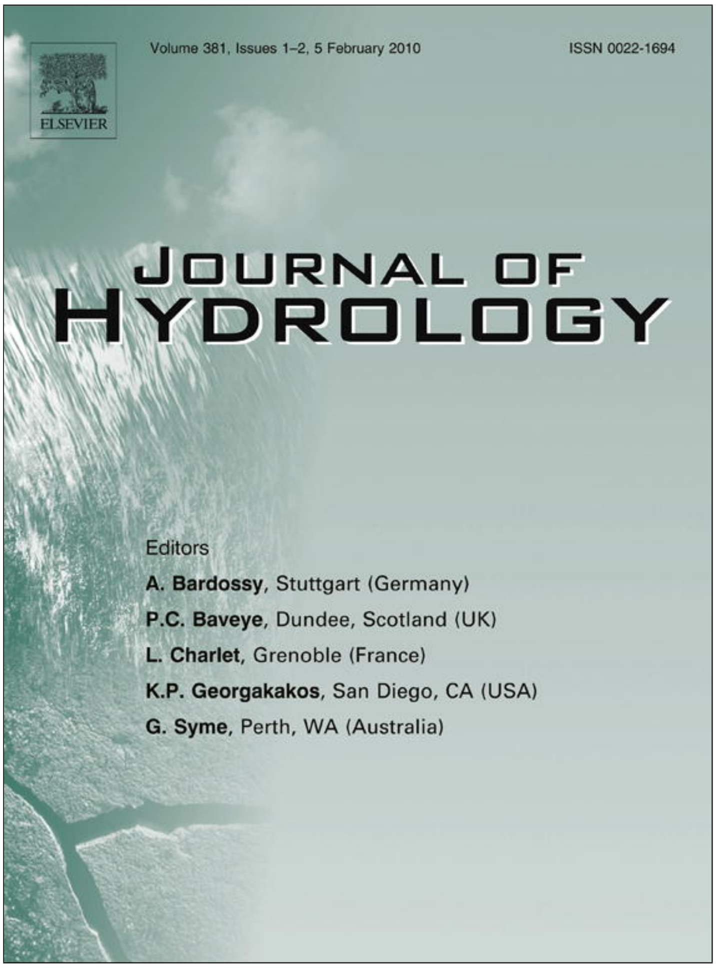

Volume 381, Issues 1–2, 5 February 2010 ISSN 0022-1694

JOURNAL OF HYDROLOGY

Journal of Hydrology 381 (2010) 65–75

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Climate informed long term seasonal forecasts of hydroenergy inflow for the Brazilian hydropower system ☆

Carlos H.R. Lima *, Upmanu Lall

Columbia University, Dept. of Earth & Environmental Engineering, 500 W, 120th street, 918 Mudd, NY 10027, USA

ARTICLE INFO

Article history:
Received 5 April 2009
Received in revised form 2 November 2009
Accepted 18 November 2009

This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Ercan Kahya, Associate Editor

Keywords:
Seasonal forecasts
Energy
Climate predictors

SUMMARY

Efficient management of water and energy is an important goal of sustainable development for any nation. Streamflow forecasts, have been used in complex optimization models to maximize water use efficiency and electrical energy production. In this paper we develop a statistical model for the long term forecasts of hydroenergy inflow into the Brazilian hydropower system, which consists of more than 70 hydropower reservoirs. At present, the planning of reservoir operation and energy production in Brazil is made with no reliable long term (one season or longer lead times) streamflow forecasts. Here we use the NINO3 index and the main modes of the tropical Pacific thermocline structure as climate predictors in order to achieve skillfull forecasts at long leads. Cross-validated results show that about 50% of the total hydroenergy inflow can be predicted with moderate accuracy up to 20 month lead time.

Introduction

With more than 70 interconnected (hydraulically and through transmission lines of electrical energy) hydropower reservoirs across the country, the Brazilian hydropower system uses complex optimization models (e.g. Barros et al., 2003) to produce electrical energy to maximize reliability and minimize cost. For instance, if in September most hydropower plants have low levels of energy storage (i.e. low levels of water head in the hydropower reservoirs and hence more water needed to produce a unit of energy) and if the upcoming rainy season (January–April) is expected to be drier than normal, then the System National Operator (ONS) usually starts producing higher cost energy from stand-by thermal plants in order to reduce the hydroenergy production and avoid a complete depletion of the reservoirs, which would lead to a more dramatic rise in the cost of electrical energy in the subsequent season.

However, the use of thermal plants and reduction in hydropower production before an upcoming above normal wet season may lead to unnecessary losses of hydroenergy through spillage and evaporation and to an avoidable increase in the electrical energy cost due to the use of thermal plants. In order to make better decisions during such situations, forecasts of inflow (e.g. Costa et al., 2003; Guilhon et al., 2007; Maceira et al., 2005; Silva et al., 2007) have been used extensively by ONS to anticipate unusual scenarios of energy supply and demands. In particular, energy supply forecasts have been made through the use of streamflow forecasts for the short (daily, weekly), medium (monthly) and long term (seasonal, annual).

More recently, the Brazilian community of researchers and hydrologists has attempted to improve the classical, auto-regressive moving average (ARMA) based streamflow model utilized by ONS (for instance, Guilhon et al., 2007). Most alternatives have been based on the coupling of dynamical models of rainfall (general or regional circulation models) and hydrological models (e.g. Silva et al., 2007). Since dynamical models of rainfall do not produce reliable forecasts with more than two weeks in advance, their use has been limited to short term forecasts. Recently developed mathematical tools (e.g. artificial neural networks) have also been used for seasonal streamflow forecasts, but again with limited success (e.g. Guilhon et al., 2007).

Here we use the leading modes of the Tropical Pacific thermocline structure identified using maximum variance unfolding (MVU, see Weinberger and Saul, 2006, 2004) and also used (Lima et al., 2009) for long term forecasts of El Niño-Southern Oscillation (ENSO) events, as climate predictors of seasonal hydroenergy inflow across Brazil. Since ENSO has a marked influence on rainfall and streamflow patterns in South America (e.g. Diaz et al., 1998; Grimm, 2004; Grimm et al., 2000; Grimm et al., 1998), it is natural to consider a climate predictor originally developed for ENSO forecasts in a prediction model for streamflow whose inter-annual

☆ Supported by fulbright and CAPES(Brazil) through Grant # 1650-041, and the Pepsico Foundation.

* Corresponding author. Tel.: +1 212 854 7219.
E-mail addresses: chr2107@columbia.edu (C.H.R. Lima), ula2@columbia.edu (U. Lall).

0022-1694/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.11.026

variability is directly affected by ENSO. This paper is organized as follows: in the next section we describe the hydroclimate data. In "Methodology" we present the methodology used to obtain the seasonal hydroenergy inflow from streamflow series. The forecast model framework is also presented in this section. Cross-validated results are finally shown in "Results".

Data

Streamflow data

Naturalized monthly series of streamflows from 54 hydropower reservoirs in Brazil (location shown in Fig. 2) are provided by the Brazilian National Operator of the Electrical System (ONS), which is the institution in charge of defining operational rules for the Brazilian system of hydropower reservoirs. The time series cover the January 1931–December 2006 period and span a large range of power capacities (80–14000 MW) and catchment areas ($322-823555\ \mathrm{km}^2$). A consolidation and consistency process is used by ONS to obtain the naturalized flows from artificial and natural streamflow gauges. Reservoir operations upstream of the streamflow gauge are removed from the original series whereas evaporation from the hydropower reservoir and water withdraws across the reservoir basin are estimated and added to the original series. Some streamflow gauges are atypical and involve pumping, transpositions between river and canals, bypasses, etc., adding more complexity to the consistency process. More details can be found in ONS (2007).

These are naturalized time series, i.e., any anthropogenic effect upstream of the river flow gauge, such as reservoir operations and water withdraws, has been estimated and removed from the original series. The time series have been revised by the Brazilian National Water Agency (ANA) and do not have any missing values.

Climate data

As a climate predictor, we use the extended NINO3 index (Kaplan et al., 1998; Reynolds and Smith, 1994), available through the International Research Institute for Climate and Society (IRI) dataset website (http://iridl.ldeo.columbia.edu/SOURCES/.Indices/.nino/.EXTENDED/.NINO3/). The NINO3 data set covers the period from January 1980–December 2006. It is defined as the monthly mean sea surface temperature (SST) anomalies (with annual cycle removed) averaged over the geographical area 5°N–5°S latitude, 150°W–90°W longitude.

We also use the first three leading modes of the D_{20} data set, which represents the thermocline depth of the Tropical Pacific ocean at the 20 °C isotherm and is a proxy for the Tropical Pacific thermocline and heat content. The D_{20} data is derived from a model-based ocean analysis system (Ji et al., 1995; Ji and Smith, 1995; Behringer et al., 1998) and is available at the IRI website (http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.EMC/.CMB/.Pacific/.monthly/.D20eq/). The leading modes were obtained using standard principal component analysis (PCA, see Wilks (2006) for some definitions) and the nonlinear maximum variance unfolding (MVU, see Weinberger and Saul, 2006; Weinberger et al., 2004). Details about MVU and the physical interpretation of the D_{20} leading modes can be seen in Lima et al. (2009).

Methodology

Clustering and the average energy inflow

Given the large number of interconnected hydropower reservoirs, the first challenge in defining the best operational policies for the hydropower system in Brazil is to build a forecast model that is able to reproduce the entire streamflow spatio-temporal variability observed in the historical data. Such models are complex and have several parameters (including large covariance matrices) to estimate given a limited amount of data, which in turn leads inevitably to large uncertainties in model outputs and consequently unreliable models. One alternative is to consider the concept of the equivalent reservoir of energy, where one defines large clusters of hydropower reservoirs, optimizes the equivalent energy of each cluster and thereafter, given the optimal energy to be produced for each cluster, optimizes the individual energy production of the reservoirs within each cluster. It turns out that this approach significantly reduces the dimension of the problem and consequently the computational effort to optimize the entire system.

Several techniques to identify clusters of streamflow homogeneous regions and for watershed classification have been used in hydrological studies. Among them, principal component analysis (PCA) is arguably one of the most common methods used to identify clusters within data points (e.g. Chiang et al., 2002; Kahya et al., 2008), especially when the data set lie on large dimensional spaces.We use the K-means algorithm, a standard, robust and efficient technique in cluster analysis (Hastie et al., 2001) with several applications in hydrology (e.g. Isik and Singh, 2008) to reduce the problem dimension. The clusters are found by minimizing the distance between the center of the cluster and the data points that belong to that cluster. Here we choose the inflow seasonality of each hydropower reservoir as the classification criterion. This leads to clusters with similar (up to some scale) rainfall and streamflows patterns and under like climate forcings, so that forecasts of the equivalent energy of each cluster can be downscaled to each reservoir within the cluster.

In order to identify the wet and dry seasons for each streamflow site, we first define the seasonal index for site s and month j as:

$$\rho_{sj} = \frac{\hat{y}_{sj}}{\hat{y}_s} \tag{1}$$

where $\hat{y}_{sj}$ is the median flow of site s for month j and $\hat{y}_s$ is the median flow for site s across the entire historical record. The wet season for a given reservoir can be defined for values of j in which $\rho_{sj} \geqslant 1$ whereas the dry season may be defined whenever $\rho_{sj} < 1$.

The seasonal index ρ_{sj}, with $j = 1, \ldots, 12$, is considered the *signature* of each hydropower site. Fig. 1 displays the seasonal index ρ_{sj} and the clusters obtained after applying K-means ($K = 4$) to ρ_{sj}, whose computation was based on the streamflow series from 1931 to 1979. The first cluster of reservoirs shows well defined wet and dry seasons. The timings of these seasons are similar to those of the third cluster, which has a less pronounced seasonality. The second cluster has the wet season in the second half of the year. Finally, the fourth cluster with only one reservoir has a peak flow in April but the wet season period is similar to that of clusters 1 and 3. Note that in all clusters the seasonal index is relatively tight, i.e., the reservoirs within each cluster have very similar seasonal patterns of inflow.

The geographical location of the clusters (Fig. 2) is clearly demarcated. The first cluster of hydropower reservoirs is mostly located in the Southeast and Southern Northeast regions of Brazil, which have a rainy season characterized by the South-Atlantic Convergence Zone (Barros et al., 2000; Carvalho et al., 2004; Lenters and Cook, 1995). Cluster 3 is located in a transition region between cluster 1 and cluster 2, whose rainfall is largely influenced by cold fronts as well as ENSO (Grimm et al., 2000, 1998). Finally, cluster 4 (Tucurui Hydropower reservoir) is located in North Brazil. Giving its large catchment area ($757577\ \mathrm{km}^2$) that extends up to

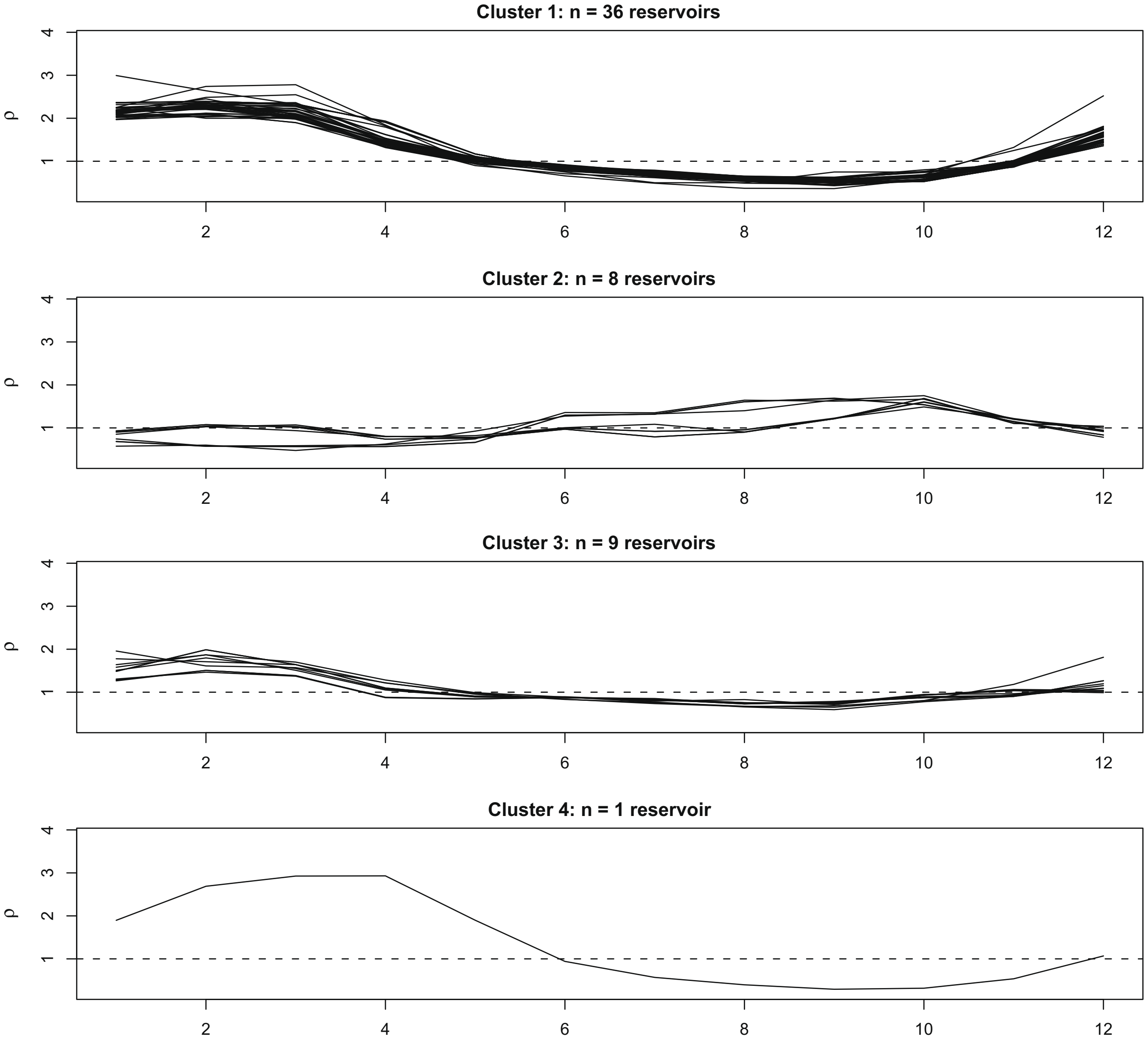

Fig. 1. Seasonal index for reservoirs within each cluster.

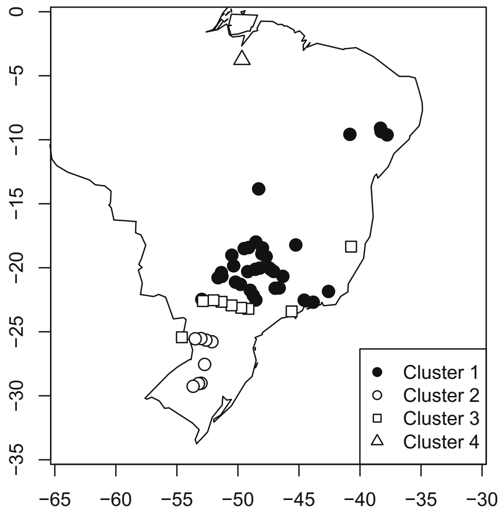

Fig. 2. Four clusters found for the 54 Brazilian hydropower reservoirs.

central Brazil, much of this cluster's energy flow is also influenced by the rainfall patterns associated with cluster 1.

Based on the literature of remote climate influences on rainfall patterns in Brazil (Barros et al., 2000; Lenters and Cook, 1995; Grimm et al., 2000), it turns out that these four clusters' series are effectively representing the interannual variability of the main rainfall regimes in Brazil. Hence, for each cluster k the average monthly inflow of hydroenergy can be defined as:

$$q_{kij} = \sum_{s=1}^{S} \mathbf{1}_k(s) \cdot y_{sij} \cdot \bar{h}_s \cdot \gamma \cdot 10^{-6} \qquad (2)$$

where q_{kij} is the average hydroenergy inflow in megawatts (MW) of cluster k at month j of year i, y_{sij} is the correspondent inflow (in m^3/s) of hydropower reservoir s, S is the total number of reservoirs, $\bar{h}_s$ is the average head (based on reservoir technical specifications) of reservoir s, $\gamma = 9.81 \times 10^3$ is the specific weight of water in SI units and $\mathbf{1}_k(s)$ is the indicator function given by:

$$\mathbf{1}_k(s) = \begin{cases} 1 & \text{if reservoir } s \text{ belongs to cluster } k \\ 0 & \text{otherwise} \end{cases} \qquad (3)$$

Fig. 3 shows the monthly energy flow q_{kij} for each cluster k. Based on a weighted average of the long term (1931–1979) mean energy flow, cluster 1, with 36 reservoirs, contributes with more than 50% of the total energy produced by the hydropower system. As we expected from Fig. 1, the second cluster has a very weak seasonal cycle. No monotonic trends appear visible in any of the cluster time series.

Finally, we can define the seasonal index of cluster k as the average of the seasonal indexes of the reservoirs that belong to cluster k:

$$\bar{\rho}_{kj} = \frac{1}{n_k}\sum_{s=1}^{S}\mathbf{1}_k(s)\rho_{sj} \tag{4}$$

where n_k is the total number of hydropower reservoirs within cluster k.

The wet season period for each cluster is then defined as the months in which $\bar{\rho}_{kj} \geqslant 1$. Table 1 shows the months corresponding to the wet and dry seasons for each cluster. For practical purposes, we define the wet season for clusters 1, 3 and 4 at year i as the period from December of year $i-1$ to May (or April in the case of cluster 3) of year i.

The wet season flow, which is more important than the dry season flow in terms of defining the operational policies for the annual planning of the hydropower system, is defined here as the average monthly flow of the wet season:

$$\bar{q}_{ki} = \frac{1}{j_e - j_0 + 1}\sum_{j=j_0}^{j_e} q_{kij} \tag{5}$$

where j_0 and j_e are, respectively, the start month and end month of the wet season of cluster k.

Table 1
Wet and dry seasons defined for each cluster according to Eq. (4).

Cluster	Wet season	Dry season
1	Dec–May	Jun–Nov
2	Jun–Nov	Dec–May
3	Dec–Apr	May–Nov
4	Dec–May	Jun–Nov

Analysis of the auto-correlation function (Brockwell and Davis, 2002) of $\bar{q}_{ki}$ shows no sign of lagged correlations (i.e. no annual persistence), suggesting that, for a fixed k, the series $\bar{q}_{ki}$ is independent across years. On the other hand, the cross-correlation functions (Fig. 4) of $\bar{q}_{ki}$ show statistically significant correlations across the clusters. As expected from the rainfall patterns across Brazil, one has significant correlation among clusters 1, 3 and 4 (open circle, filled square and open triangle curves showed in Fig. 4) and an uncorrelated behavior of cluster 2. More interesting, however, is the negative lagged correlation between the northern clusters (1 and 3) and cluster 2 (filled circle and open square curves in Fig. 4). It suggests a somewhat *seesaw* structure lagged by one and two years, where an above (below) than normal wet season in clusters 1 and 3 at year i is associated with a below (above) than normal wet season in cluster 2 at years $i+1$ and $i+2$. Note that the wet season periods are not the same (see Table 1).

Climate predictors and the forecast model

In order to take into account the effects of large scale climate forcings on rainfall and streamflow patterns across Brazil, particularly the remote influence of the tropical Pacific SST (a description

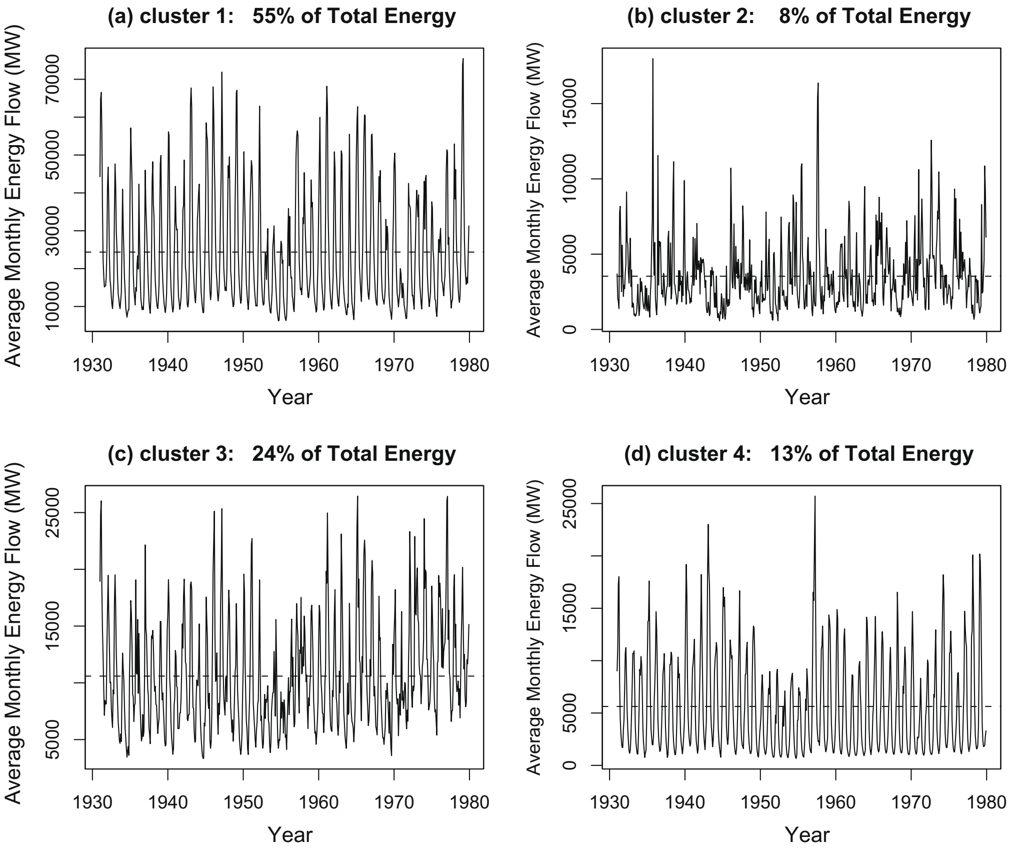

Fig. 3. Time series of monthly energy inflow (in MW) of each cluster of hydropower reservoirs. The contribution of each cluster to the total energy of the system is a weighted average based on the mean energy flow (dashed line) for the 1931–1979 period.

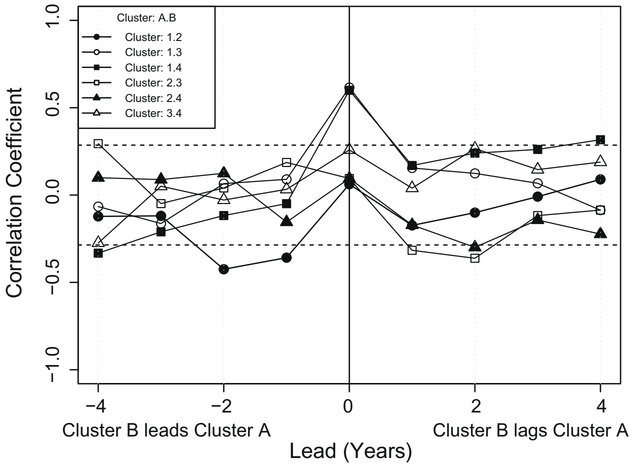

Fig. 4. Cross-correlation function of wet season energy flow across clusters. Correlations outside the dashed line interval are statistically significant at the 5% significance level. Lead and lag times are given in years. The wet season period for each cluster in shown in Table 1.

of the climate patterns and teleconnections of the region of interest can be found in Ropelewski and Halpert (1987), Diaz et al. (1998), Grimm et al. (2000), Carvalho et al. (2004), Grimm (2004), Vera et al. (2006)), we use the NINO3 index and the tropical Pacific thermocline structure as predictors in a forecast model for the wet season hydroenergy flow of each hydropower cluster. As a proxy for the tropical Pacific thermocline depth, we use here the three main modes of the Pacific thermocline data D_{20}. Both linear (obtaining through PCA) and non-linear (obtained through MVU) D_{20} modes are evaluated here. Lima et al. (2009) provides a detailed description of the first three modes of the D_{20} data set.

As a preliminary analysis, Fig. 5 shows the lagged correlations of the wet season energy flow and the climate predictors NINO3 and first three MVU modes $[Y_1\, Y_2\, Y_3]$ of D_{20}. The effect of ENSO on the energy flow appears statistically significant for short lags in all but cluster 1, which shows some influence but it is not statistically significant. The correlations with the first MVU mode Y_1 appear statistically significant only for the first cluster. A detailed look at both time series (not show here) suggests that this large correlation between them is probably associated with a monotonic trend observed in both series. The second MVU mode, which turns out to be a very good predictor for NINO3 (see Lima et al., 2009), also has significant lagged correlations with all clusters, especially clusters 2 and 3.

Lagged correlations with the first three PCs of the D_{20} data are shown in Fig. 6. Unlike the MVU modes, the PCs do not show any significant lagged correlation with the wet season flow of any of the first three clusters.

A forecast model for the wet season energy flow can now be built based on the cross-correlation functions of response variable and predictors (Figs. 5 and 6). Since Y_1 and Y_3 do not show significant correlations with $\bar{q}_{ki}$ (except for the trend correlation of cluster 1 and the first MVU mode), we consider here that the average wet season flow can be modeled as a function of NINO3 and Y_2, both lagged by some time τ_1 and τ_2:

$$\bar{q}_i = f(\mathrm{NINO3}(t_i - \tau_1(\tau)), Y_2(t_i - \tau_2(\tau))) \qquad (6)$$

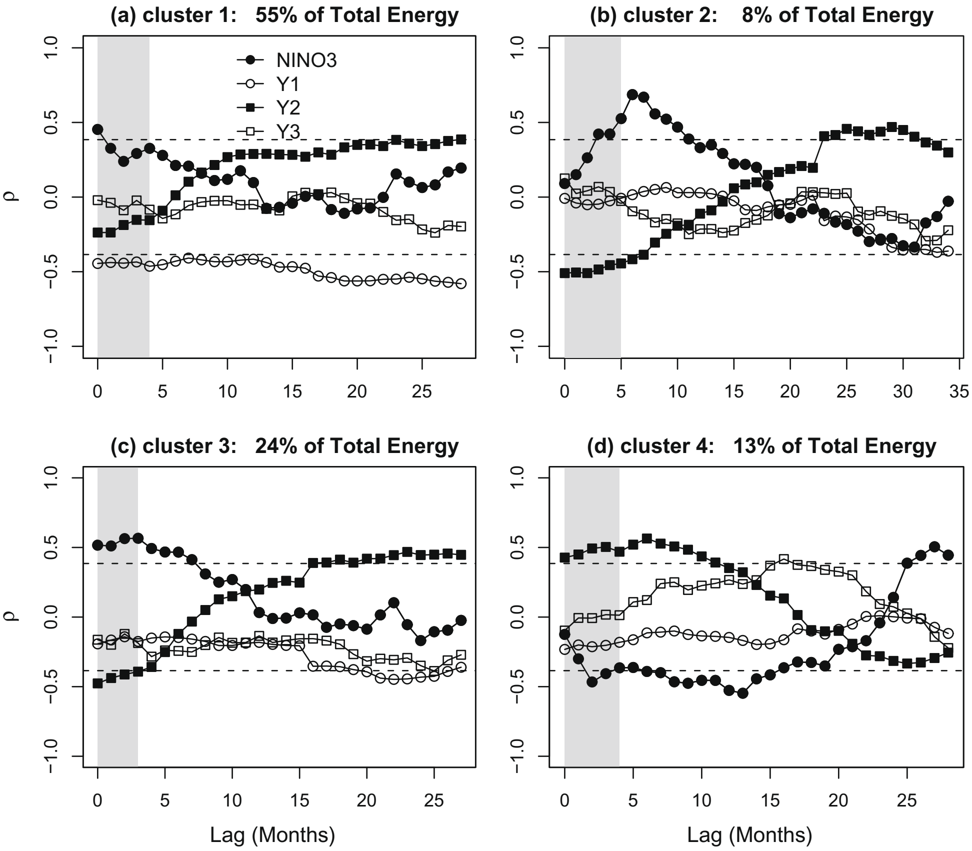

Fig. 5. Lagged correlation between climate predictors (NINO3 and the first three MVU modes Y_1, Y_2 and Y_3 of the D_{20} thermocline data) and the energy inflow series of the hydropower clusters. The gray shaded region shows the rainfall season (as defined in Table 1), where $\tau = 0$ represents the last month of the rainy season. For instance, $\tau = 6$ in panel a (cluster 1) represents the correlation between the climate predictor of November of year $i-1$ and the wet season energy inflow (average of the Dec–May flows) of cluster 1 at year i.

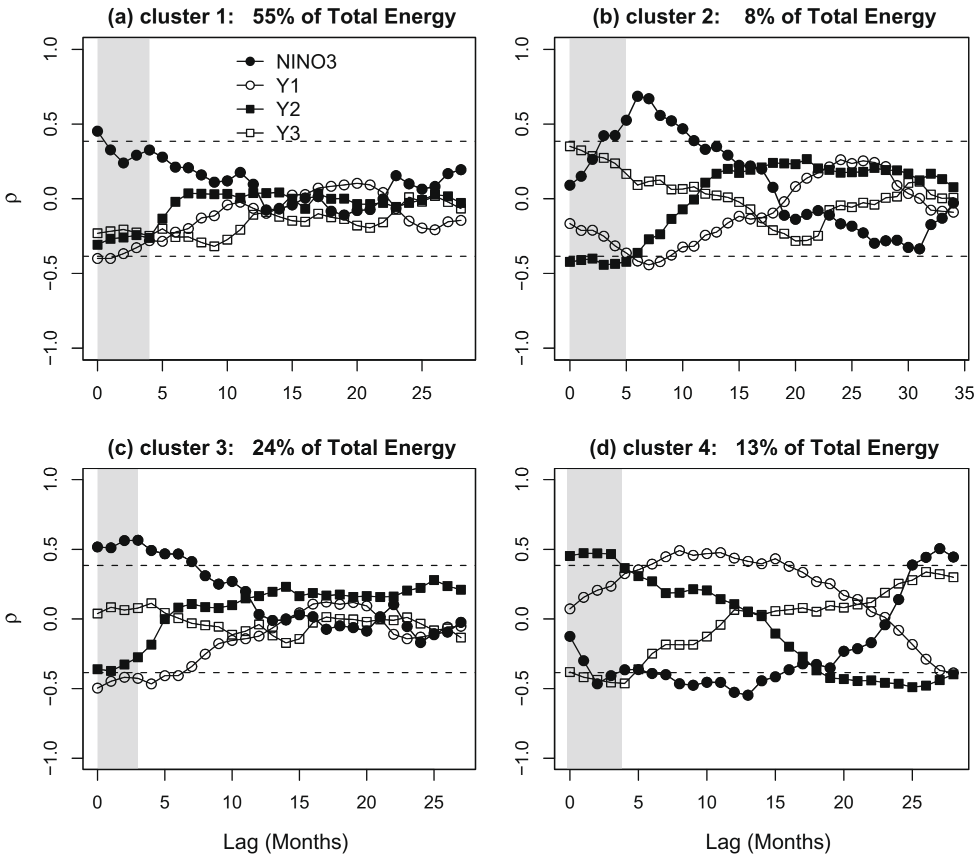

Fig. 6. Same as in Fig. 5, but for the first three PC's of the D_{20} thermocline data.

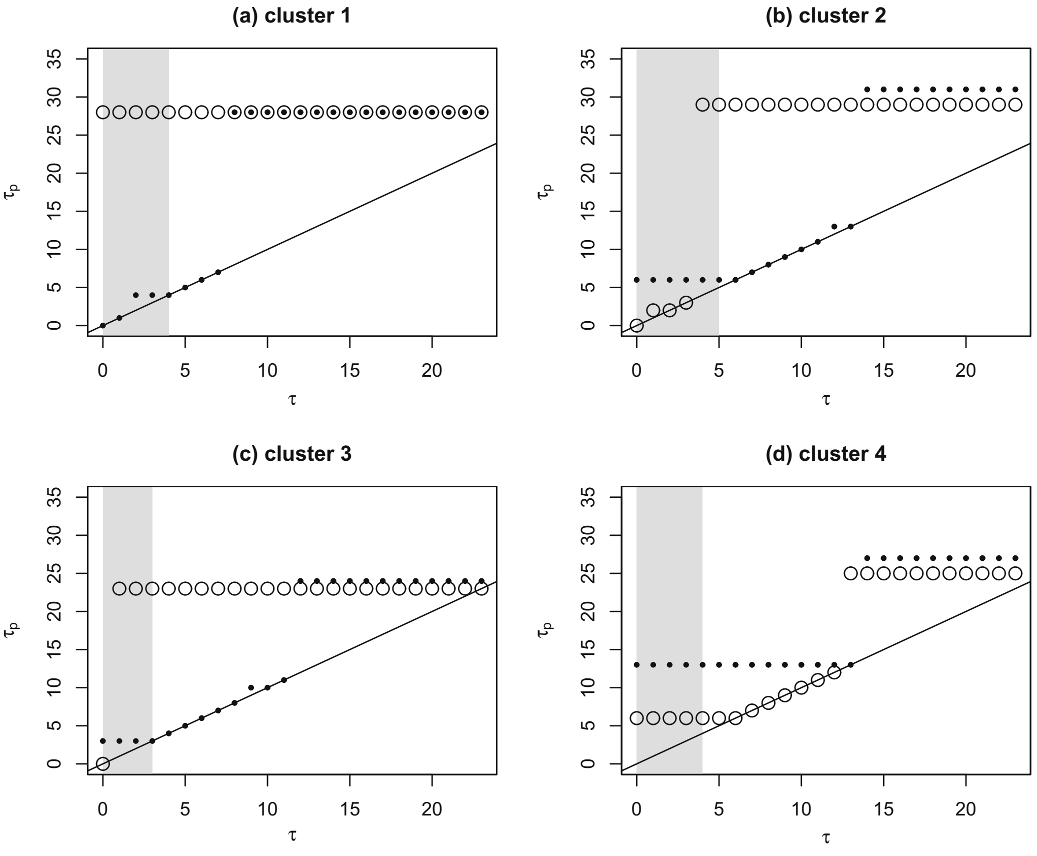

Fig. 7. Lag times (τ_1, τ_2 along the *y* axis) selected for the NINO3 (τ_1, solid circles) and second MVU mode (τ_2, open circles) according to the lead time (τ, along the *x* axis) of the forecast for each cluster. The gray shaded region shows the rain season period, where $\tau = 0$ represents the last month of the rainy season.

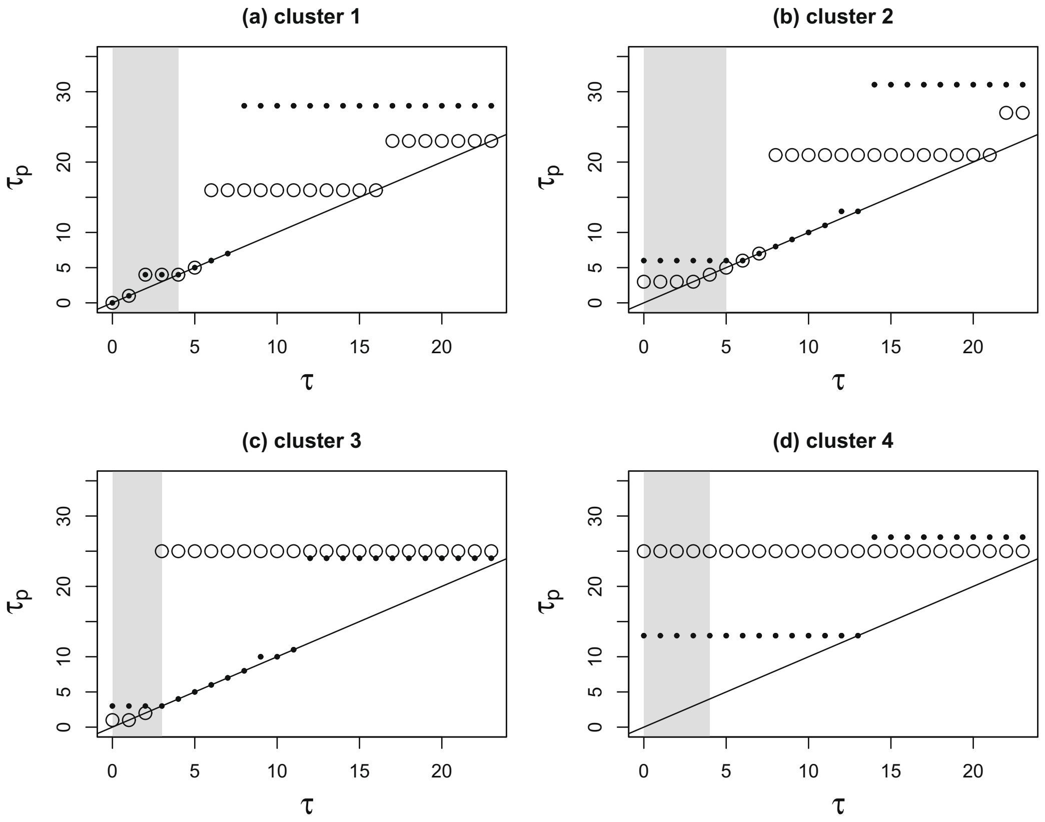

Fig. 8. As in Fig. 7, but for the NINO3 (solid circles) and second PC mode (open circles).

where τ refers to the lead time of the forecast, t_i is the end month of the wet season (as defined in Table 1) of year i, and $\tau_1 = \tau_2 = \{0, 1, 2, \ldots, 32\}$ are referred as the set of lag times of NINO3 and Y_2, which turns out to be functions of the lead time of the forecast ($\tau_1, \tau_2 \geqslant \tau$). We have also omitted the cluster subscript k of all variables and parameters for the sake of simplicity.

In order to reduce the number of parameters and simplify model (6), we first assume that the function f in (6) is linear. For a given cluster and lead time τ of forecast, we also consider only one element of the set of lag times τ_1 and τ_2, but we let it vary according to the predictor used in the model (i.e. τ_1 and τ_2 need not to be the same). The most common alternative to select those lag times is just to consider them equal to the lead time of the forecast, such that $\tau_1 = \tau_2 = \tau$. However, in some cases longer lag times are preferred than shorter ones since they may have better correlations (see, for instance, the cross-correlation of Y_2 in panel (5a)). Hence, we define τ_1 and τ_2 as the lag time $\hat{\tau} \geqslant \tau$ in which the correlation function ρ (Figs. 5 and 6) between response variable and predictors is maximum:

$$\tau_1(\tau) = \hat{\tau} \,|\, \rho(\text{NINO3}(t - \hat{\tau}), \bar{q}) \text{ is maximum and } \hat{\tau} \geqslant \tau \quad (7)$$

$$\tau_2(\tau) = \hat{\tau} \,|\, \rho(Y_2(t - \hat{\tau}), \bar{q}) \text{ is maximum and } \hat{\tau} \geqslant \tau. \quad (8)$$

and t is the end month of the wet season (as defined in Table 1).

Note that seasonal forecasts imply that we must have the inequality $\tau_1, \tau_2 \geqslant \tau$. For instance, a March–May wet season with $\tau = 3$ and $\tau_1 = \tau_2 = 4$ would imply forecasts issuing in February using NINO3 and Y_2 dated back to January of the concurrent year. Typically, we vary τ from 0 to 24 months while τ_1 and τ_2 can go up to 32 months.

Fig. 7 shows the lag times τ_1 and τ_2 selected for NINO3 and the second MVU mode $Y2$, respectively. Clusters 1 to 3 present similar lag times, with the NINO3 lag time being close to τ for short lead forecasts, whereas larger lags are selected for $Y2$. On the other hand, a variety of lag times is selected for the Y_2 obtained trough PCA (Fig. 8).

A general framework for a long term forecast model for the wet season flow $\bar{q}_i$ can finally be defined as:

$$\bar{q}_i \sim N(\alpha + \beta \cdot \text{NINO3}(t_i - \tau_1(\tau)) + \theta \cdot Y_2(t_i - \tau_2(\tau), \sigma^2) \quad (9)$$

where τ is the forecast lead time that we are interested in, τ_1 and τ_2 are the individual lag times of each predictor, Y_2 is the second MVU or PC mode of the D_{20} data and we have omitted the cluster subscript k of all variables and parameters for simplicity.

Results

Cross-validated r^2 and rmse scores

We use the *leave-one-out* cross-validation scheme (Michaelsen, 1987; Hastie et al., 2001) to compare the forecast skills of three models (description in Table 2) derived from (9). The cross-validation is done as follows: (i) one data point is withheld and the model parameters are estimated using the remaining data; (ii) forecasts are made for the removed data point; (iii) this process is repeated until all data points are predicted. The wet season energy flow from 1980 to 2006 is used as validation data. Note that the

Table 2
Seasonal forecast models tested.

Model #	Predictors	Y_2 from	Equation
1	NINO3	–	$E(\bar{q}) = \alpha + \beta \cdot \text{NINO3}$
2	NINO3, Y_2	MVU	$E(\bar{q}) = \alpha + \beta \cdot \text{NINO3} + \theta \cdot Y_2^{mvu}$
3	NINO3, Y_2	PCA	$E(\bar{q}) = \alpha + \beta \cdot \text{NINO3} + \theta \cdot Y_2^{pca}$

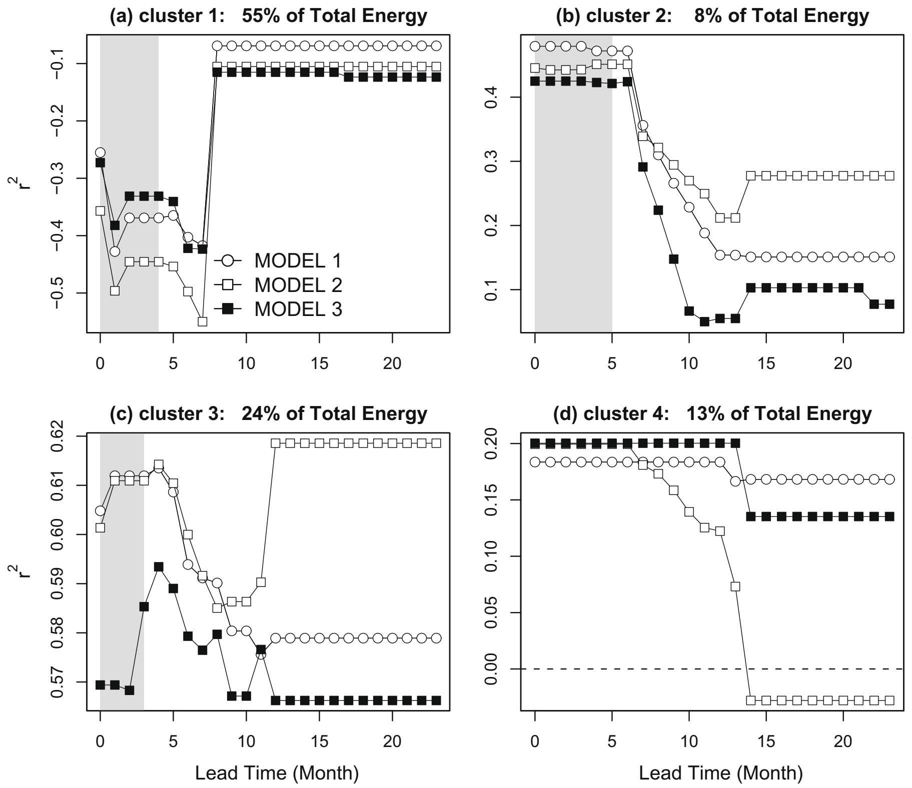

Fig. 9. Cross-validated r^2 score for each cluster as a function of lead time (months) of forecast and model used. Model definitions are shown in Table 2.

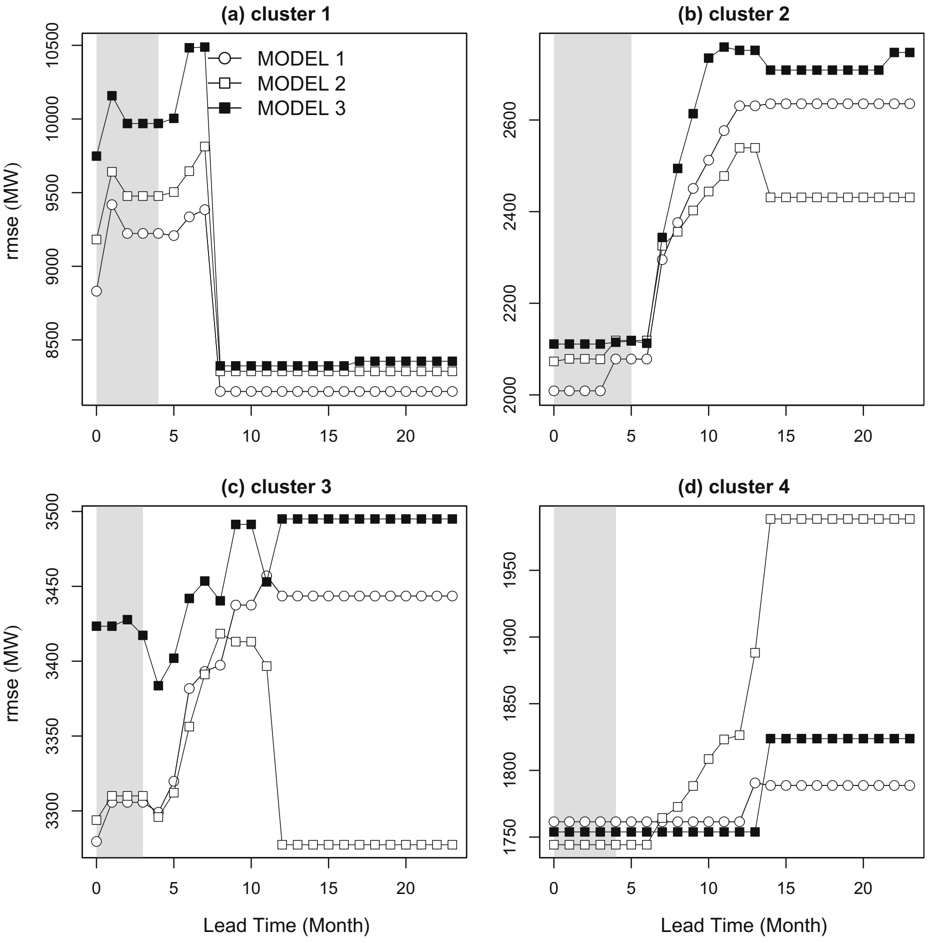

Fig. 10. Cross-validated *rmse* score for each cluster as a function of lead time (months) of forecast and model used. Model definitions are shown in Table 2.

unavailability of D_{20} data before 1980 limits the validation period to a short extent.

In order to compare the forecast skills of the proposed models over the simplest model (or null model) based on the average wet season flow, we use the r^2 score defined as:

$$r^2 = 1 - \frac{\sum_{i=1}^{N}(\hat{q}_i - \bar{q}_i)^2}{\sum_{i=1}^{N}(Q - \bar{q}_i)^2} \qquad (10)$$

where $\hat{q}_i$ is the forecast for year i, Q represents the historical average (based on the 1931–1979 period) wet season hydroenergy flow, $\bar{q}_i$ is the observed wet season energy flow at year i and N is the total number of years in the validation period. For $0 < r^2 \leqslant 1$, the model performs better than if one had used a model based on chance, in this case based on the average wet season flow.

Fig. 9 displays the r^2 score for each cluster as function of lead time of forecast and model utilized as defined in Table 2. The inclusion of climate predictors is able to improve the skill over the null model in all clusters but cluster 1, whose long term average provides a better forecast. NINO3 turns out to be a good predictor for clusters 2–4, especially for leads less than 8 months. Clusters 1–3 have improvements in the forecast skill when the second MVU mode Y_2 is also used as predictor (model 2), in particular for lead times greater than 8 months. Cluster 4 shows similar performances across all models for short lead times (<7 months), but model 2 has a sharp drop in the skill as the lead time increases.

The skills of the models obtained for the r^2 score are corroborated by the rmse scores showed in Fig. 10. Cluster 1 has a high forecast error across all models. Cluster 2 shows similar forecast errors across models for lead times less than 7 months. Thereafter model 2 has the lowest rmse. Models 1 and 2 have close rmse for cluster 3 up to 8 months, where for leads greater than this the inclusion of Y_2 (from MVU) through model 2 leads to a sharp decrease in the rmse. Cluster 4 shows no difference in the model skills up to seven months lead. Beyond that models 1 and 3 have similar rmse up to 13 months lead and thereafter model 1 has the lowest error.

Cross-validated forecasts

Cross-validated forecasts at six month lead are made using model 2, which overall showed the best skill among the models as evidenced in Figs. 9 and 10. Fig. 11 shows the 95% confidence interval for the mean forecast. Cluster 1, as expected, does not have good forecast results. Forecasts tend to follow the observed data for clusters 2–4, with substantial improvement over the null model

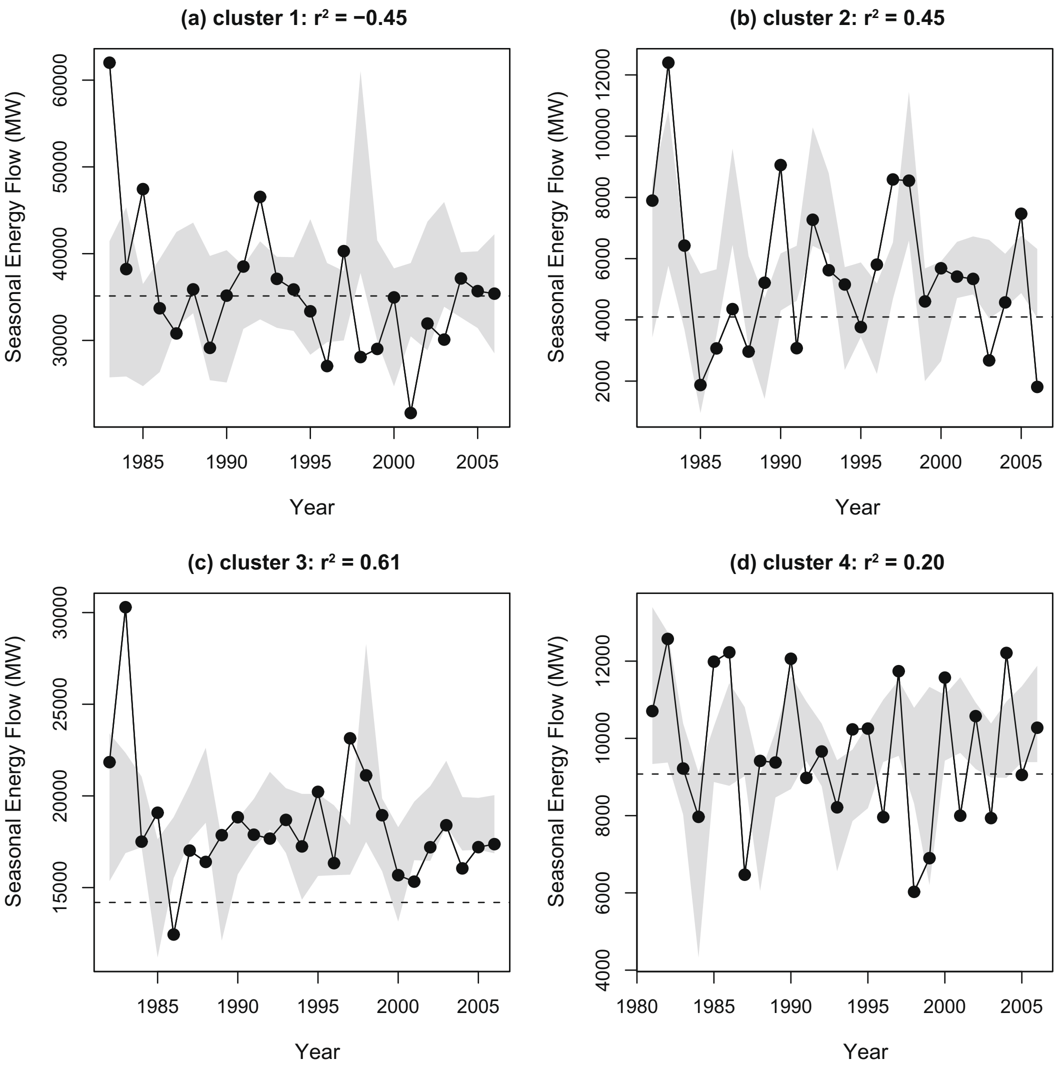

Fig. 11. Cross-validated forecasts for lead time = 6 months. The gray region shows the 95% confidence interval for the mean forecast. The dashed line shows the historical average based on the 1931–1979 period.

based on the average flow. Interestingly cluster 3 shows a somewhat trend of flows above the average for the validation period that is well reproduced by the forecasts. A drop in the model skill evidenced by the r^2 score is observed for clusters 1, 2 and 4 when forecasts are made with 18 month lead (Fig. 12). One does not see this deterioration in the skill for cluster 3.

Summary

A climate informed statistical model was developed and applied to predict wet season hydroenergy inflow of clusters of hydropower reservoirs located in Brazil. Climate predictors were used to reduce the variance of the forecasts. The NINO3 index and the tropical Pacific thermocline D_{20} were used in the model. Due to the high-dimensionality of the D_{20} data, the maximum variance unfolding and the principal component methods were used to obtain the first three modes of variability of D_{20} in order to have uncorrelated climate predictors.

A preliminary analysis of cross-correlation between seasonal flows and climate predictors showed statistically significant lagged correlations between the seasonal energy flow of all clusters and the climate predictors NINO3 and second MVU mode. Lagged correlations obtaining with the PCs were usually very weak. The observed 20-month lagged correlation of NINO3 and the second MVU mode as described in Lima et al. (2009) suggests that the seasonal flow of clusters 2–3, located respectively in south and southeast Brazil, might be affected by the second MVU mode through ENSO. Cluster 4 shows relatively high correlations with NINO3 and the second MVU mode at lag times less than 8 months. A monotonic decreasing trend observed in the seasonal flow of cluster 1 correlates well with the uptrend of the first MVU mode (Lima et al., 2009). Whether such high correlations are pure by chance or linked to some physical mechanisms is not clear yet and deserves further investigations by climate based models.

Finally, cross-validated probabilistic forecasts show that model 2 (NINO3 and second MVU as predictors) performs better than the null model (based on the long term average flow) for clusters 2, 3 and 4. The performance of the models was not dramatically affected as the lead time increases and forecast of the wet season energy flow across Brazil were made up to 20 months in advance with moderate skill. Therefore, ONS could operationally use the models proposed here for long term forecasts of seasonal hydroenergy flow. Future work could focus on reducing parameter uncertainties through hierarchical Bayesian modeling, where regression parameters of each cluster are shrunk towards a common mean. Concurrent and lagged correlations across clusters could also be considered through a multivariate forecast model,

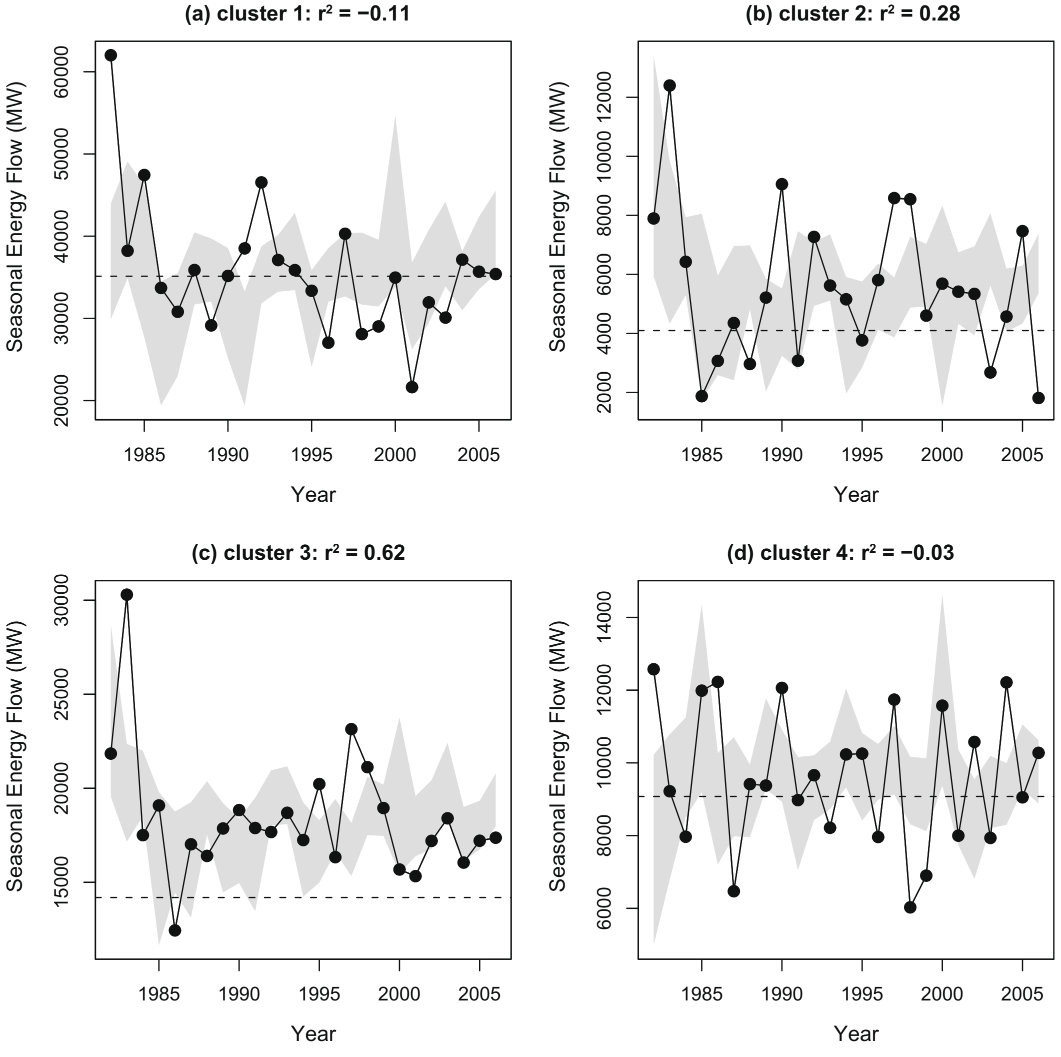

Fig. 12. As in Fig. 11, but for lead time = 18 months.

where inter-cluster variability would also be modeled. The predictability of the dry season flow could also be analyzed and explored.

References

Barros, V., Gonzalez, M., Liebmann, B., Camilloni, I., 2000. Influence of the South Atlantic convergence zone and South Atlantic Sea surface temperature on interannual summer rainfall variability in Southeaster South America. Theor. Appl. Climatol. 67, 123–133.

Barros, M.T.L., Tsai, F.T.-C., Yang, S.-L., Lopes, J.E.G., Yeh, W.W.-G., 2003. Optimization of large-scale hydropower system operations. J. Water Resour. Plann. Manage. 1290 (3), 178–188.

Behringer, D.W., Ji, M., Leetmaa, A., 1998. An improved coupled model for ENSO prediction and implications for Ocean initialization. Part I: The Ocean data assimilation system. Mon. Weather Rev. 126, 1013–1021.

Brockwell, P.J., Davis, R.A., 2002. Introduction to Time Series and Forecasting. Springer.

Carvalho, L.M.V., Jones, C., Liebmann, B., 2004. The South Atlantic convergence zone: intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall. J. Climate 17, 88–108.

Chiang, S.-M., Tsay, T.-K., Nix, S.J., 2002. Hydrologic regionalization of watersheds. I: methodology development. J. Water Resour. Plann. Manage. 128 (1), 3–11.

Costa, F.S., Damázio, J.M., Maceira, M.E.P., Denício, M., Guilhon, L.G., Silva, S.B., 2003. Stochastic model of monthly streamflow forecast – PREVIVAZM. In: XV Brazilian Symposium of Water Resources. (In Portuguese).

Diaz, A.F., Studzinski, C.D., Mechoso, C.R., 1998. Relationships between precipitation anomalies in Uruguay and Southern Brazil and sea surface temperature in the Pacific and Atlantic Oceans. J. Climate 11, 251–271.

Grimm, A.M., 2004. How do La Niña events disturb the summer monsoon system in Brazil? Clim. Dynam. 22, 123–138.

Grimm, A.M., Ferraz, S.E.T., Gomes, J., 1998. Precipitation anomalies in Southern Brazil associated with El Niño and La Niña events. J. Climate 11, 2863–2880.

Grimm, A.M., Barros, V.R., Doyle, M.E., 2000. Climate variability in Southern South America associated with El Niño and La Niña events. J. Climate 13, 35–58.

Guilhon, L.G.F., Rocha, V.F., Moreira, J.C., 2007. Comparison of hydropower inflow forecat models. In: IONS Workshop of Streamflow Forecasts. (In Portuguese).

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer.

Isik, S., Singh, V.P., 2008. Hydrologic regionalization of watersheds in Turkey. J. Hydrol. Eng. 13 (9), 824–834.

Ji, M., Smith, T.M., 1995. Ocean model response to temperature data assimilation and varying surface wind stress: intercomparisons and implications for climate forecast. Mon. Weather Rev. 123, 1811–1821.

Ji, M., Leetmaa, A., Derber, J., 1995. An Ocean analysis system for seasonal to interannual climate studies. Mon. Weather Rev. 123, 460–481.

Kahya, E., Kalayci, S., Piechota, T.C., 2008. Streamflow regionalization: case study of Turkey. J. Hydrol. Eng. 13 (4), 205–214.

Kaplan, A., Cane, M., Kushnir, Y., Clement, A., Blumenthal, M., Rajagopalan, B., 1998. Analyses of global sea surface temperature 1856–1991. J. Geophys. Res. 103, 18567–18589.

Lenters, J., Cook, K.H., 1995. Simulation and diagnosis of the regional summertime precipitation climatology of South America. J. Climate 8, 2988–3005.

Lima, C.H.R., Lall, U., Jebara, T., Barnston, A.G., 2009. Statistical prediction of ENSO from subsurface sea temperature using a nonlinear dimensionality reduction. J. Climate 22, 4501–4519.

Maceira, M.E.P., Penna, D.D.J., Damázio, J.M., 2005. Synthetic generation of energy and streamflow scenarios for the energetic operation planning. In: XVI Brazilian Symposium of Water Resources. (In Portuguese).

Michaelsen, J., 1987. Cross-validation in statistical climate forecast models. J. Climate Appl. Meteor. 26, 1589–1600.

ONS, 2007. Streamflow series update-1931–2006 period. Tech. rep., National Operator of the Electrical System. (In Portuguese).

Reynolds, R.W., Smith, T.M., 1994. Improved global sea surface temperature analyses using optimum interpolation. J. Climate 7, 929–948.

Ropelewski, C.F., Halpert, M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. Mon. Weather Rev. 115, 1606–1626.

Silva, B.C., Collischonn, W., Tucci, C.E., Clarke, R.T., Corbo, M.D., 2007. Short term streamflow hydroclimatic forecast for the São Francisco Basin. In: I ONS Workshop of Streamflow Forecasts. In Portuguese.

Vera, C., Baez, J., Douglas, M., Emmanuel, C.B., Marengo, J., Meitin, J., Nicolini, M., Nogues-Paegle, J., Paegle, J., Penalba, O., Salio, P., Saulo, C., Dias, M.A.S., Dias, P.S., Zipser, E., 2006. The South American low-level jet experiment. Bull. Am. Meteorol. Soc. 87, 63–77.

Weinberger, K.Q., Saul, L., 2006. Unsupervised learning of image manifolds by semidefinite programming. Int. J. Comput. Vision 70 (1), 77–90.

Weinberger, K.Q., Sha, F., Saul, L.K., 2004. Learning a kernel matrix for nonlinear dimensionality reduction. In: Proceedings of the Twenty First International Conference on Machine Learning. Banff, Canada, pp. 839–846.

Wilks, D.S., 2006. Statistical Methods in the Atmospheric Sciences. Elsevier, New York.