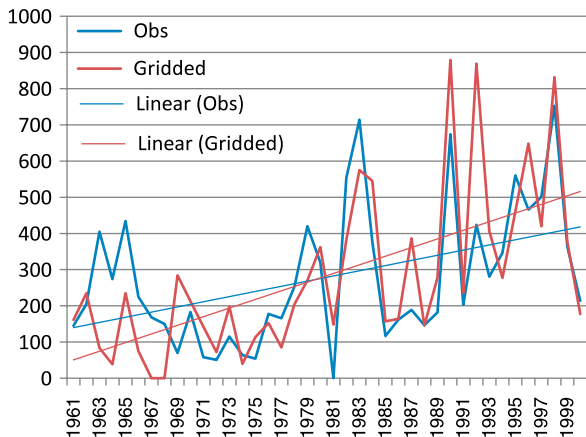


Figure 13. An example time-series comparison for an extreme high positive-trend case of R95p, showing the grid-box series with that for the nearest station. The grid box is centred at 25.75°S, 50.50°W and the station is located at 25.93°S, 50.58°W. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

Acknowledgements

The authors thank Brant Liebmann and the other reviewer for their thoughtful comments. This work was supported by the European Community's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement 212492 (CLARIS LPB: A Europe–South America Network for Climate Change Assessment and Impact Studies in La Plata Basin). The authors acknowledge the many data providers across the region who are listed at <http://www.claris-eu.org/> (under collaborators and stakeholders).

References

- Barancourt C, Creutin JD. 1992. A method for delineating and estimating rainfall fields. *Water Resources Research* **28**: 1133–1144.
- Boulanger JP, Brasseur G, Carril AF, Castro M, Degallier N, Ereño C, Le Treut H, Marengo JA, Menendez CG, Nuñez MN, Penalba OC, Rolla AL, Rusticucci M, Terra R. 2010. A Europe–South America network for climate change assessment and impact studies. *Climatic Change* **98**: 307–329.
- Di Luzio M, Johnson GL, Daly C, Eischeid JK, Arnold JG. 2008. Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *Journal of Applied Meteorology and Climatology* **47**: 475–497.
- Dulière V, Zhang Y, Salathé EP Jr. 2011. Extreme precipitation and temperature over the U.S. Pacific Northwest: a comparison between observations, reanalysis data, and regional models. *Journal of Climate* **24**: 1950–1964.
- Ensor LA, Robeson SM. 2008. Statistical characteristics of daily precipitation: comparisons of gridded and point datasets. *Journal of Applied Meteorology and Climatology* **47**: 2468–2476.
- Frei C, Schär C. 1998. A precipitation climatology of the Alps from high-resolution rain-gauge observations. *International Journal of Climatology* **18**: 873–900.
- Haylock MR, Hofstra N, Klein Tank AMG, Klok EJ, Jones PD, New M. 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research* **113**: D20119, DOI: 10.1029/2008JD010201.
- Haylock MR *et al.* 2006. Trends in total and extreme South American rainfall 1960–2000 and links with sea surface temperature. *Journal of Climate* **19**: 1490–1512.
- Hofstra N, Haylock MR, New M, Jones PD, Frei C. 2008. Comparison of six methods for the interpolation of daily European climate data. *Journal of Geophysical Research* **113**: D21110, DOI: 10.1029/2008JD010100.
- Hofstra N, Haylock M, New M, Jones PD. 2009. Testing E-OBS European high-resolution gridded data set of daily precipitation and surface temperature. *Journal of Geophysical Research* **114**: D21101, DOI: 10.1029/2009JD011799.
- Liebmann B, Allured D. 2005. Daily precipitation grids for South America. *Bulletin of the American Meteorological Society* **86**: 1567–1570.
- Quenouille MH. 1952. *Associated Measurements*. Butterworth Scientific Publications: London, 242pp.
- Penalba OC, Robledo F. 2010. Spatial and temporal variability of the frequency of extreme daily rainfall regime in the La Plata Basin during the 20th century. *Climatic Change* **98**: 531–550.
- Rajeevan M, Bhate J, Kale JD, Lal B. 2006. High resolution daily gridded rainfall data for the Indian region: analysis of break and active monsoons. *Current Science* **91**: 296–306.
- Re M, Barros VR. 2009. Extreme rainfalls in SE South America. *Climatic Change* **96**: 119–136.
- Svensson C, Jones DA. 2010. Review of methods for deriving area reduction factors. *Journal of Flood Risk Management* **3**: 232–245.
- Tencer B, Rusticucci M, Jones PD, Lister DH. 2011. A southeastern South American daily gridded data set of observed surface minimum and maximum temperature for 1961–2000. *Bulletin of the American Meteorological Society* **92**: 1339–1346.
- Yatagai A, Arakawa O, Kamiguchi K, Kawamoto H, Nodzu MI, Hamada A. 2009. A 44-year daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *SOLA* **5**: 137–140, DOI: 10.2151/sola.2009-035.