Groundwater monitoring of a hidric shortage crisis in Brazil based on LS-SVM forecasts for the city of São Paulo

Ademir L Xavier Jr, Sergio Celaschi

Abstract—We evaluate the groundwater (GW) storage under the state and the city of São Paulo using remote data provided by NASA's Gravity Recovery and Climate Experiment (GRACE) in the context of the water shortage crisis in 2013-2014. The study provides a GW forecasts based on Least Square Support Vector Machine (LS-SVM) for GRACE and Global Land Data Assimilation System (GLDAS). The study emphasizes the strong correlation between GRACE data and the differential volumes of the city main water supply system further enhancing the reliability of the LS-SVM forecast. We provide an equivalent water thickness forecasts of ground water time series and show that GRACE generalization is better than that obtained with GLDAS data together with error estimates. A minimum around 22 months was found as the optimum amount of past information to be used in the regressor to minimize forecast error.

Index Terms—GLDAS model, GRACE mission, gravimetry, groundwater usage, LS-SVM, remote sensing, support vector machines,watershortage crisis.

1 Introduction

The distribution of water in the atmospheric circulation has shown significant modifications in the last decade possibly as a side effect of climate change, besides population growth [1][2]. This phenomenon manifests itself as an abrupt change in the rain regimes, posing significant risks to many populations[3]. Both severe droughts in over populated areas[4][5] and extreme rain events in lowlands [6], often irregularly occupied, constitute threatening scenarios against which intergovernmental actions are being discussed [7]. However, from the point of view of urban planning and government, lack of water is far more damaging: it is estimated that, by 2025, more than two billion people will be living under water shortage regimes [8][9]. Since water demand in the world is not equally distributed among the several economic sectors, significant challenges are expected in the case of an enduring and generalized water crisis.

Local water crises have been registered in several parts of the world, some examples are: Canada [10], Europe [11], the Middle East [12], Australia[13], China [14], India [15][16] and Africa[17]. Global trends in population growth are the major factor contributing to these episodes; however, it is not less important to monitor the role played by climate change [18][19].Latin America is recognized as a region of water abundance, which does not mean water is a readily available resource [20]. In Brazil, in spite of the existence of places with chronic water shortage under semi-arid regimes [21][22][23], water has been regarded as an abundant resource especially in the southern parts of the country where most of the population lives. The presence of a giant aquifer linked to the Paraná basin in south Brazil [24][25], together with special hydrologic cycles associated to the Amazon basin [26][27], significantly increase the hydric potential of four important states in the south which are: São Paulo, Paraná, Santa Catarina and Rio Grande do Sul besides north Argentina, Paraguay and Uruguay. However, water abundance in a country does not imply in climate change invulnerability. Since 50% of Brazil GDP is linked to renewable resources[28], water dependence makes slight changes in its availability a potential threat to the economic basis. Lack of water will surely affect agriculture and other eco sustainable resources in Brazil as already predicted in several works [29][30].

Table 1 São Paulo city WSS with information about number of inhabitants (in millions), water flow (liters per hour), historical average precipitation (HAP) and accumulated precipitation (AP) for the 2013-2014 period. Observed variations at each WSS in the crisis period are also given in the last column.

São Paulo WSS	*Total Inh. (mil)*	*~ Flow (l/h)*	*HAP (mm)*	*AP (mm)*	*Percent change*
Alto Cotia	0.41	1200	1390	13337	-3.8%
Alto Tietê	4.5	15000	1463	1054	-28.6%
Cantareira	6.5	33000	1569	905	-42.3%
Guarapiranga	4.9	15000	1342	1189	-11.4%
Rio Claro	1.5	40000	2176	2309	6.1%
Rio Grande	1.2	5000	1559	1337	-14.2%

- *A. L. Xavier Jr is technologist at the Brazilian Space Agency, Área 5, Bloco A, CEP 70610 200, Brasília, DF, Brazil. E-mail: ademir.junior@aeb.gov.br*
- *S. Celaschi is technologist at Renato Archer Center. E-mail: sergio.celaschi@cti.gov.br*

An iniquitous side of a continuous water shortage manifested already with a water supply crisis never seen in the last 85 years in the largest Brazilian city. São Paulo city climate is generally humid, with most rains (and floods) concentrated in January-March. Yet, by mid-2013, São Paulo frequent floods ceased to be a problem with an abnormal lack of rain which caused a drop of more than 40% in the average volumes of the city main Water Supply System (WSS). Coincidently, the situation had some parallel elsewhere[34]given the location size and the amount of people involved. Detailed information about São Paulo city WSS [31][32] is provided by SABESP (São Paulo State Sanitation Company) and can be seen in Table 1, with water flows up to 40000 liters per hour, together with historical average precipitation (HAP) and recorded values for 2013-2014[33]. As this table shows, the most severe water deficit was observed in Cantareira system, which serves the largest population.

The aim of this work is to correlate and forecast possible trends in the water pluviometric cycle (as available through SABESP time series) with equivalent water thickness (EWT) as a measure of groundwater (GW) under the city of São Paulo using NASA Gravity Recovery and Climate Experiment (GRACE [35][36]) and Global Land Data Assimilation Systems (GLDAS[37]). With the availability of GRACE and GLDAS data, we use GW time series as an alias for pluviometric data, disregarding losses due to evapotranspiration and water consumption by the population, and by adjusting a special regressor using Support Vector Machines. The use of statistical learning tools for climate forecast renders possible the implementation of predictive machines for the entire land surface of the globe. Due to the intrinsic complexity of climate systems, the use of methods not dependent on first principles but on data learning seems specially indicated to further reduce forecast uncertainty.These methods may be used to extract information from the pseudo-periodicity found in many climate signals. In this sense, the lack of a recurrent peak in GRACE EWT in the time series of São Paulo city was a unique opportunity to test such methods in their ability to predict system behaviour in the post-crisis time. GRACE EWT results are presented in Section 2.3. This work studies the generalization error associated to such data driven methods as a potential tool to improve the forecast power for the GW dynamics in time.

2 Data and Methods

2.1 Problem description

The State of São Paulo is located in the southeast part of Brazil, between the States of Paraná and Rio de Janeiro. Its capital, the city of São Paulo, southeast of the state, is the largest city in Latin America whose economic and social importance may be measured by some of its numbers: the metropolitan area hosts 11 million people (it is the largest city in the southern hemisphere) and is responsible for 11% of Brazil's GDP. São Paulo city occupies an area of about 8000 km^2 in the Tietê river basin with two tributaries, Tamanduateí and Pinheiros. With the city proximity to the ocean, the climate is mainly humid (80%), with mild temperatures in winter (May-Aug, 17.1^{o}C), and a moderate summer (Dec-Mar, 22.1^{o}C) as compared to other Brazilian cities. Average monthly precipitation runs from 43 mm in winter to over 200 mm in summer, causing frequent urban floods that are further enhanced by the growing imperviousness of urban soil. The city area is circumscribed by a square of coordinates 23.401^o S, 46.791^o W and 23.981^{o}S and 46.151^{o}W.

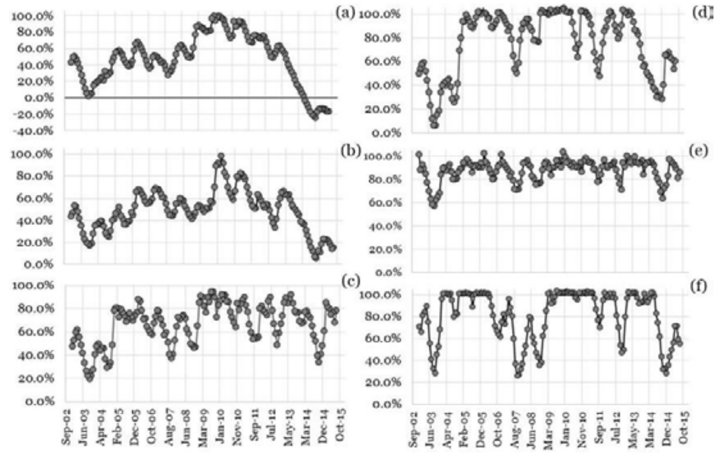

Figure 1Recorded volumes in percentagefrom Sep-02 to Sep-15 for São Paulo main WSS: (a) Cantareira, (b) Alto Tietê, (c) Guarapiranga, (d) Alto Cotia, (e) Rio Grande and (f) Rio Claro. For Cantareira WSS, values below 0% mark the end of the gravity assisted pumping level.

The city management and distribution water systems are well described in [38]. Following[33], here we provide some updates after the drought of 2013 and concentrate on Cantareira WSS [39] which, being the largest municipal storage system (responsible for supplying 34.1% of the city population), shown the most severe water shorage as indicated by Table 1. Cantareira WSS has a total volume of 1495hm[3] and is located north of the city. The system is formed by six dams along five basins (rivers: Atibainha, Cachoeria, Jacareí, Jaguariand Juquery) interconnected by a complex network of tunnels, channels and pumping stations necessary to overcome the physical barrier of Cantareira ridge. SABESP is a public-held concessionary responsible for exploring watersheds, according to a directive plan encompassing a large region including the metropolitan area and Cantareira. However, companies are only responsible for collecting, treating and distributing water. According to this model, the federation (through ANA, or the Federal Water Agency) is responsible for managing the usage of untreated water or deciding the level of priority among several users (e. g., switch between irrigation and water storage). A federal law warrants the priority of humans and animals in case of severe drought.

From Oct-13 to Feb-14, the accumulated precipitation on the city amounted to 444 mm while the average level for the same period is 995 mm, triggering the crisis, Fig. 1(a). The previous minimum occurred by the end of 2003, with a similar behaviour in other systems (Figure 1(b)-(f)). On the other hand, the operation of WSS such as in Fig. 1(d), (e) and (f), close to volume saturation for long periods in the past, indicate that water has not been properly stored or, at least, water excess has not been properly managed. Others could interpret such plots as another manifestation of climate change: how can one explain the existence in time (< 5-year interval) of such highly disparate hydrologic regimes?

Cantareira WSS water throughput in 2014 reduced to 1/4 of the annual average. Since this average is 44.1+/-12.4 m^3/s, statistically, the drought had 0.4% chance of occurrence. The former "absolute" minimum recorded for this system was in the year 1953 (~25 m^3/s against 11 m^3/s in 2014). On Oct-14, the minimum volume was reached corresponding to 2.9% of the total capacity contrasting with the last maximum on Jan-10 (~100%). The crisis was managed by applying several policies including supply rotations (compulsory water supply reduction per capita), incentives to saving water (bonus program), transference of treated water from other systems, reduction in the average time for fixing leakages and other loss prevention actions. The city of São Paulo has near 2000 wells pumping about 10 m^3/s from the São Paulo acquifer (associated to the Alto Tietê basin[40][41]), nearly the same throughput value extracted from Cantareira after reduction. This is not an updated number, since the crisis increased the number of wells considerably, thus enhancing the importance of assessing GW potential.

2.2 LS-SVM

The least square support vector machine (LS-SVM) technique [42] is the evolution of a binary classification method in the scope of Vapnik-Chervonenkis statistical learning theory[43].The main application is to propose generalizations functions for time series. The method is recognized as providing accurate generalization [44] for a large class of time series, including chaotic ones[45]. The method has been tested in a variety of contexts ranging from financial[46]to climate systems [47][48][49][50][51].

Essentially, the training process is reduced to an optimization problem in which input vector of the type $\{\boldsymbol{x}_i, y_i\}, i = 0, \ldots, N$, is used to find the optimal solution for the function

$$y = \boldsymbol{w}^{\boldsymbol{t}}\varphi(x) + b \quad (1)$$

where N is the size of the training set, $x_i \in \mathbb{R}^p, y \in \mathbb{R}$ and $\varphi(x)$ is a predefined function. Here p is the dimension of the training input vector and y is the output. The solution provides suitable values for the vector $\boldsymbol{w}$ and the scalar b along with the choice of a kernel function $\varphi(x)$ which depends on other parameters. Due to its unique generalization features, only the so called "Gaussian kernel" was regarded in this work, that is, functions of the type

$$K(x', x'') = exp\left(\frac{-\|x' - x''\|^2}{2\sigma^2}\right), \quad (2)$$

with σ a tuning parameter. From the practical point of view, after optimization, the estimated value will be given by

$$y(x) = \sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b, \quad (3)$$

where the set $\{\alpha\}$ contains Lagrange multipliers of the original problem[42]. Such problem may be reduced to solving the $(N+1) \times (N+1)$ linear system

$$\begin{pmatrix} 0 \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} 0 & \boldsymbol{1} \\ \boldsymbol{1^t} & \Omega + \gamma^{-1}\boldsymbol{1} \end{pmatrix} \begin{pmatrix} b \\ \boldsymbol{\alpha} \end{pmatrix}, \quad (4)$$

with γ a optimization parameter, $\Omega_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{1} = (1,1,\ldots,1)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)$.

We call $n$the regression order or the position of the time advanced point for which a numerical forecast must be obtained[52]. That is, (1), the value of y refers to the point x_n not necessarily the immediate advanced point after the last series element. The regressor dimension p is essentially the number of previous points in the training set, therefore, is proportional to the amount of past information. Solving (4) for a given pair (σ, γ) does not guarantee the best solution. The search inthis space requires a validation set for which a measure of distance [53]between validation and forecast should be minimized. Thus, input data are divided into two sets of size N_T (training) and N_V (validation), so that the effective instances of available points for training and validation are, respectively, $N_T - p - n$and$N_V - p - n$. We take the size of $N_V$as nearly $N/3$ and, for each regressor order n, distinct optimal values for the pair(σ, γ) must be found. Also since, during validation, the variance of the regression values in relation to the real data can be calculated for each order (which is proportional to the residue of the distance), estimates for the regression errors can be calculated as the best standard deviation found during validation. The residue should be minimized not only for the (σ, γ)pair, but also for p or the number of recursive past points to be included in the regression function (3). The search for optimum parameters was undertaken using Particle Swarm Optimization, PSO[54][55]. New tuning parameters randomly chosen on the (σ, γ)plane were found. These were based on an iterative procedure during which the value of the distance residue is used to orient the "swarm" toward the optimum solution. Optimum values are found and compared among several p, so that an optimal regressive model is found.

2.3 Applying LS-SVM to GRACE and GLDAS data

As result of a NASA and DLR (Deutsche Forschungsanstaltfür Luft und Raumfahrt) partnership, March 2002 saw the launching of the twin satellite GRACE[56] as a very successful enterprise resulting in a time dependent matrix of gravimetric data for the entire surface of the Earth. GRACE is capable of estimating the stored GW as an EWT [57] after a series of steps [58][59] in which mass anomalies (measured as anomalies in the gravity field) are inferred based on differences in the tracked orbital positions of both satellite elements. GLDAS is distinct join effort by NASA and Goddard Space Flight Center to integrate remotely sensed and ground data and constraint surface state models in order to render maps of ground water and energy storage[59][60]. These fields provide valuable information for assessing a variety of processes such as predicting climate change, productivity in agriculture, weather and other hydrologic phenomena. Moreover, groundwater maps can be compared to GRACE water equivalents[60]. Therefore, regression functions for both GRACE and GLDAS time series were obtained in this work.

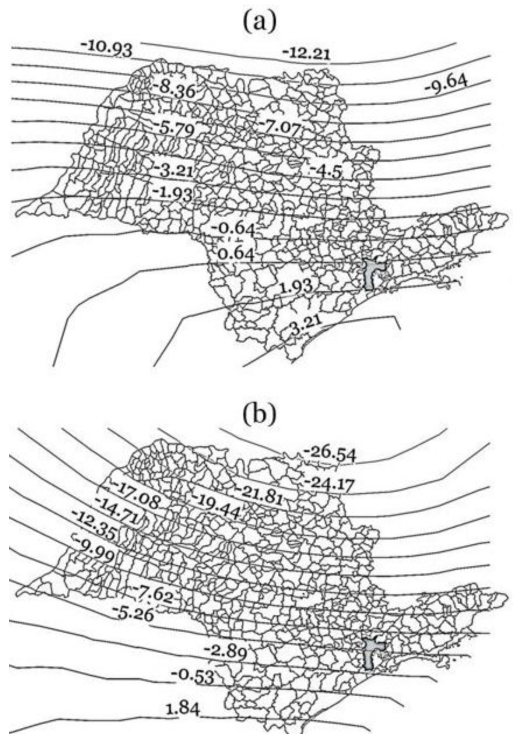

Figure 2 "Iso-thickness" line using GRACE data for EWT (in cm) for the São Paulo state in (a) Oct-10 and (b) Oct-14. Map border indicates the positionof all São Paulo municipalities with São Paulo city highlighted in grey at the southwest part of the state.

However, before blindly accepting GRACE and GLDAS data, it is important to remark the following facts: (i) in order to reduce errors, GRACE data must be averaged over large areas. In the present study, the influence of draught was so severe that such averaging need is of lesser concern in spite of still be present. For this reason, we concentrate our analysis to Cantareira WSS which showed the worst situation. Thus for São Paulo WSS typical errors at 1 km scale are about 6 cm [35]; (ii) GRACE time scale is alsonot accurate and not even represents a monthly average, once GRACE signal is the result of distinct satellite passages which are optimized to enhance low power signals;(iii) there is an intrinsic time delay in the release of GRACE data which contributes to reduce GRACE data as forecast tool; (iv) GLDAS is a modeling tool and does not constitute measure data. Water amounts in this model can be adjusted to suit the atmosphere state without any water balance preservation.With all these observations in mind, in this paper, we pioneered in applying LS-SVM methods to GRACE and GLDAS data with a backtesting perspective.

GRACE EWT data under the State of São Paulo for Oct-10 and Oct-14 are shown in Fig. 2. Negative values are associated to a negative trend in relation to the average baseline. Thus, for the month of October, EWT is usually negative, but the scenario of Oct-14 was unique. It is also apparent that the thickness gradient toward the north of the state has also increased from 2010 to 2014 with a much more severe water depletion on Oct-2014. Monthly data from GRACE and

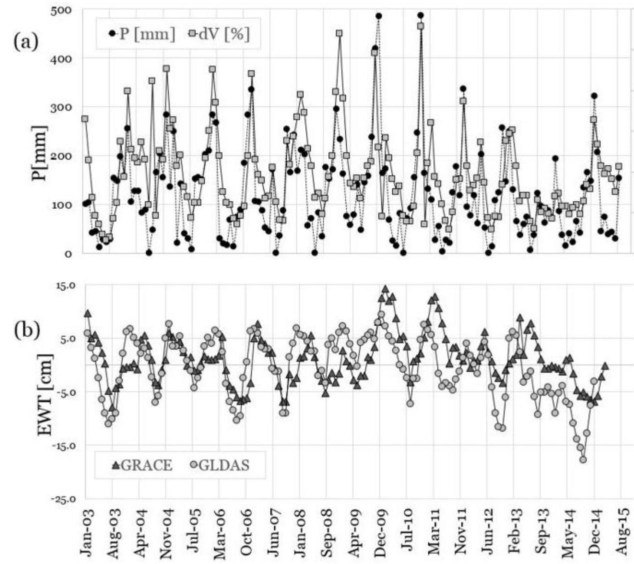

Figure 3(a) Monthly accumulated precipitation (P) in mm and Cantareira volume percent derivative (dV) from Fev-03 to Sep-15. (b) GRACE and GLDAS EWT for the city of São Paulo from Fev-03 to Apr-15.

GLDAS encompassing a period of ~12 years and the equivalent SABESP time series for Cantareira WSS are shown in Fig. 3. This plot compares Cantareira precipitation in mm (Fig. 3(a) - left axis), Cantareira WSS percent differencial volume dV (Fig.4(b) - right axis) from Fev-03 to Sep-15 and two plots of GRACE and GLDAS EWTs (Fig.4(b)) for nearly the same period (GRACE lacking points are due to data unavailability). Average and standard deviation for these series are 1±6 cm for GRACE, 0±5 cm for EWT GLDAS, 123±98 mm for the precipitation and 0±5% for dV in the period Jan-03 to Jan-15. The largest and smallest values, respectively, for this period were 14.2 cm (Feb-03) and -7.9 cm (Sep-03) for EWT GRACE, 9.38 cm (Jan-10) and -17.79 cm for EWT GLDAS, 486 mm (Jan-11) and 0.0 mm (Aug-07) for the precipitation and 17.8% (Jan-11) and -8.4% (Aug-03) for dV. The second smallest EWT GRACE level was -6.90 cm in Jan-15. While the steepest change in dV was observed on Aug-03, the fall by late 2014 was the most severe in the total volume as shown by Fig. 1(a). Comprehensively, dV is affected by water consumption; nevertheless, there is a strong correlation between the differential volume (the derivative of Fig. 1(a)), precipitation and GRACE/GLDAS series data as the following measures indicate.

Admitting $R_{xy}(x_{k+l}, y_k)$ as Pearson's correlation coefficient between two sequence of datax_{k+l}andy_k, shifted by a chosen integer value l, Table 2 brings the main values found for São Paulo data. The series are well correlated for pairs such as dV-P and GLDAS-dV, but maximum correlation was found for $l = 1$ in dV-GRACE and $l = 2$ in P-GRACE.

Table 2Correlation coefficient for pairs of time series data as indicated. The value of l is an integer to representing time lag among adjacent data.

	GRACE	P	dV
GLDAS	0.53 (l=0)	0.55 (l=0)	0.69 (l=0)
GRACE		0.5(l=2)	0.41 (l=1)
P			0.69 (l=0)

3 Results and Discussion

3.1 Training and validating data

The lack of Dec-13 to Mar-14 recurrent peak as shown in Fig. 3 provided the unique opportunity for testing the application of a LS-SVM regression in 2015. The first attempt of using GRACE EWT was made around Jan-15 in order to estimate peak future presence (intensity and time). Since GRACE data are provided with four months delay, SABESP time series was used as a proportional indicator for EWT enabling updated confirmation of equivalent forecasts. LS-SVM training and validation size sets were:$N_T = 153$(dates from May 1 2002 to January 1 2015) and$N_V = 51$. A search for the best regression dimension p for a given n was done in order to minimize the measure:

$$\sigma_V^2(n,p) = \frac{1}{N_V}\sum_{i=1}^{N_V}\left[y - \bar{y}_{\sigma,\gamma}(x_i,n,p)\right]^2, \qquad (5)$$

where $\bar{y}$ is given by (3). Search results are shown in the top of Figs. 4 and 5. As expected, the regression error increases with n- the number of months in the future for the forecast. However, for the regression dimension, there is a minimum around $p = 22$ months, implying that if more information than 22 months is used, the regression loses accuracy. The further decrease in σ_V for$p > 26$ is due to the reduction in the number of validation points in the training set. As a measure of period relevance, bottoms of Figs. 4 and 5 bring plots of the absolute value of the FFT for GRACE and EWT autocorrelation functions[61], respectively:

$$r(k) = \langle (y_i - \langle y_i \rangle)(y_{i+k} - \langle y_i \rangle) \rangle \qquad (6)$$

as a function of the number of months, with the symbol $\langle\ \rangle$ denotes the expected value function.The second maximum

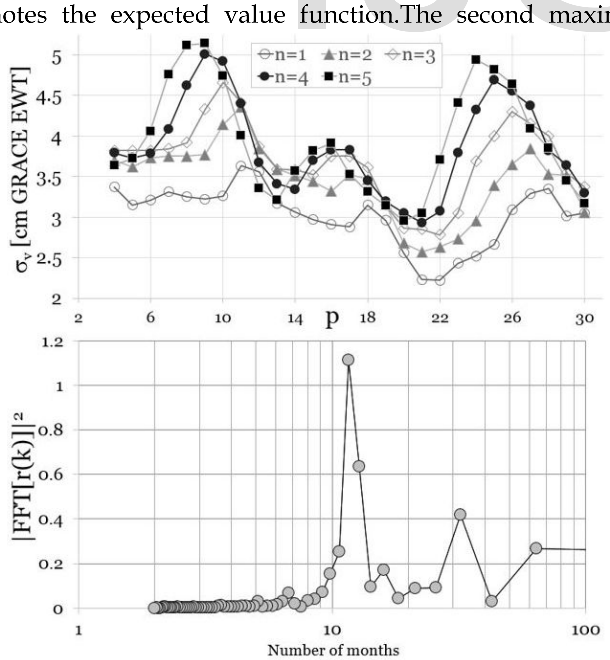

Figure 4(Top) Validation regresson errors for several LS-SVM models as a function of p and n, the regressor dimension and order for data of Fig. 2. (Bottom) FFT of GRACE EWT signal autocorrelation as a function of mong number for $\sigma = 2$ and $\gamma = 22$.

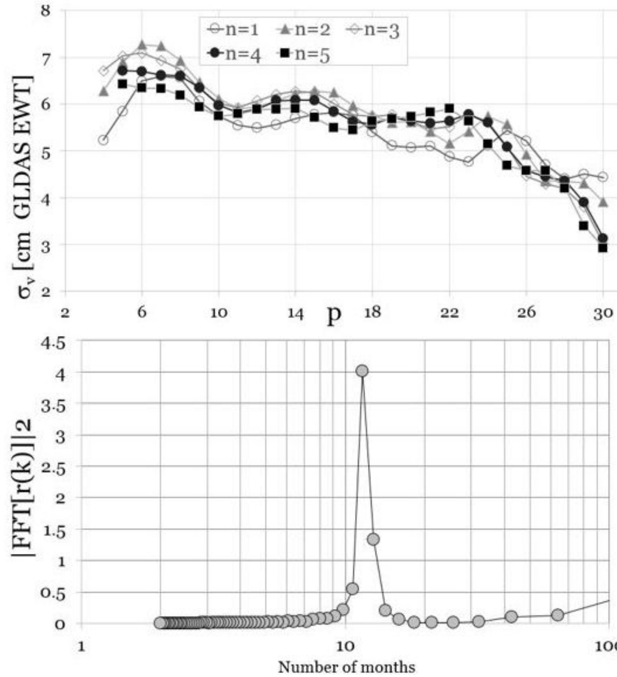

Figure 5(Top) Validation regresson errors for several LS-SVM models as a function of p and n, the regressor dimension and order for data of Fig. 2. (Bottom) FFT of GLDAS signal autocorrelation as a function of mong number for $\sigma = 3$ and $\gamma = 20$.

occurs around 32 months (2.6 years) and correspond probably to weak long period recurrence. The corresponding spectrum for GLDAS only shows the main peak, which is understandable since LDAS does not include GW and surface components such as lakes and rivers [62][63]. Although GRACE and GLDAS are often described as in agreement with each other[60], their respective power spectra are quite distinct.

3.2 LS-SVM forecasts for GRACE data

Fig. 6 and 7, respectively, exhibits superimposed GRACE and GLDAS data to their LS-SVM forecasts during validation. As can be observed in each of these figures, the regression results for GLDAS (Fig. 7) are not as good as those obtained using GRACE (Fig. 6). Each plot corresponds to different values of n
follow the trend of GLDAS data along several months, there is an apparent detachment in the last months of 2014. In this calculation, the regressor dimension p was changed so that the best fitting was attained for the pair (n, p).

The accumulated monthly precipitation over Cantareira can be used as an advanced indicator in good correlation with the EWT. Hence GRACE data series until Jan-15 was used to obtain EWT forecasts for the first three months in 2015 as shown in Fig. 8(a). This plot shows, on the left axis, EWT and LS-SVM forecasts in cm together with the precipitation in mm on the right axis. GRACE signal lags behind precipitation so that, by Jan-15, the perspective of a peak resurgence could be read in the precipitation data and was confirmed by the LS-SVM regressor using GRACE data. It is apparent in Fig. 6, that LS-SVM slightly increased from Jan-13 to Jul-14 when the peak should have occurred in the original GRACE data. For the last

ISSN 2229-5518

GRACE EWT available (until Apr-15), the same series continuation is shown in Fig. 8(b), indicating an EWT intensity equivalent to the May-Jun of 2013, and also in accordance to the corresponding precipitation of that period.

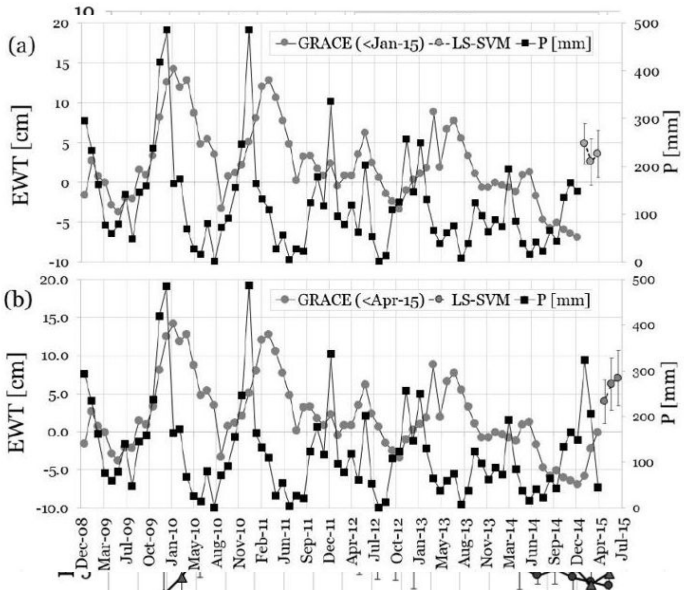

Figure 8(a) GRACE data ending in Jan-165 superimposed to precipitation on Cantareira WSS (in mm). The last three points correspond to LS-SVM forecasts for GRACE (Feb, Mar, Apr of 2015). (b) The same of (a) for last available GRACE data (Abr-15) and , Jun and

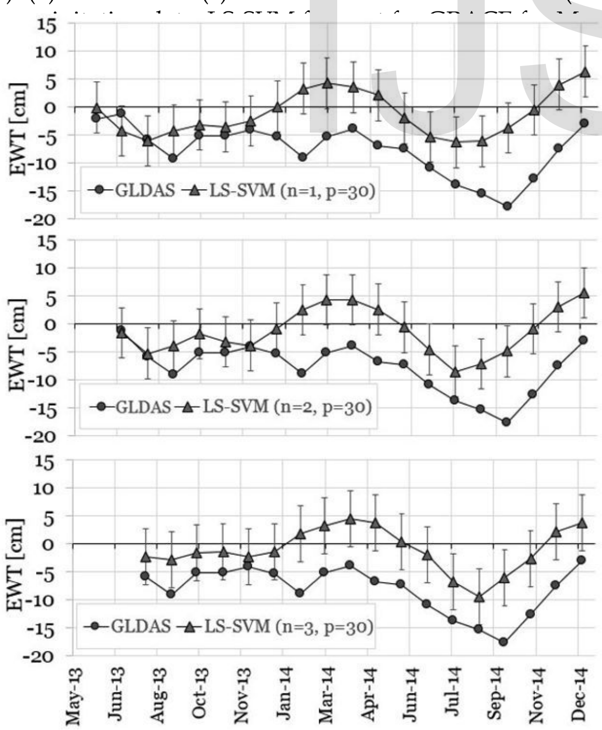

Figure 7 GLDAS data superimposed to LS-SVM forecast for several values of n to expose regressorvalidation (using test data, $\sigma = 3$ and $\gamma = 20$).

4 Conclusion

In this paper we applied remote sensing data such those as provided by GRACE satellite to study their correlation with surface (SABESP) and modelled data (GLDAS). We confirm the correlation between precipitation and watershed differential volume using Cantareira as the worst case scenario observed during 2013-2014 crisis. As shown in Table 2, such correlation is particularly stronger between the differential volume and GLDAS.As for GRACE, the search had to shift the time basis possibly due the way GRACE data are produced (no time scale accuracy). The comparison of the autocorrelation power spectra for both time series indicates that the only common feature is the strong annual peak of$p = 12$. Low period features are particularly absent from GLDAS data.

The application of the LS-SVMregressor during validation (which corresponds to the application of the obtained regressor after training) confirmed the generalization power of the method with all the cautionary notes associated to the validity of GRACE and GLDAS data. The LS-SVM method reproduces the 2013-2014 peak absence for various values of the advanced month (n) as shown in Fig. 6. However, LS-SVM was more successful with GRACE data than GLDAS, although, with this last time series, the overal trend was captured (Fig. 7). Also, during regressor validation, a continuous decrease in the generalization error for both GRACE and GLDAS was found as the number of past values (that is, past information) was increased. However, for GRACE, the error shown a minimum around $p = 22$ months (top of Fig. 5). Since GRACE data is limited to 153 points in this study (with many lacking months being interpolated between adjacent values), it is hard to make any generalization and to conclude that 22 months is the optimal number of periods to be used in LS-SVM for GRACE forecasts. In other words, this result may be an artifact of the implicit erros associated to the time series. Nor were the authors able to find any associated process that could explain such time period. Perhaps by including more cells, it would be possible to confirm the trend. LS-SVM allows data to be entered in an arbitrary way so that time series from adjacent GRACE cells could be added to the input training set$\{\boldsymbol{x}_i, y_i\}$ to further reduce the regressor error at a specific location. This procedure, and their resulting minimized errors - which should implement a type of cell search in order to find the most contributing ones - could map the regions around a given cell that are most relevant to the generalization, thus establishing a stronger type of “learning" correlation. Such search must not be limited to neighbourEWT time series but could use other climate-related data as, for instance, monthly sampled ocean's temperature oscillations, El Niño and La Niña[64][65]. According to this idea, first principle information assists in the choice of relevant data sets to be used in the input vector and possibly the regressor dimension while the method establishes the generalization rule in accordance to its internal formalism.

The period between 2013-2014 was characterized by a water shortage crisis never seen in a city with 11 million people. As a consequence, it becomes particularly important to assess all available data in order to better characterize the situation, both in terms of atmospheric or climate causes as of their ground water potentials. The availabledata show a rich but yet partially unknown dynamics. As the number of pumping wells in the city of São Paulo has increased since the onset of the crisis, the assessment of the groundwater potential adds importance to the task. On the other hand, the resurgence of the peak and the continuous increase in the LS-SVM EWT values after Apr-15 indicated that the water crisis was an exception rather than an enduring situation. Its impact is, however, enduring as the State Government was obliged to reorganize the way water is distributed to the city of São Paulo, reducing the dependence on Cantareira WSS and campaigning for population water saving. Furthermore, from now on lows in differential volumes of all WSS should be closely monitored in pair with actions to avoid a critical state with far more dangerousconsequences. This new hydric regime is established in southern Brazil together with an increase in the droughts in the northern parts (which historically are characterized as semi-arid regions).

In short, the application of GRACE EWT data, using statistical learning tools and the time series for São Paulo city were shown as an opportunity to test such methods in their ability to predict system behaviour in the post-crisis period time.

Acknowledgment

We would like to thank financial support from the Brazilian Ministry of Science, Communication and Inovation (MCTIC). Special thanks go to NASA JPL teams for providing GRACE and GLDAS land data as available at http://grace.jpl.nasa.gov. GRACE is supported by the NASA MEaSUREs Program and GLDAS used the “Goddard Earth Sciences Data and Information Services Center”.

References

[1] S. Corti, F. Molteni and T. N. Palmer. “Signature of recent climate change in frequencies of natural atmospheric circulation regimes”. *Nature*, 398 (6730), pp. 799-802, 1999.

[2] B. Bates, Z. W. Kundzewicz, S. Wu, and J. Palutikof. “Climate Change and Water”. *Technical Paper of the Intergovernmental Panel on Climate Change*, IPCC Secretariat, Geneva, pp. 210, 2008.

[3] C. J. Vörösmarty, P. Green, J. Salisbury and R. B. Lammers. “Global water resources: vulnerability from climate change and population growth”. *Science*, 289(5477), pp. 284-288, 2000.

[4] D. Groenfeldt and J. J. Schmidt. “Ethics and water governance”. *Ecology and Society*, 18(1), p. 14, 2013.

[5] L. S. Pereira, I. Cordery and I. Iacovides. “Coping with water scarcity: Addressing the challenges”. *Technical Documents in Hydrology*, No. 58, 2002.

[6] G. McGranahan, D. Balk and B. Anderson. “The rising tide: assessing the risks of climate change and human settlements in low elevation coastal zones”. *Environment and urbanization*, 19(1). pp.17-37, 2007.

[7] I. Von Frantzius. "World Summit on Sustainable Development Johannesburg 2002: A critical analysis and assessment of the outcomes". *Environmental Politics*, 13(2), pp. 467-473, 2004.

[8] United Nations. “Coping with water scarcity: challenge of the twenty-first century”. *Report for World Water Day 2007*. United Nations, Rome, Italy, pp. 1-29, 2007.

[9] M. W. Rosegrant, X. Cai and S. A. Cline. *World water and food to 2025: dealing with scarcity*. International Food Policy Research Institute, p. 2, 2002.

[10] D. W. Schindler and W. F. Donahue. “An impending water crisis in Canada's western prairie provinces”. *Proceedings of the National Academy of Sciences*, 103(19). pp. 7210-7216, 2006.

[11] A. Hamdy A, M. Abu-Zeid and C. Lacirignola. “Water crisis in the Mediterranean: agricultural water demand management”. *Water International*, vol. 20, no. 4, pp. 176-187, 1995.

[12] E. Weinthal, A. Vengosh, A. Marei, A. Gutierrez and W. Kloppmann. “The water crisis in the Gaza strip: prospects for resolution”. *Groundwater*, vol. 43, no. 5, pp. 653-660, 2005.

[13] P. Troy (Ed).*Troubled waters: Confronting the water crisis in Australia's cities*. Camberra: ANUE Press, p. 19, 2008.

[14] X. J. Zhang, C. Chen, J. Q. Ding, A. Hou, Y. Li, Z. B. Niu, X. Y. Su, Y. J. Xu & E. A. Laws. “The 2007 water crisis in Wuxi, China: analysis of the origin”. *Journal of Hazardous Materials*, vol. 182, no. 1, pp. 130-135, 2010.

[15] H. Thakkar. “Future water solutions for India”. *Development*, vol. 51, no. 1, pp. 68-71,2009.

[16] V. Narain. “India's water crisis: the challenges of governance”. *Water Policy*, vol. 2, no. 6, pp. 433-444, 2000.

[17] J. Skinner. “Where every drop counts: tackling rural Africa’s water crisis”. London, UK: *International Institute for Environment and Development*, 2009. Available at http://pubs.iied.org/17055IIED/

[18] F. R. Rijsberman. “Water scarcity: Fact or fiction?”*Agricultural water management*, 80(1), pp. 5-22, 2006.

[19] M. Kummu, P. J. Ward, H. de Moel and O. Varis. "Is physical water scarcity a new phenomenon? Global assessment of water shortage over the last two millennia". *Environmental Research Letters*, vol. 5, no. 3, 034006, 2010.doi:10.1088/1748-9326/5/3/034006.

[20] P. Rogers. “Water governance in Latin America and the Caribbean”. Inter-American Development Bank, 2002.https://publications.iadb.org/handle/11319/3966

[21] A. D. Moura and J. Shukla. “On the dynamics of droughts in northeast Brazil: Observations, theory and numerical experiments with a general circulation model”. *Journal of the Atmospheric Sciences*, vol. 38, no. 12, pp. 2653-2675, 1981.

[22] V. B. Rao, P. Satyamurty and J. I. O. De Brito. “On the 1983 drought in North-East Brazil”. *Journal of climatology*, vol. 6, no. 1, pp. 43-51, 1986.

[23] V. P. R. Da Silva. “On climate variability in Northeast of Brazil”. *Journal of Arid Environments*, vol. 58, no. 4, pp. 575-596, 2004.

[24] L. Amore. ‘The Guarani Aquifer: from knowledge to water management”. *International Journal of Water Resources Development*, vol. 27, no. 3, 463-476, 2011.

[25] F. Sindico. “The guarani aquifer system and the international law of transboundary aquifers”. *International Community Law Review*, vol. 13, no. 3, pp. 255-272, 2011.

[26] E. H. Berbery and E. A. Collini. “Springtime precipitation and water vapor flux over southeastern South America”. *Monthly Weather Review*, vol. 128, no. 5, pp. 1328-1346, 2000.

[27] E. H. Berbery and V. R. Barros. “The hydrologic cycle of the La Plata basin in South America”. *Journal of Hydrometeorology*, vol. 3, no. 6, pp. 630-645, 2002.

[28] C. A. Nobre, G. Sampaio, L. Salazar. “Scenarios for climate change

for South America for the end of the 21st century". In: *Brazil and climate change: vulnerability, impacts and adaptation*. Ed. CGEE, Brasília, pp. 21-36, 2009.

[29] M. A. De Lima, B. J. R. Alves. "Vulnerabilities, impacts and adaptations to climate change in the agricultural sectors and agricultural soils". In: *Brazil and climate change: vulnerability, impacts and adaptation*. Ed. CGEE, Brasilia, pp. 69-95, 2009.

[30] M. A. Vasconcelos de Freitas, J. L. da Silva Soito. "Energy and water resources: vulnerability, impacts and possibilities for adapting hydroelectric power generation in Brazil to global climate change". In *Brazil and climate change: vulnerability, impacts and adaptation*. Ed. CGEE, Brasília, pp. 165-191, 2009.

[31] B. P. F. Braga. Integrated urban water resources management: a challenge into the 21st century. *International Journal of Water Resources Development*, vol. 17, no. 4, pp. 581-599, 2001.

[32] B. P. F. Braga, M. F. A. Porto and R. T. Silva. Water management in metro politan São Paulo. *Water Resources Development*, vol. 22, no. 2, pp. 337-335, 2006.

[33] SABESP. CHESS: Crise hídrica, estratégias e soluções SABESP para a região metropolitana de São Paulo. São Paulo. SABESP, 2015. Availableat: http://site.sabesp.com.br/site/uploads/file/crisehidrica/chess_crise_hidrica.pdf

[34] D. L. Swain, M. Tsiang, M. Haugen,D. Singh, A. Charland, B. Rajaratnam and N. S. Diffenbaugh. "The extraordinary California drought of 2013-2014: Character, context, and the role of climate change". *Bull Am MeteorolSoc*, vol. 95, no. 7, S3-S7, 2014.

[35] F. W. Landerer and S. C. Swenson. "Accuracy of scaled GRACE terrestrial water storage estimates". *Water Resources Research*, vol. 48, no. 4, 2012.doi: 10.1029/2011WR011453

[36] S. Swenson and J. Wahr. "Post-processing removal of correlated errors in GRACE data". *Geophysical Research Letters*, 33(8), 2006.doi: 10.1029/2011WR011453

[37] M. Rodell, P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. J. Meng, K. Arsenault, B. Cosgrove, K. Radakovich, M. Bosilovich,J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll. "The Global Land Data Assimilation System". *Bulletin of the American Meteorological Society*, vol. 85, no. 3, pp. 381-394, 2004.

[38] B. P. F. Braga, M. F. A. Porto and R. T. Silva. "Water management in metropolitan São Paulo". *Water Resources Development*, vol. 22, no. 2, pp. 337-35, 2006.

[39] D. Rodrigues, H. Gupta, A. Serrat-Capdevila, P. Oliveira, E. Mario Mendiondo, T. Maddock III and M. Mahmoud. "Contrasting American and Brazilian Systems for Water Allocation and Transfers". *J Water Resour Plan Man.*, vol. 141, no. 7, 04014087, 2015.

[40] I. Wahnfried and R. Hirata. "Comparação dos métodos de estimativa de recarga de aquíferos em uma planície aluvionar na bacia hidrográfica do Alto Tietê (São Paulo) ". *Revista Brasileira dos Recursos Hídricos*. vol. 1, no. 1, pp. 15-25, 2005.

[41] R. Hirata, A. Suhogusoff and A. Fernandes. "Groundwater resources in the State of São Paulo (Brazil): the application of indicators". *Anais da Academia Brasileira de Ciências*, vol. 79, no. 1, pp. 141-152, 2007.

[42] J. A. Suykens, T. Van Gestel, B. De Moor and J. Vandewalle. "Basic Methods of Least Squares Support Vector Machines". In: *Least Squares Support Vector Machines*. World Scientific Publishing Co. Pte. Ltd, pp. 71-116, 2002.

[43] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2nd Ed, pp. 247-249, 2013.

[44] N. Sapankevych and R. Sankar. "Time series prediction using support vector machines: a survey". *Computational Intelligence Magazine, IEEE*, vol. 4, no. 2, pp. 24-38, 2009.

[45] M. Espinoza, J. A. Suykens and B. De Moor. "Short-term chaotic time series prediction using symmetric LS-SVM regression". In: *Proc. of the 2005 International Symposium on Nonlinear Theory and Applications* (NOLTA), pp. 606-609, 2005.

[46] T. Van Gestel, J. A. Suykens, D. E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor and J. Vandewalle. "Financial time series prediction using least squares support vector machines within the evidence framework". *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 809-821, 2001.

[47] G. J. Huffman, R. F. Adler, B. Rudolf, U. Schneider and P. R. Keehn. "Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information". *Journal of Climate*, vol. 8, no. 5, pp. 1284-1295, 1995.

[48] C. Sivapragasam, S. Liong and M. F. K. Pasha. "Rainfall and runoff forecasting with SSA-SVM approach". *Journal of Hydroinformatics*, vol. 3, no. 3, pp. 141-152, 2001.

[49] S. Tripathi, V. V. Srinivas and R. S. Nanjundiah. "Downscaling of precipitation for climate change scenarios: a support vector machine approach". *Journal of Hydrology*, vol. 330, no. 3, pp. 621-640, 2006.

[50] J. Wu, M. Liu and L. Jin. Least square support vector machine ensemble for daily rainfall forecasting based on linear and nonlinear regression. In Advances in neural network research and applications. Springer Berlin Heidelberg, pp. 55-64, 2010.

[51] U. Okkan and Z. A. Serbes. "Rainfall–runoff modeling using least squares support vector machines". *Environmetrics*, vol. 23, no. 6, pp. 549-564, 2012.

[52] L. J. Herrera, H. Pomares, I. Rojas, A. Guillén, A. Prieto and O. Valenzuela. "Recursive prediction for long term time series forecasting using advanced models". *Neurocomputing*, vol. 70, no. 16, pp. 2870-2880, 2007.

[53] E. Deza and M. M. Deza. *Encyclopedia of Distances*. Springer Berlin Heidelberg. p. 94, 2009.

[54] M. Clerc. Particle swarm optimization, John Wiley & Sons, vol. 93, p. 21, 2010.

[55] G. Zhiqiang, W. Huaiqing and L. Quan L. "Financial time series forecasting using LPP and SVM optimized by PSO". *Soft Computing*, vol. 17, no. 5, pp. 805-818, 2013.

[56] O. L. Colombo. "The global mapping of gravity with two satellites". *RijkscommissievoorGeodesie*, vol. 7, no. 3, p. 13, 1984.

[57] R. Schmidt, P. Schwintzer, F. Flechtner, C. Reigber, A. Güntner, P. Döll and J. Wünsch. "GRACE observations of changes in continental water storage". *Global and Planetary Change*, vol. 50, no. 1, pp. 112-126, 2006.

[58] M. Rodell and J. S. Famiglietti. "The potential for satellite-based monitoring of groundwater storage changes using GRACE: The High Plains aquifer, Central US."*Journal of Hydrology*, vol. 263, no. 1, pp. 245-256, 2002.

[59] M. Rodell, P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. J. Meng, K. Arsenault, B. Cosgrove, K. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll. "The Global Land Data Assimilation System". *Bulletin of the American Meteorological Society*, vol. 85, no. 3, pp. 381-394, 2004.

[60] T. H. Syed, J. S. Famiglietti, M. Rodell, J. Chen and C. R. Wilson. Analysis of terrestrial water storage changes from GRACE and GLDAS. Water Resource Research, vol. 44, no. 2, 2008. doi: 10.1029/2006WR005779

[61] P. M. T. Broersen. *Automatic autocorrelation and spectral analysis*. Springer Science & Business Media, p. 1, 2006.

[62] E. Rangelova, W. Van der Wal, A. Braun, M. G. Sideris and P. Wu. "Analysis of Gravity Recovery and Climate Experiment time variable mass redistribution signals over North America by means of principal component analysis". *Journal of Geophysical Research: Earth Surface*. (2003–2012), vol. 112, no. F3, 2007.

[63] A. Güntner. "Improvement of global hydrological models using GRACE data". *Surveys in Geophysics*, vol. 29, no. 4-5, pp. 375-397, 2008.

[64] C. F. Ropelewski and M. S. Halpert. "Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation". *Monthly weather review*, vol. 115, no. 8, pp. 1606-1626, 1987.

[65] A. M. Grimm, S. E. Ferraz and J. Gomes. "Precipitation anomalies in southern Brazil associated with El Niño and La Niña events". *Journal of Climate*, vol. 11, no. 11, pp. 2863-2880, 1998.