HEPEX – Flow forecasting in the Rio Grande watershed:

some preliminary results

1. TEST BED DESCRIPTION

The Rio Grande drains an area of about 145,000 km^2 of the Brazilian states of Minas Gerais and São Paulo, lying within a region roughly defined by 19^o to 22^o S and 43^o to 53^o W (Figure 1). The river is the main tributary of the River Paraná in its upper basin, and is used extensively for hydropower generation. The main hydropower installations are Marimbondo, Água Vermelha, Furnas and Estreito, each of which has installed capacity greater than 1,000 MW. In total, the Rio Grande basin has an installed capacity of about 8,780 MW, which corresponds to approximately 12.3% of the Brazilian total. Mean annual rainfall over the basin is approximately 1400 mm and is highly concentrated during 6 months from November to April.

Rainfall records of variable length and quality are available for 620 stations; flow records are available for 159 stations; and natural flows have been reconstructed for 19 sites extending back in some cases to 1931. Medium range forecasts (up to 15 days) and longer-term forecasts (up to a month or longer) are required for inflows into reservoirs from which hydropower is generated. Shorter-term forecasts (up to 7 days) are also of interest for local flood control purposes.

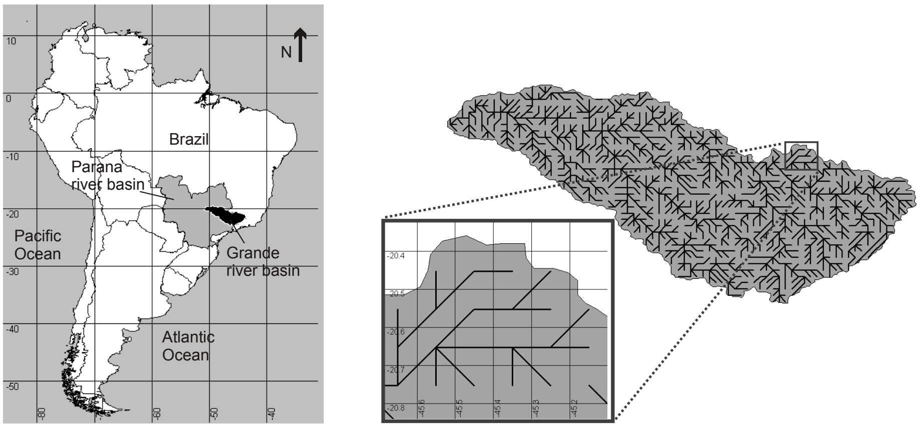


Figure 1 – Location of the Rio Grande watershed (left) and figure (right) showing division into cells with resolution 0.1º x 0.1º with the drainage pattern used in the hydrologic model.

2. OVERVIEW OF THE MGB-IPH MODEL

2.1 Model description

The Rio Grande basin is being modeled using the large-scale hydrologic model MGB-IPH (Collischonn and Tucci, 2001; Allasia et al., 2005), which is based on the LARSIM (Bremicker, 1998) and VIC-2L (Liang et al., 1994) models. It consists of modules for calculating soil water budget; evapotranspiration; flow propagation within a cell, and flow routing through the drainage network. The drainage basin is divided into elements of area (cells in a square grid), with vegetation and land use within each element categorized into several classes, the number of vegetation and land use types being at the choice of the user. To reduce computational load, the Grouped Response

Unit (GRU) (Kouwen *et al.*, 1993) approach is adopted. This consists of grouping all areas with a similar combination of soil and land cover, such that a cell contains a limited number of distinct GRUs. A soil water budget is computed for each GRU, and runoff generated from the different GRUs in the cell is then summed and routed to the river network. A soil water budget is calculated following the variable contributing area of the ARNO model (Todini, 1996). Evapotranspiration is calculated separately for each GRU in each cell following the model by Wigmosta (1994). Routing through the river network uses the Muskingum-Cunge method (Miller & Cunge, 1975). The time step normally used is one day, because rainfall is often only available in South America at a daily basis; however the model has been adapted to receive input and perform simulations in smaller time steps when such rainfall data are available (Collischonn *et al.*, 2005). In the work reported here, a daily time-step was used. During recent years the model has been tested and used in South American basins ranging from the sub-tropical, rapid response basins of Southern Brazil, to the basins in the Pantanal region, marked by seasonal rainfall, and, in some cases, slow response hydrographs (Collischonn & Tucci, 2001; Allasia et al., 2005). Ongoing applications include modeling the São Francisco river basin which lies partly in the semiarid region of Northeast Brazil (Tucci *et al.*, 2005), the Madeira river, one of the most important tributaries of the Amazon (Ribeiro *et al.*, 2005), and the Tapajos river, also an important tributary of the Amazon, where satellite-derived rainfall information is being used to run the model.

2.2 Use of the model for forecasting

When the MGB model is used for flow forecasting, variables are updated using flow data observed at certain strategic gauging sites, for which correction factors are determined using differences between observed and calculated values. These correction factors are applied to the flow from each cell immediately upstream of each strategic gauging station to the cells most upstream from it, using as a weighting the area drained by each cell. This ensures that the correction is complete in the cell in which the gauging station lies, and is least in the cells furthest upstream (Collischonn et al. 2005). The variables corrected are the flows in channels and the volumes stored in simple linear reservoirs (surface, sub-surface, and aquifer storage).

3. FIT OF THE MGB-IPH MODEL IN THE RIO GRANDE WATERSHED

To use the MGB-IPH model, the Rio Grande basin was divided into 16 sub-basins (Figure 2) with areas ranging from 1900 a 25900 km^2 (Table 1), determined by the sites of the main hydropower installations and the availability of streamflow data. Model parameters were calibrated for each sub-basin (using the corresponding outfall cell for the comparison between observed and calculated daily flows), for the period from 1970 to 1980. The period from 1981 to 2001 was used for model validation.

The model was found to fit well, with the goodness of fit measured by the coefficients Nash-Sutcliffe applied to flows (R2), to log flows (R2log), and to the volume error (ΔV), calculated from equations 1, 2 and 3 respectively. In both calibration and verification, the values obtained for R2 and R2log were about 0.9 in all the sub-basins, being only slightly less for the sub-basin Ponte Guatapara, for which the quality of streamflow data is suspect. The errors in volume between observed and fitted hydrographs were also acceptable, with values less than 0.05% during calibration and less than 7% at validation. These results from fitting the MGB-IPH model to the Rio Grande watershed are similar to results obtained in earlier work when it was fitted to other basins. The coefficients R2, R2log and ΔV are even slightly better to those obtained in other applications of the model. Figures 3 and 4 show the good agreement between calculated and observed hydrographs at the outfalls of the Furnas and Água Vermelha sub-basins respectively, for the calibration and validation phases.

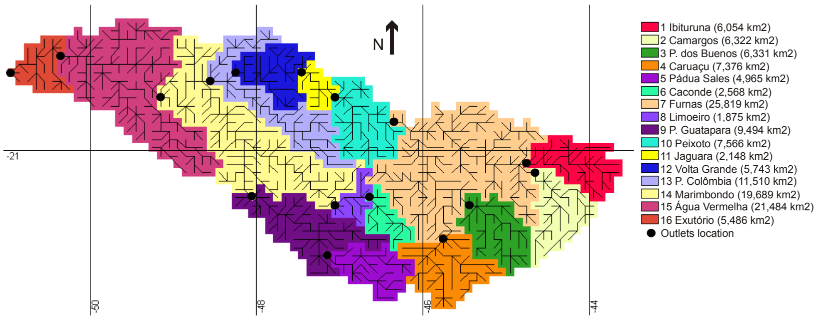


Figure 2 – Division of the Rio Grande basin into sub-basins for hydrologic modeling.

$$R2 = 1 - \frac{\sum_{i=1}^{N}(Q_{OBS} - Q_{CALC})^2}{\sum_{i=1}^{N}(Q_{OBS} - \overline{Q_{OBS}})^2} \qquad \text{equation 1}$$

$$R2\log = 1 - \frac{\sum_{i=1}^{N}(\log Q_{OBS} - \log Q_{CALC})^2}{\sum_{i=1}^{N}(\log Q_{OBS} - \overline{\log Q_{OBS}})^2} \qquad \text{equation 2}$$

$$\Delta V = \frac{\sum_{i=1}^{N} Q_{CALC} - \sum_{i=1}^{N} Q_{OBS}}{\sum_{i=1}^{N} Q_{OBS}} \qquad \text{equation 3}$$

Table 1 – Statistics showing goodness of model fit in the Rio Grande basin, during calibration and validation.

Basin	Aea	Calibration (1970-1980)			Validation (1981-2001)		
	(km2)	R2	R2log	ΔV (%)	R2	R2log	ΔV (%)
Ibituruna	6,054	0.89	0.89	<0.01	0.90	0.91	-3.6
Camargos	6,322	0.91	0.93	0.02	0.89	0.90	6.7
P. dos Buenos	6,331	0.91	0.93	<0.01	0.89	0.92	2.7
Caruaçu	7,376	0.85	0.85	0.01	0.82	0.85	-1.5
Pádua Sales	4,965	0.83	0.84	0.03	0.84	0.81	-0.9
Caconde	2,568	0.85	0.87	<0.01	0.85	0.87	3.3
Furnas	25,819	0.93	0.93	<0.01	0.91	0.91	5.9
Limoeiro	1,875	0.87	0.88	<0.01	0.89	0.90	2.2
P. Guatapara	9,494	0.76	0.69	<0.01	0.84	0.88	-3.8
Peixoto	7,566	0.93	0.92	<0.01	0.92	0.91	6.4
Jaguarão	2,148	0.93	0.92	0.01	0.92	0.91	6.1
Volta Grande	5,743	0.93	0.92	<0.01	0.92	0.92	5.1
P. Colômbia	11,510	0.93	0.92	0.02	0.92	0.93	2.4
Marimbondo	19,689	0.93	0.93	0.05	0.94	0.95	0.7
A. Vermelha	21,484	0.92	0.91	0.05	0.95	0.95	0.7
Exutório	5,486	*	*	*	*	*	*

* Sub-basin not calibrated because of poor quality stremflow data.

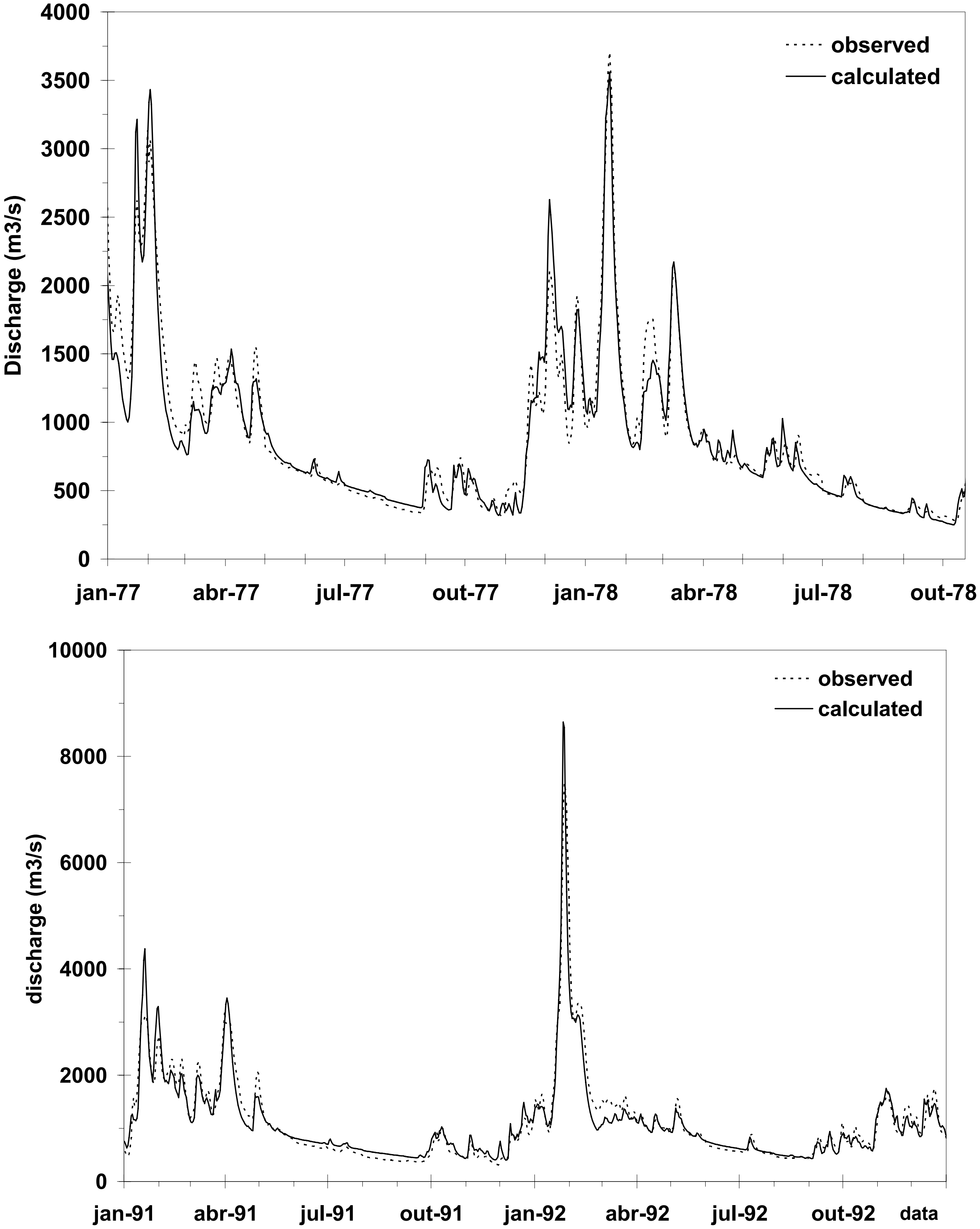


Figure 3 – Calculated and observed hydrographs during (a) calibration and (b) verification periods of the Rio Grande in the outlet of Furnas sub-basin.

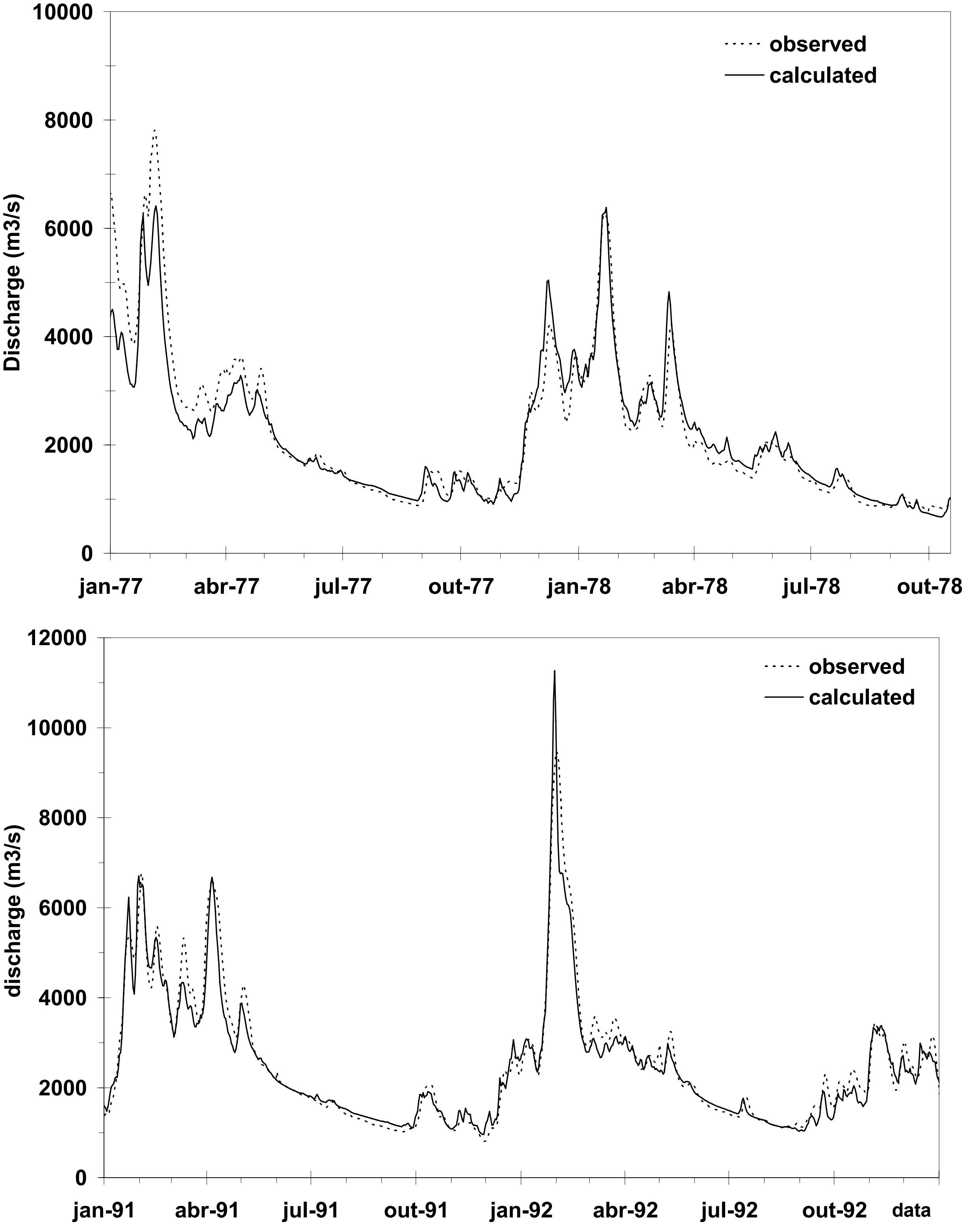


Figure 4 – Calculated and observed hydrographs during (a) calibration and (b) verification periods of the Rio Grande in the outlet of Água Vermelha sub-basin.

4. SHORT-TERM FORECASTING

4.1Rainfall forecasting

For the purpose of short-term forecasting, rainfall forecasts obtained from the ETA model were used. Their spatial resolution was 40 km (Figure 5), with a time horizon from 1 to 10 days. Forecasts were produced at weekly intervals and issued every Wednesday. Forecast rainfall was accumulated from 12:00 Z on one day to 12:00 Z on the next, corresponding in Brazilian time to an interval from 9h on one day to 9h on the next, the period for which daily rainfall is measured in Brazil. The period for which rainfall forecasts were available extended from January 1996 to November 2001.

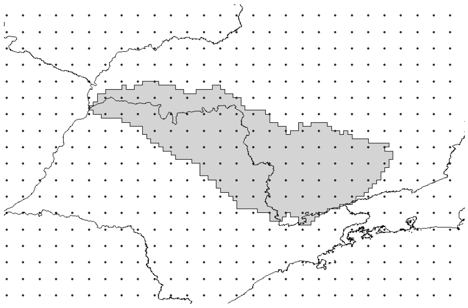
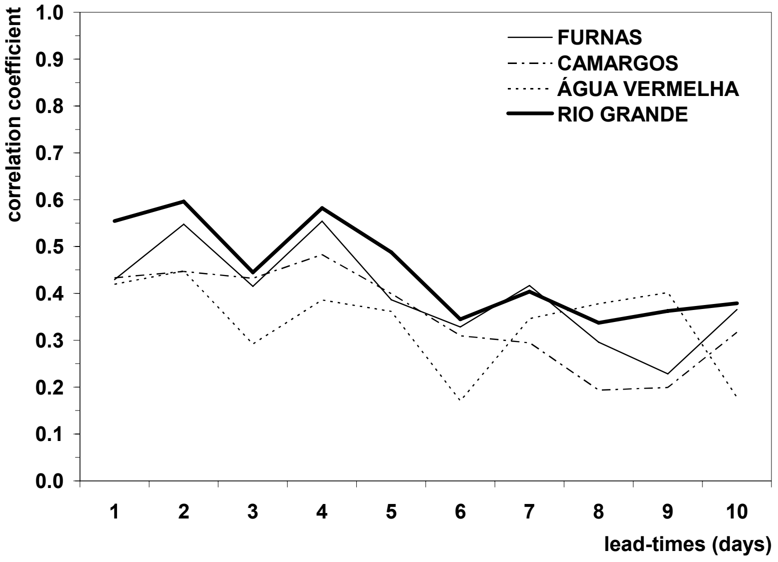


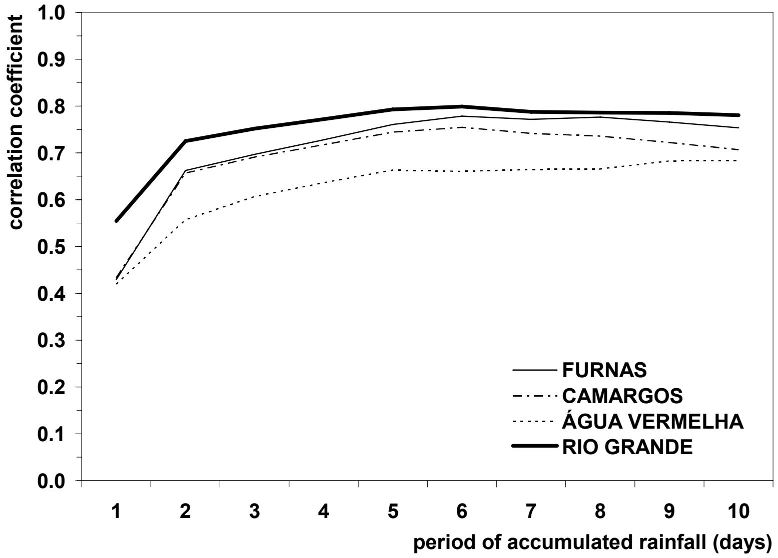
Figure 5 – Grid used by the ETA model - 40km cells of the Rio Grande watershed.

The quality of rainfall forecasts given by the ETA model was analysed for different lead-times, by comparing forecasts with mean rainfall over the whole of the Rio Grande watershed and over each sub-basin (Figure 6-a). The general tendency is for the correlation between measured and forecast rainfalls, calculated using equation 4, to decrease with increasing lag. However, contrary to what was expected, the highest correlation occurs two days after the forecast is issued, and not after a lag of one day, over the whole of the Rio Grande basin (r = 0.6). In some cases, another peak in the correlation occurred at a lag of 4 days. As expected, the correlation calculated using measured and forecast rainfall for the entire Rio Grande basin is higher than the correlations obtained for the sub-basins Camargos, Furnas e Água Vermelha, due to their greater area, which filters out some of the local errors. The graph in Figure 6-b shows how the correlation coefficient r between measured and forecast rainfall varies, when rainfall is accumulated over periods defined by increasing lead-times. As expected, the correlation increases with increasing length of accumulation period, reaching a maximum around 6 or 7 and remaining at this plateau value up to 10 days' accumulation. As before, the highest correlation is obtained for the entire Rio Grande basin. The graphs in Figure 7 show the agreement between rainfall forecasts and mean observed rainfalls for the whole of the Rio Grande basin, for accumulation periods from 3 to 7 days, over the whole of the forecasting period (January 1996 to November 2001).

$$r = \frac{\sum_{k=1}^{n} \left(P - \overline{P}\right) \cdot \left(O - \overline{O}\right)}{n \cdot S_P \cdot S_O} \qquad \text{equation 4}$$



(a)



(b)

Figure 6 – Evaluation of the quality of rainfall forecasts given by the 40km-grid ETA model, for the whole of the Rio Grande basin, and for some sub-basins: (a) correlation coefficient as a function of lead-time; (b) correlation coefficient as a function of length of period over which rainfall is accumulated.

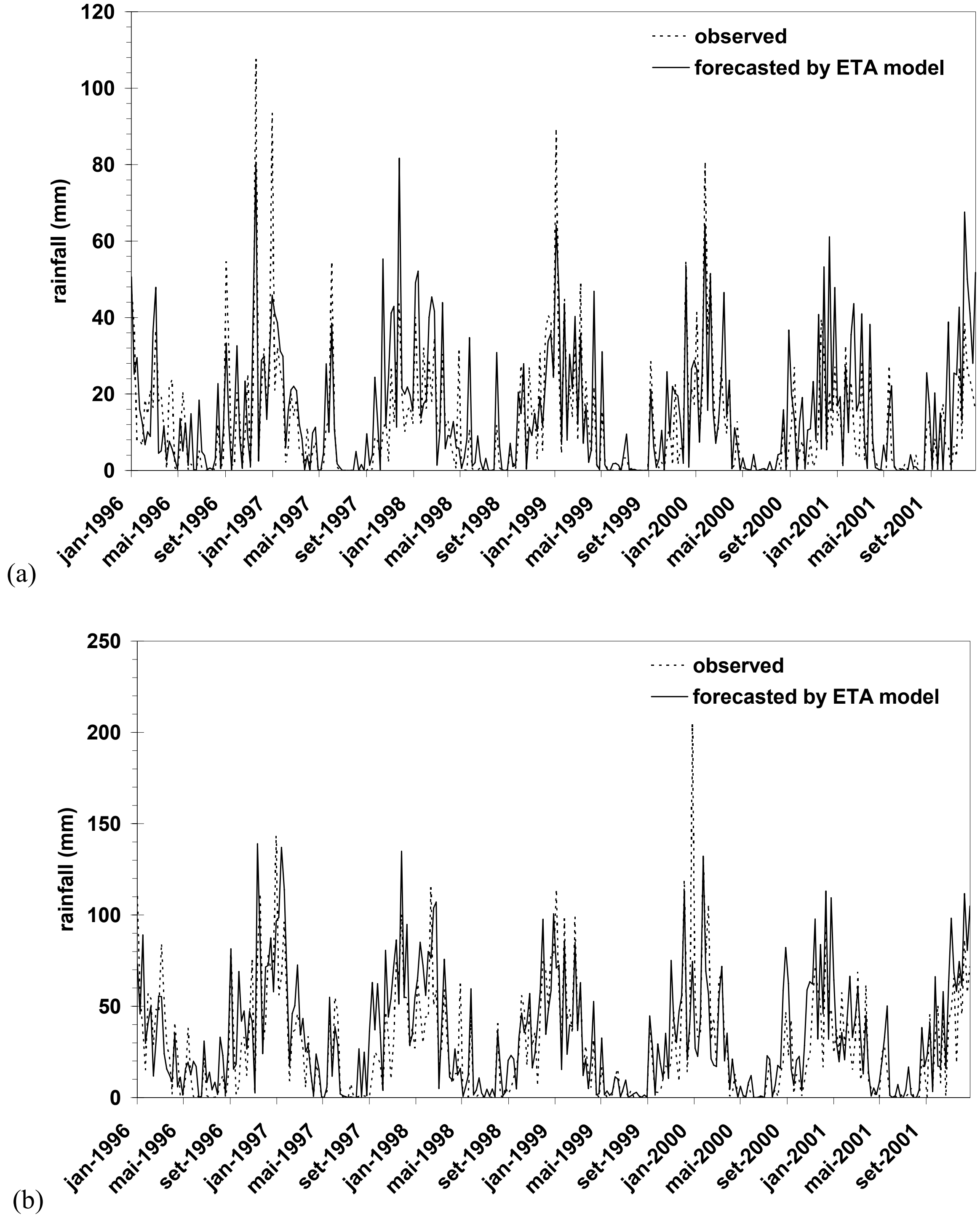


Figure 7 – Comparison between observed rainfall and rainfall forecasts given by the 40km-grid ETA model. Accumulated values of mean areal rainfall over the Rio Grande basin, (a) accumulated over 3 days, and (b) accumulated over 7 days.

4.2 Flow forecasts

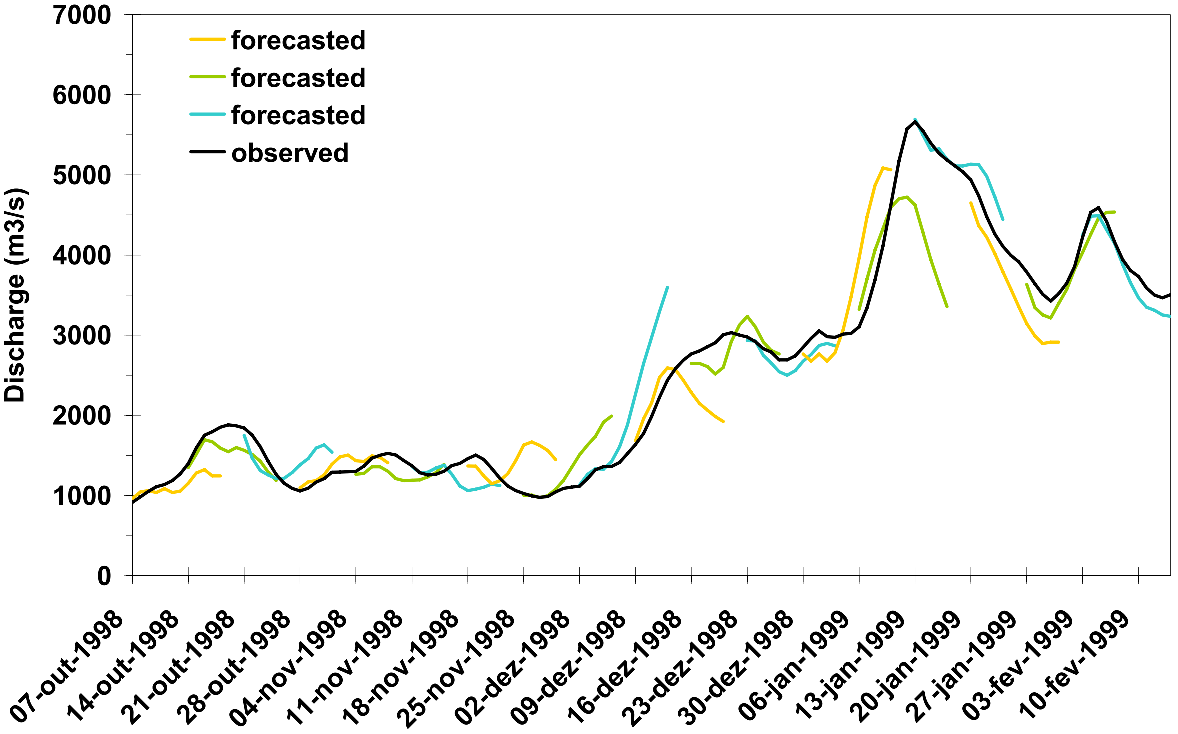
Since the rainfall forecasts given by the ETA model were issued weekly, forecasts of flow were also calculated on a weekly basis, beginning every Wednesday. The time horizon for flow forecasts extended to 12 days ahead, according to the following procedure:

i. The hydrologic model was initialized and flow was simulated using the rainfall that had been observed up to the instant at which rainfall forecasts were issued;

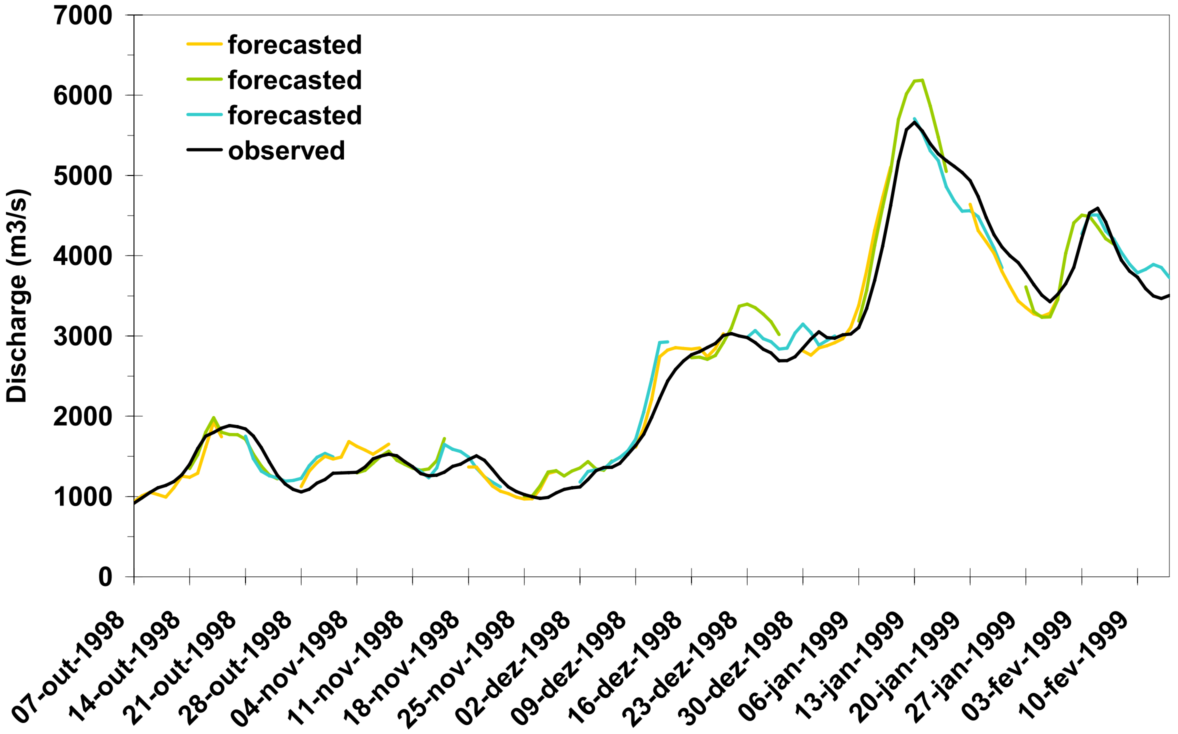
ii. Forecasts of flow were calculated for the next 10 days, using the rainfall forecasts from the ETA model, interpolated to the grid-points of the hydrologic model;

iii. Flow forecasts for the last 2 days were calculated, assuming that there was no further rainfall.

To illustrate the performance of the hydrologic model for short-term flow forecasting, results are presented for the period 7 October 1998 to 10 February 1999 for the outfall of the Água Vermelha sub-basin (Figures 8-a and 8-b). The traces on the graphs show the forecasts of hydrographs issued weekly (on Wednesdays) up to the 12-day time horizon, so that there are several such traces shown, one for each run of the hydrologic model, and shown by a different color. The agreement is fairly good, principally at forecasting the time of hydrograph rise and the magnitude of the peak corresponding to each rainfall event (Figure 8-a). For comparison, flow forecasts were also made using the rain actually observed to fall over the period of forecast, thus simulating the ideal case in which future rainfalls were exactly known (Figure 8-b). This simulation allows evaluation of the error that can be ascribed specifically to the hydrologic model, and constitutes the best possible result. Another part of the project is to evaluate the use of factors to correct for possible systematic errors in rainfall forecasts given by the ETA model; variations in the forecasting procedure to allow the hydrologic model to be updated are also being explored.



(a)



(b)

Figure 8 – Short-term flow forecasts at the outfall of the Água Vermelha sub-basin: (a) using rainfall forecats from 40km-grid ETA model (b) using rainfall actually observed over the period of forecasts. Colors are used only to distinguish between different runs of the hydrologic model, giving the 12-day-ahead forecasts initiated every Wednesday.

5. LONG-TERM FORECASTS

5.1 Rainfall forecasts

Long-term rainfall forecasts were provided by CPTEC (Centro de Previsão e de Tempo e Estudos Climáticos), part of the Brazilian national space agency INPE (Instituto Nacional de

Pesquisas Espaciais), using the CPTEC global atmospheric model AGCM. The AGCM model is based on model code used by the USA's Center for Ocean-Land-Atmosphere Studies (COLA),with adaptations (Cavalcanti et al., 2002; Marengo et al., 2005), using a spatial resolution of approximately 200 km and 28 layers in the vertical. Figure 9-a shows the spatial grid used by the AGCM model over a region including the Rio Grande watershed.

Forecasts were made using persisting TSM anomalies and 9 initial conditions (an ensemble of 9 forecasts). Starting from the initial condition, the AGCM runs in simulation mode for about 2.5 months using the observed TSM. Forecasts are then calculated for the following six months. Forecasts available for use in the Project were for intervals of 6h, and were integrated to obtain daily values for the period July 1997 to March 2003.

The climatology of the global model was analyzed over a 50-year period (1951-2001) (Marengo et al.,2005) so as to identify possible systematic errors. Comparing observed daily rainfalls with those given by the 9 members of the global climate model, it was found that the latter tends to overestimate in the eastern part of the basin (the area of the sub-basins Ibituruna, Camargos, Porto dos Buenos and Furnas) and to under-estimate in the extreme west (area of the sub-basins Marimbondo, Água Vermelha and Exutório) (Figures 10 and 11), with a slight tendency to over-estimate the mean rainfall over the basin as a whole (Figure 9-b).

To use the forecast rainfall as input to the hydrologic model, systematic errors in rainfall forecasts must be corrected or minimized. A statistical technique based on a transformation of the probability distribution (Hay and Clark, 2003; Wood et al., 2002) was therefore used. In the present study, the probability distributions of observed rainfall and of model climatology were used, with separate probability distributions calculated for each month of the year. For each cell of the hydrologic model, 12 probability distributions of daily observed rainfalls were therefore computed, together with 12 probability distributions of daily rainfalls given by model climatology. Both observed daily rainfalls and climatology rainfalls were interpolated to obtain values for grid-points in the hydrologic model, interpolating according to the inverse of the squared distance. The correction procedure was as follows (Figure 12): (i) given a daily rainfall forecast, interpolated to give a value at the cell of the hydrologic model, a probability associated with it was determined from the climatological cumulative probability distribution; (ii) the corrected rainfall was given by the value having the same cumulative probability in the cumulative probability distribution of observed daily rainfalls.

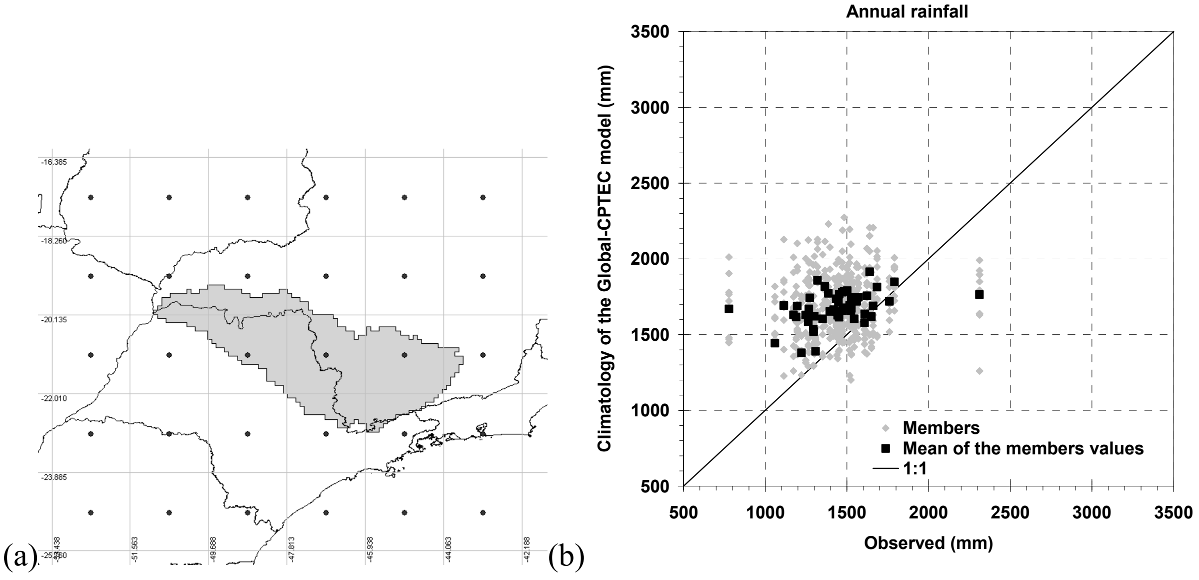


Figure 9 – (a) Grid used by the Global-CPTEC model in the region covering the basin of the Rio Grande; (b) observed mean annual rainfalls for the entire Rio Grande basin, plotted against annual rainfall calculated from the climatology of the Global-CPTEC model. In (b), grey points refer to the 9 members of the Global-CPTEC model, and black points are the means of the 9 members.

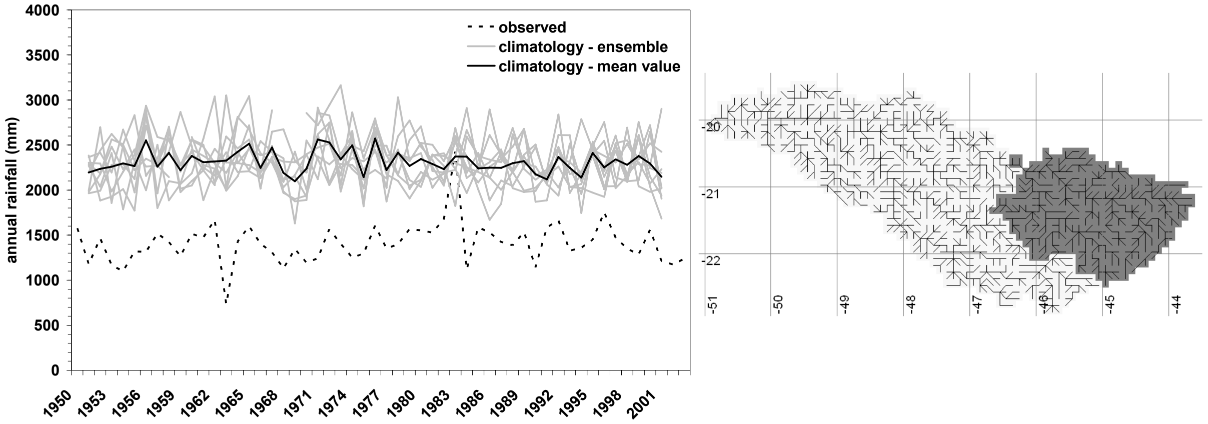


Figure 10 – (a) Comparison between observed annual rainfall and annual rainfall given by climatology of the global model, using the ensemble of forecasts and the mean of the ensemble. Mean values are over the area shown in grey (sub-basins Ibituruna, Camargos, Porto dos Buenos and Furnas) in (b).

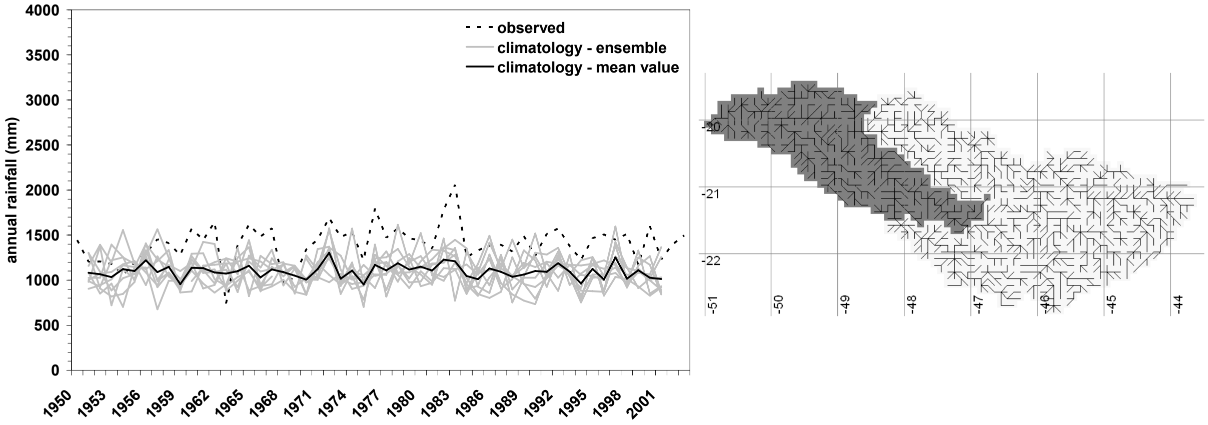


Figure 11 – (a) Comparison between observed annual rainfall and annual rainfall given by climatology of the global model, using the ensemble of forecasts and the mean of the ensemble. Mean values are over the area shown in grey (sub-basins Marimbondo, Porto Colômbia and Exutório) in (b).

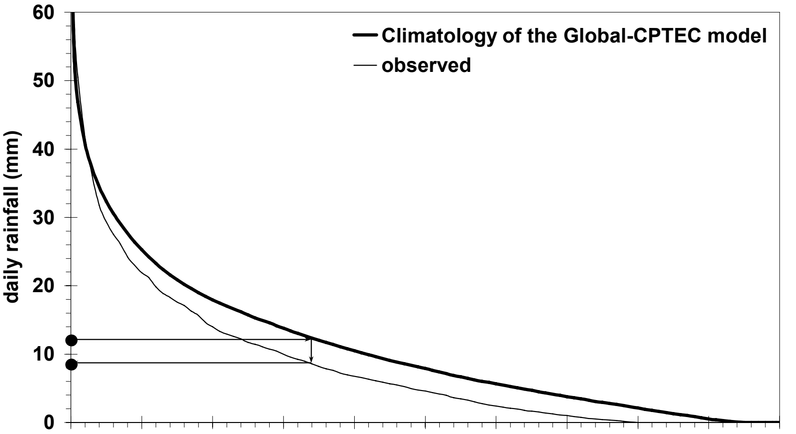


Figure 12 – Procedure for correcting systematic errors in rainfall forecasts using cumulative probability distributions of observed and climatological daily rainfalls obtained from the global model. The example shown is for the month of January at the outfall cell of the Furnas sub-basin: for a forecast rainfall of 12 mm on a given day, the cumulative probability curve obtained from model climatology shows a probability of 34%; this probability corresponds to a rainfall of 8 mm on the cumulative probability distribution of observed values. The corrected rainfall forecast is therefore 8 mm.

5.2 Flow forecasting

Long-term flow-forecasting uses rainfall forecasts from CPTEC's AGCM model and begin on the first day of each month, extending for the following six months. Up to the last day before forecasts begin, the hydrologic model is run using observed rainfall data, interpolated to the model grid-points. The model state variables are updated to the day immediately preceding the first forecast day using the weighting procedure described earlier. For example, to obtain flow forecasts for the period 1 October 1997 to 28 March 1998 (Figures 13 and 14), the hydrologic model was initialized on 1 January 1990 and run with observed rainfall data as input up to 30 September 1997. From 1 October 1997 flow forecasts are calculated using rainfall forecasts from the global model, after statistical correction using the cumulative probability distributions.

Figures 13 and 14 show forecasts of daily flow at the outfalls of the Furnas and Água Vermelha sub-basins respectively, from 1 October 1997 to 28 March 1998. In both figures, flow forecasts are shown with and without statistical correction of rainfalls. It can be seen that the statistical correction obtained from the cumulative probability distributions improves the flow forecast, although even after correction the forecast is not very satisfactory. Within the Project context, alternative procedures for correcting rainfall forecasts are being studied so as to improve flow forecasts. The grey band in Figures 13 and 14 shows the interval between the highest and lowest flow forecasts obtained from the ensemble of rainfall forecasts, together with the mean of the forecasts obtained from the ensemble. The band shows a relatively wide dispersion of flow forecasts, but it includes the observed flow sequence after the statistical correction of rainfall sequences. The mean value of the ensemble of forecasts can be considered satisfactory when compared with observed flows, given the long lead-time of forecasts (extending up to 6 months).

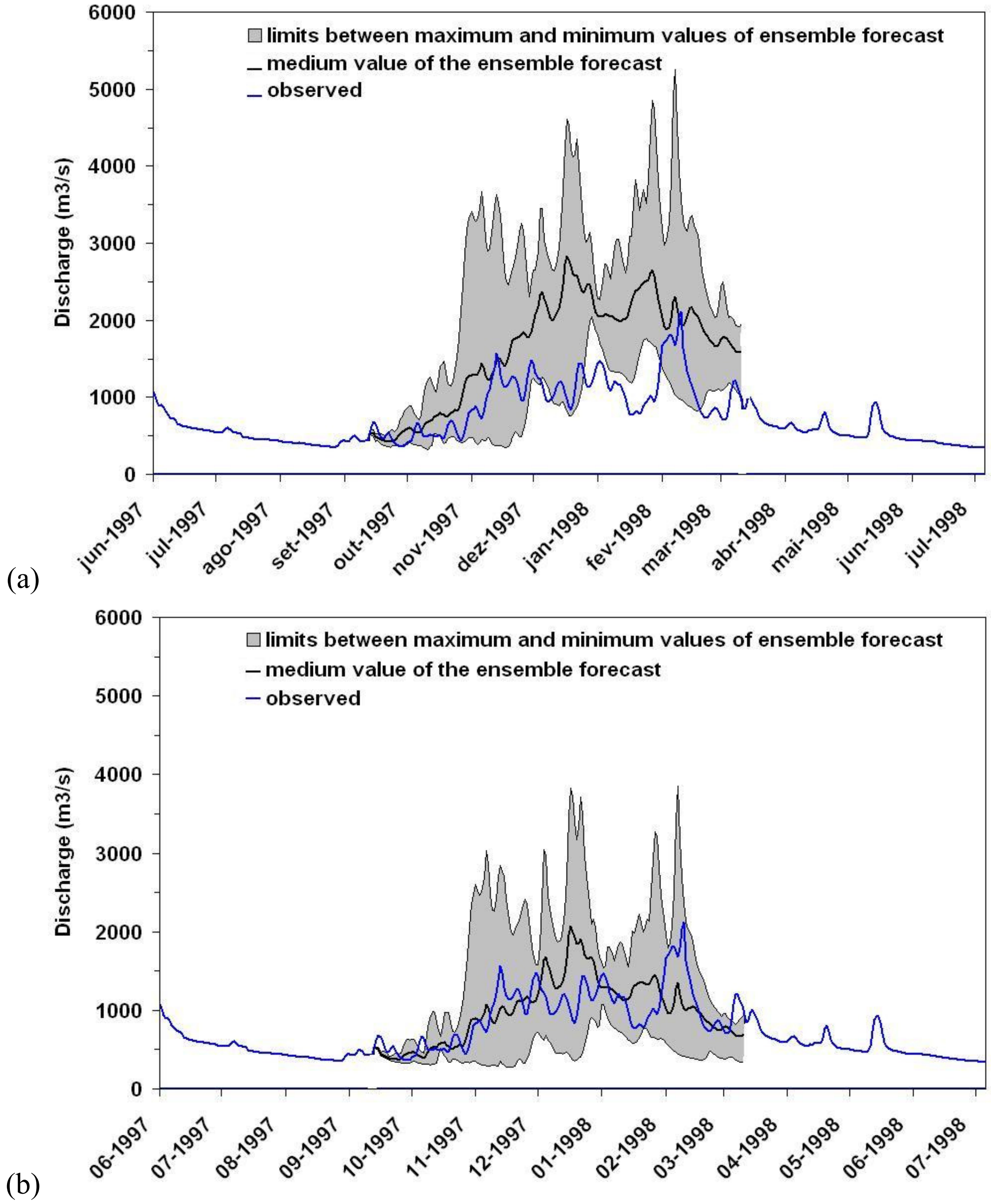


Figure 13 – Long-term flow forecasts (up to lead-time 6 months) obtained from rainfall forecasts given by the CPTEC AGCM model (a) without statistical correction of rainfall forecasts using cumulative probability distributions, (b) with statistical correction. Forecasts start on 1 October 1997, for the outfall from the Furnas sub-basin.

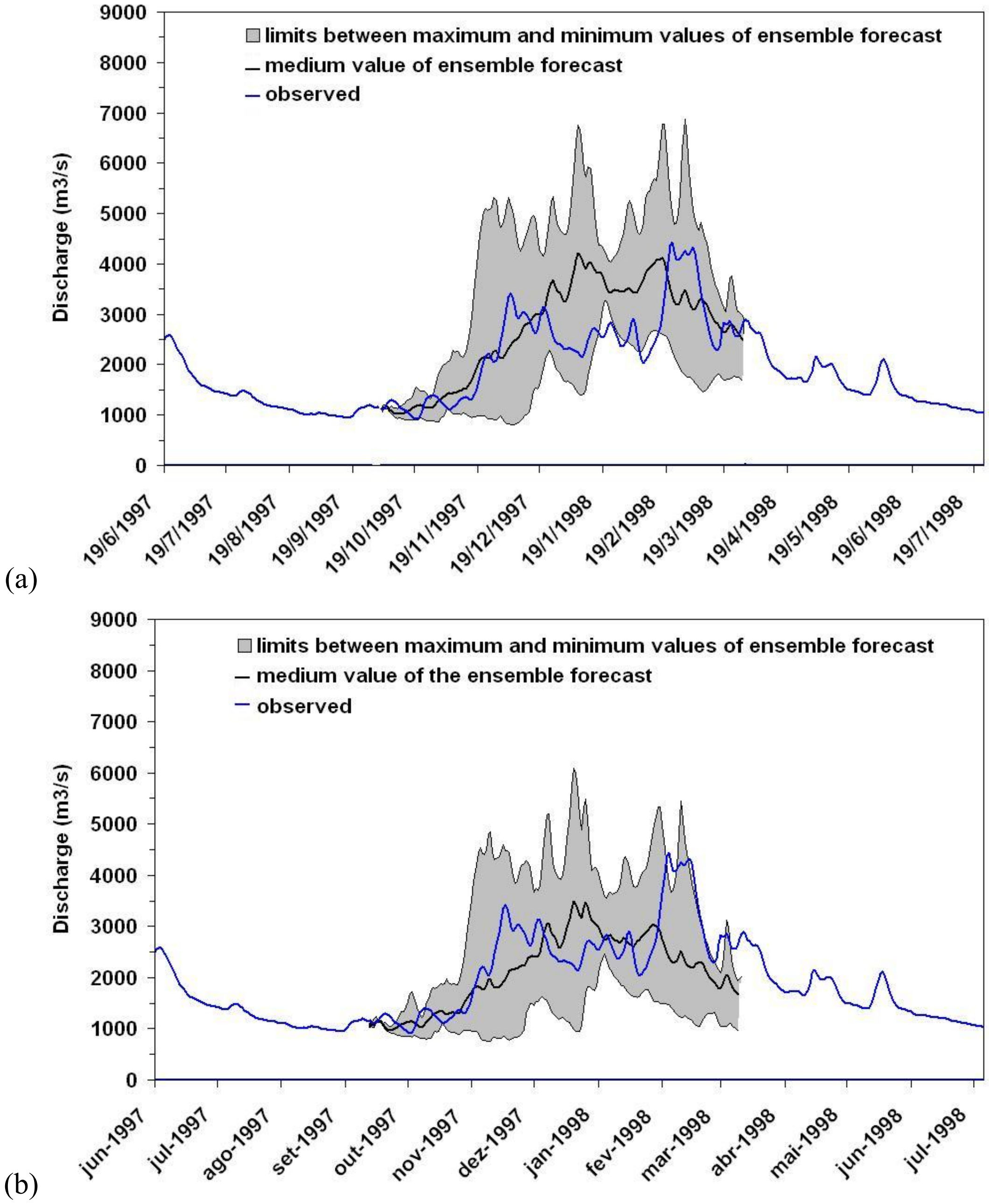


Figure 14 – Long-term flow forecasts (up to lead-time 6 months) obtained from rainfall forecasts given by the CPTEC AGCM model (a) without statistical correction of rainfall forecasts using cumulative probability distributions, (b) with statistical correction. Forecasts start on 1 October 1997, for the outfall from the Agua Vermelha sub-basin.

6. REFERENCES

Bremicker, M. (1998) Aufbau eines Wasserhaushaltsmodells für das Weser und das Ostsee Einzugsgebi*et al*s Baustein eines Atmosphären-Hydrologie-Modells. Dissertation Doktorgrad, Geowissenschaftlicher Fakultät der Albert-Ludwigs-Universität. Freiburg. Germany.

Cavalcanti I F A, Marengo J.A., Satyamurty P, Trosnikov I., Bonatti J, Nobre C.A., D'Almeida, C., Sampaio G., Castro C A C, Camargo H, Sanches MB, Global climatological features in a simulation using CPTEC/COLA AGCM. J Climate, 15: 2965–2988, 2002.

Collischonn, W. & Tucci, C.E.M. (2001) Large Basins Hydrologic Simulation (in Portuguese). *Brazilian Journal of Water Resources* **6**(1), 15-35.

Collischonn, W; Haas, R.; Andreolli, I. & Tucci, C.E.M (2005). Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *J. Hydrol.* **205**, 87-98.

Hay LE and Clark M.P (2003). Use of statistically and dynamically downscaled atmospheric model output for hydrologic simulations in three mountainous basins in the western United States. Journal of Hydrology. V. 282, p.56-75.

Kouwen, N., Soulis E.D., Pietroniro A., Donald J., & Harrington R.A. (1993) Grouped response units for distributed hydrologic modeling. *J. of Water Resour. Planning and Maagemen.,* **119**(3), 289-305.

Liang, X.; Lettenmaier, D.P.; Wood, E.F. & Burges, S. J. (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res.* **99**(7), 14415-14428.

Marengo J A, Alves L M, Camargo H (2005) An overview of global climate predictability at seasonal to interannual time scales, GEWEX Newsletter.

Miller, W. A.& Cunge, J. A. (1975) Simplified equations of the unsteady flow in open channels. In: *Unsteady flow in open channels* (ed. by Mahmood, K. & Yevjevich, V). Water Resources Publications. Fort Collins.

Ribeiro A., Vieira da Silva, R., Collischonn, W. & Carlos E. M. Tucci. (2005) Hydrological Modelling in Amazonia - Use of the MGB-IPH Model and Alternative Data Base. *Proceedings of VII IAHS Scientific Assembly, Foz do Iguaçu*, Brasil.

Todini, E. (1996) The ARNO rainfall – runoff model. *J. Hydrol.* **175**, 293-338.

Tucci, C.E.M.; Clarke, R.T.; Collischonn, W.; Dias, P.L.S. & Sampaio, G.O. (2003) Long term flow forecast based on climate and hydrological modeling: Uruguay river basin. *Water Resour. Res.* **39**, n.7, 3(1-11).

Tucci, C.E.M.; Marengo, J.A.; Silva Dias, P.L.; Collischonn, W.; Silva, B.C.; Clarke, R.T.; Cardoso, A.O.; Juarez, R.N.; Sampaio, G.; Chan, C.S & Tomasella, J. (2005) Streamflow Forecasting in São Francisco River Basin Based in the Climatic Forecasting. *Technical Report ANEEL/WMO/98/00*. Porto Alegre, Brazil (in Portuguese).

Wigmosta, M.S.; Vail, L.W. & Lettenmaier, D.P. (1994) A distributed hydrology-vegetation model for complex terrain. *Water Resour. Res.,* **30(**6), 1665-1679.

Wood E.F, Maurer, E.P.; Kumar A & Lettenmaier DP (2002) Long-range experimental hydrologic forecasting for the eastern United States. Journal of Geophysical Research. V.107, n.20, p.6(1-15).