

Artigos

Utilização de máquina de vetor de suporte para previsão de níveis de água subterrânea

Use of support vector machine for prediction of groundwater levels

Thiago Boeno Patricio Luiz¹; Guilherme Freitas Gaiardo¹; José Luiz Silvério da Silva¹ ✉

¹ Universidade Federal de Santa Maria (UFSM), Santa Maria, RS, Brasil.

✉ thiagoboeno@hotmail.com, ggaiardo@inf.ufsm.br, silverioufsm@gmail.com

Resumo

Palavras-chave:

Hidrogeologia.
Lençol freático.
Modelos de regressão.
Gestão de recursos hídricos.

Este trabalho teve como objetivo utilizar uma técnica de regressão com Máquina de Vetor de Suporte para prever níveis diários de água subterrânea de um poço de monitoramento em aquífero livre no oeste do Rio Grande do Sul. Para tanto, empregou-se séries temporais de nível estático e de precipitação pluviométrica, compreendidos em um período de monitoramento diário de 1468 observações. O intuito foi criar um modelo autorregressivo capaz de estimar os níveis de água subterrânea, a partir das séries históricas para horizontes de 120 e de 180 dias. Na etapa de teste do modelo, os ajustes das previsões realizadas apresentaram coeficientes de determinação estatístico (R^2) de 0,89 e coeficientes de Nash-Sutcliffe (CNS) de 0,90, atestando a qualidade preditiva da metodologia empregada. Após a análise dos resíduos, constatou-se maior eficiência para a previsão de eventos de descarga, principalmente, para níveis estáticos altos. O uso da metodologia mostrou grande precisão para estimar os níveis de água subterrânea utilizando apenas as variáveis de nível estático e de precipitação, demonstrando grande aplicabilidade na área da hidrologia subterrânea.

Abstract

Keywords:

Hydrogeology.
Water-table.
Regression models.
Water resources management.

This work had a purpose to use a regression technique with Support Vector Machine (SVM) to predict daily water-table levels of groundwater in a monitoring well on unconfined aquifer in the west of Rio Grande do Sul. For that, was used time series of static water level and rainfall, comprising daily monitoring period of 1468 observations. The aim was to create an autoregressive model capable of estimate groundwater levels from the time series for horizons of 120 and 180 days. At the test step of the model, the adjustment of prediction realized presented statistical coefficients (R^2) of 0,89 and Nash-Sutcliffe coefficients (CNS) of 0,90, attesting the predictive quality of the methodology used. After the analysis of the residues, it was verified greater efficiency for the prediction of discharge events, mainly, for high static water levels. The use of present methodology showed high accuracy to estimate the groundwater level using only the static water level and rainfall variables, demonstrating great applicability in groundwater hydrology field.

Revisado por pares.

Recebido em: 23/08/2017.

Aprovado em: 08/01/2018.

DOI: <http://dx.doi.org/10.14295/ras.v32i1.28921>

1. INTRODUÇÃO

A problemática mundial de escassez hídrica vem levantando uma série de desafios aos gestores dos recursos hídricos, haja vista que a exploração das reservas de água para consumo humano depende, fundamentalmente, da sua acessibilidade e de suas características qualitativas. Para as águas subterrâneas, o uso sustentável do manancial está sujeito a diversos aspectos, entre eles as condições de armazenamento, a potencialidade aquífera e as taxas de recarga.

No Brasil, muitas cidades utilizam água subterrânea para garantir o abastecimento público, de maneira que o controle das reservas hídricas nessas localidades são assuntos de interesse para os órgãos públicos responsáveis. A exploração excessiva de água subterrânea, principalmente em épocas de maior estiagem, pode trazer situações de esgotamento hídrico, além de possíveis

impactos a sua qualidade natural. A extração em demasia de água subterrânea pode acarretar, entre outros, na redução sensível das descargas aquíferas, nos volumes de lagoas e do fluxo de base de rios além de comprometer as características ambientais de sistemas lacustres, brejos e pântanos (FEITOSA et al., 2008).

No Estado do Rio Grande do Sul, a correta gestão dos recursos hídricos subterrâneos possui um papel central para o desenvolvimento sustentável das atividades econômicas da região, uma vez que a maioria dos municípios do Estado utiliza águas subterrâneas para abastecimento público (ANA, 2010). Além deste, outro grande uso da água subterrânea é destinado para a agricultura, onde a maior parte das outorgas são para atividades de irrigação de culturas como soja, milho e arroz.

Nesse contexto, metodologias que ofereçam suporte às tomadas

de decisão por parte da sociedade quanto à gestão dos recursos hídricos subterrâneos são importantes, pois disponibilizam ferramentas capazes de estimar a ocorrência de situações críticas, podendo evitar assim, maiores danos ambientais.

Desde o início da década de 1990, modelos baseados em Máquinas de Vetores de Suporte (*Support Vector Machine - SVM*) vêm ganhando maior destaque na academia e indústria por apresentar resultados semelhantes ou até mesmo melhores quando comparadas à outras técnicas em algumas tarefas de classificação e regressão (SCHÖLKOPF et al., 1996). Em muitos casos, os resultados da aplicação do SVM têm se mostrado superiores quando comparados aos obtidos por outros algoritmos, como por exemplo as redes neurais artificiais, k-vizinho mais próximo ou árvores de decisão (RAGHAVENDRA & DEKA, 2014).

Nos últimos anos, as técnicas baseadas em inteligência artificial ou aprendizagem de máquina (*machine learning*) têm evoluído em grande escala devido ao avanço computacional e ao forte interesse de fazer prognósticos em cenários hipotéticos para o futuro. Na área da hidrologia, essas metodologias têm sido amplamente utilizadas com sucesso para predições de dados pluviométricos, regimes de vazões em bacias hidrográficas, níveis de lagos e de reservatórios (ASEFA et al., 2006; LIN et al., 2006; KHAN & COULIBALY, 2006; TRIPATHI et al., 2006; BEHZAD et al., 2008; RAGHAVENDRA & DEKA, 2014).

Basicamente, as metodologias que utilizam aprendizagem de máquinas para previsão são técnicas de classificação e de regressão em que o enfoque principal é “aprender” a partir dos da-

dos ou das amostras disponíveis. Trata-se de um campo de pesquisa que utiliza a inteligência artificial computacional como fator de desenvolvimento de métodos capazes de extrair conceitos a partir de amostras de dados (MITCHELL, 1997).

Dessa forma, o presente estudo objetivou aplicar um modelo de regressão baseado em Máquina de Vetores de Suporte em dados de nível de água subterrânea em um poço de monitoramento e da precipitação pluviométrica com o intuito de criar um modelo que permita prever as oscilações diárias dos níveis deste poço. Esse trabalho visa contribuir para o desenvolvimento de metodologias e técnicas baseadas em aprendizado de máquinas para a gestão dos recursos hídricos subterrâneos.

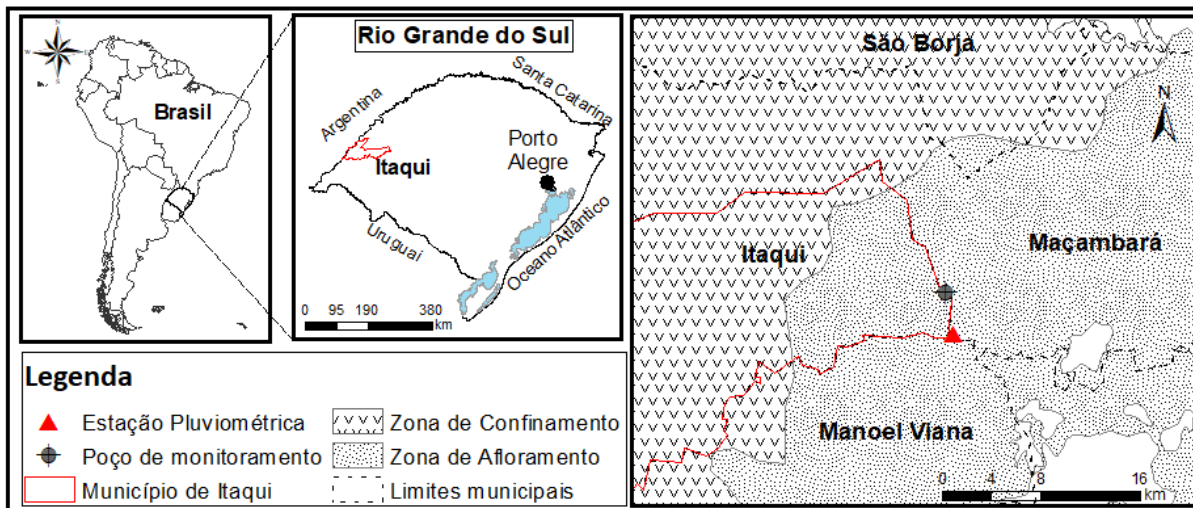
2. MATERIAIS E MÉTODOS

2.1. Localização e descrição da área de estudo

O poço de monitoramento e a estação pluviométrica utilizada para este estudo estão representados na Figura 1, localizam-se no município de Itaqui, no oeste do estado do Rio Grande do Sul, sul do Brasil.

O município de Itaqui possui uma extensão territorial de aproximadamente 3406 km², está inserido predominantemente na Bacia do Rio Ibicuí na Região Hidrográfica do Uruguai, possuindo uma população de 38.159 pessoas (IBGE, 2010). A base da economia do município é a agropecuária, onde destacam-se a criação de animais e o cultivo de arroz irrigado, a principal fonte econômica do município.

Figura 1 - Localização do poço de monitoramento e da estação pluviométrica no município de Itaqui/RS



Fonte: MACHADO & FREITAS (2005); RIMAS (2016); HIDROWEB (2016)

O poço tubular utilizado no estudo (Figura 2), está localizado na região de Puitã (RIMAS, 2016), a cerca de 3,5 km de distância da estação pluviométrica utilizada. Em parte do município de Itaqui, ocorre a Zona de Afloramentos do Sistema Aquífero Guarani (SAG), que são rochas porosas responsáveis pela recarga/des-

carga de água subterrânea. Na porção oeste do município, o SAG encontra-se confinado sob as rochas da Formação Serra Geral, formando a Zona de Confinamento do SAG (MACHADO & FREITAS, 2005).

Figura 2 - Poço de monitoramento no município de Itaqui/RS



Fonte: RIMAS (2016)

O poço de monitoramento possui 40,5 metros de profundidade, em uma cota topográfica de 107,06 metros acima do nível do mar. Seu perfil litológico compreende camadas intercaladas de arenitos finos a médios, sendo considerado um poço com nível freático raso, com alta porosidade e de condição livre.

O clima na região é definido segundo a classificação de Köppen como subtropical úmido, tipo Cfa, caracterizado pela presença de invernos e verões bem definidos e inexistência de estação seca, com precipitações bem distribuídas em todos os meses do ano (ALVARES et al., 2013).

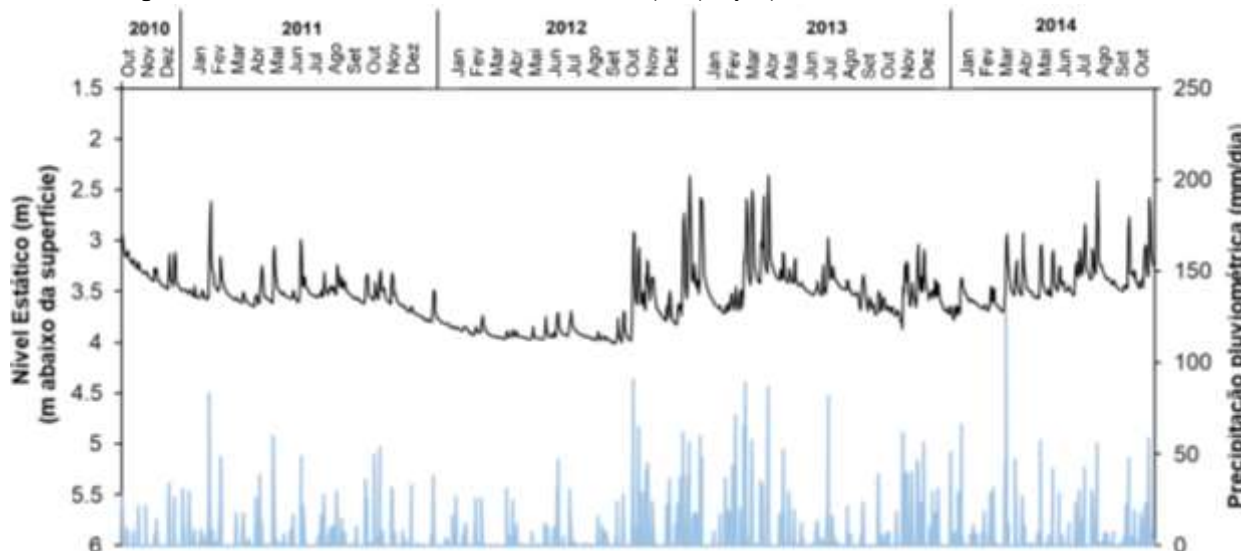
2.2. Caracterização dos dados

Para construção do modelo, considerou-se a série histórica de 1468 observações diárias do nível estático e da precipitação pluviométrica, compreendidos entre 7 de outubro de 2010 a 13 de

outubro de 2014, abarcando um período de 4 anos e 6 dias de monitoramento contínuo. Os dados dos níveis estáticos foram adquiridos através da Companhia de Pesquisa de Recursos Minerais (CPRM) por meio da Rede Integrada de Monitoramento de Águas Subterrâneas (RIMAS), estação n° 4300020119 localizada na porção leste do município de Itaqui. As medições dos níveis de água subterrânea foram pareadas a cada 24 horas, obtidas por medidor automático de nível equipado com *datalogger*.

Os dados de precipitação pluviométrica foram adquiridos junto à Agência Nacional das Águas (ANA) do Brasil através do Sistema de Informações Hidrológicas (HIDROWEB) (HIDROWEB, 2016), por meio de pluviômetro da estação Cachoeira Santa Cecília n° 02955002. As séries diárias estão representadas na Figura 3 e não apresentaram falhas de medições no período considerado.

Figura 3 - Gráficos diários das séries de nível estático e de precipitação pluviométrica utilizadas no estudo



Em grande parte do período, as chuvas ficaram distribuídas de maneira uniforme e apresentaram, em geral, boas correlações visuais com as recargas subterrâneas. As precipitações pluviométricas médias anuais registradas para o município de Itaqui são de 1812 milímetros (ALVARES et al., 2013), todavia os totais

anuais dos anos 2011, 2012 e 2013 registradas pela estação Cachoeira Santa Cecília ficaram em 1248, 1363 e 1905 milímetros respectivamente. A amplitude dos níveis estáticos no período de análise atingiu 1,65 metros.

2.3. Revisão bibliográfica

Algoritmos computacionais baseados em SVM fundamentam-se na teoria de aprendizagem de máquinas (*machine learning theory*) (VAPNIK, 1995; 1998). Esses algoritmos “aprendem” com a série de dados, isto é, extraem informações que identificam a influência ou peso de cada parâmetro (e.g. precipitação, níveis estáticos anteriores) no resultado final. Algoritmos baseados em SVM são aplicados em diversos ramos da ciência em problemas de classificação e regressão (RAGHAVENDRA & DEKA, 2014).

A metodologia de SVM é uma técnica de aprendizado de máquina supervisionado, ou seja, recebe um conjunto de N amostras na forma par ordenado (X, Y) , onde $X \in R_m$ é um vetor de m características que descrevem a amostra e $Y \in R$ é o valor esperado da amostra. Com base nessas informações será escolhida uma função, ou classificador, que tomando como entrada um vetor de características, prevê o valor esperado da amostra que é desconhecido. Um classificador linear é da forma $f(X) = \langle W, X \rangle + b$, onde W é um vetor de pesos. Para determinar a função, Vapnik (1995) propôs o seguinte problema de otimização:

$$\begin{aligned} & \text{Minimizar } \mathbf{W}, b, \xi, \xi^* \text{ em } \frac{1}{2} \|\mathbf{W}\|^2 + C \sum_{k=1}^N (\xi_k + \xi_k^*) \\ & \text{Sujeito a } \left\{ \begin{array}{l} y_k - \mathbf{W}^T \phi(\mathbf{X}_k) - b \leq \varepsilon + \xi_k \\ \mathbf{W}^T \phi(\mathbf{X}_k) + b - y_k \leq \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \geq 0 \end{array} \right\} \quad k = 1, 2, \dots, N \end{aligned} \quad (1)$$

Nesta formulação existe o parâmetro ε que é o erro máximo permitido para as estimativas. As variáveis de folga, ξ e ξ^* permitem erros na função. Isso consente que sempre exista uma solução para o problema de minimização, mesmo que existam casos onde a restrição de ε não é respeitada. $C \geq 0$ é uma constante que determina o quão plana será a função e a quantidade que os desvios maiores que ε podem ser tolerados.

Para uma classificação/regressão não linear é introduzido o conceito de função *Kernel*, que serve para mapear implicitamente os vetores de características para um espaço de maior dimensão onde supõe-se que os dados são linearmente separáveis. Ao mapear os vetores de forma não linear e aplicando a técnica, resulta uma função não linear no espaço original, mesmo que ela seja linear em outro espaço (SMOLA & SCHÖLKOPF, 2004).

Na área da hidrologia subterrânea, metodologias baseadas em regressão ou aprendizado estatístico têm produzido resultados relevantes no que concerne à simulação de séries temporais de níveis de água subterrânea, principalmente em escalas de tempo mensais, semanais e diárias. A utilização deste tipo de metodologia tem sido utilizada para a previsão de dados piezométricos em diferentes contextos geológicos como em Taormina et al. (2006), Yoon et al. (2011), Shirmohammadi et al. (2012) e Shiri et al. (2013).

Para Yoon et al. (2011) e Shiri et al. (2013), a utilização da técnica baseada em SVM permitiu a obtenção de bons resultados para realizar previsões de dados piezométricos de água subterrânea proveniente de aquíferos livres. Os autores alcançaram altos índices de acerto nas etapas de calibração e de previsão dos modelos criados, em diferentes aquíferos da Coréia do Sul (costeiro e aluvionar respectivamente).

No Brasil, destaca-se a utilização de modelos estocásticos em séries de níveis de água subterrânea em trabalhos de Manzione et al. (2010) na região do Cerrado no estado de Goiás e Manzione et al. (2012) na região da Zona de Afloramentos do SAG no estado de São Paulo. Por meio da aplicação do modelo autorregressivo PIRFICT, os autores simularam as flutuações dos níveis da água subterrânea e elaboraram possíveis cenários nas áreas estudadas.

Metodologias baseadas em SVM na hidrogeologia ainda são pouco empregadas, sobretudo no Brasil. O uso da técnica em dados de sistemas aquíferos brasileiros e sob diferentes condições climáticas podem trazer alterações em sua performance, uma vez que as diferentes correlações entre as variáveis usadas na calibração do modelo tendem a gerar diferentes resultados.

2.4. Estrutura de programação empregada e estimadores utilizados

Para calcular a função de regressão dos dados, utilizou-se a ferramenta LIBSVM (CHANG & LIN, 2011), a qual possui uma interface de programação em Python (PYTHON SOFTWARE FOUNDATION, 2016), que foi a linguagem escolhida para o processamento das informações.

Inicialmente, os dados brutos de nível estático foram analisados a partir da aplicação da Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP), procurando identificar a melhor estrutura de entrada para a série de nível a fim de realizar o treinamento (calibração) dos dados em SVM. Nesta etapa, utilizou-se as primeiras 1288 observações de medições do nível estático e da precipitação, reservando os 180 dias finais da série histórica para a etapa de verificação, a fim de comparar o valor observado com o valor estimado pelo SVM no poço de monitoramento nesse período.

A definição da melhor estrutura de entrada no algoritmo para os dados de precipitação pluviométrica foi determinada a partir do teste de várias combinações e, observando seus ajustes amostrais. Dessa forma, testou-se as estruturas baseadas em 1, 3, 5, 7 e 10 dias, ou seja, utilizaram-se como vetores os valores das medições de chuvas em até 10 dias anteriores ao dia que se deseja prever o nível. O objetivo desse procedimento foi o de avaliar a estrutura que melhor se adapte para a modelagem dos dados observados, constituindo-se como fator central para o desenvolvimento do modelo, uma vez que a melhor estrutura tende a gerar uma previsão com menor erro.

Após calibrado, o modelo permite que sejam feitas previsões para tantos dias quanto necessário, onde o erro cometido durante o processo de previsão tende a acumular-se em cada iteração. As previsões foram traçadas para um intervalo de 120 e 180 dias, compreendidos entre os dias 16/04/2014 a 14/08/2014 e 16/04/2014 a 13/10/2014 respectivamente, totalizando 4 e 6 meses que não foram considerados na calibração do modelo. Após realizadas as previsões, avaliou-se os erros cometidos no processo por meio da análise de resíduos, buscando validar a utilização do modelo e observar possíveis tendências no processo.

Para estimar os ajustes em cada etapa da modelagem bem como a eficiência do modelo, calcularam-se além do coeficiente de determinação (R^2), o coeficiente de Nash-Sutcliffe (CNS) buscando-se avaliar o poder preditivo do modelo hidrológico criado e, o Erro Quadrático Médio (EQM). Os estimadores de ajuste utilizados estão representados por meio das equações (2), (3) e (4), da seguinte forma:

$$R^2 = \frac{\sum_{i=1}^k (o_i - \bar{o})^2}{\sum_{i=1}^k (t_i - \bar{t})^2 + (t_i - o_i)^2} \quad (2)$$

$$CNS = 1 - \frac{\sum_{i=1}^k (t_i - o_i)^2}{\sum_{i=1}^k (t_i - \bar{t})^2} \quad (3)$$

$$EQM = \frac{1}{k} \sum_{i=1}^k (t_i - o_i)^2 \quad (4)$$

Onde o_i são os dados observados e t_i são os valores estimados. O CNS varia de $-\infty$ a 1, sendo que a unidade corresponde a uma equivalência perfeita, enquanto uma eficiência igual a zero indica que as previsões do modelo são tão precisas quanto a mé-

dia dos dados. Já o EQM foi calculado a fim de determinar em que medida o modelo não se ajustou aos dados.

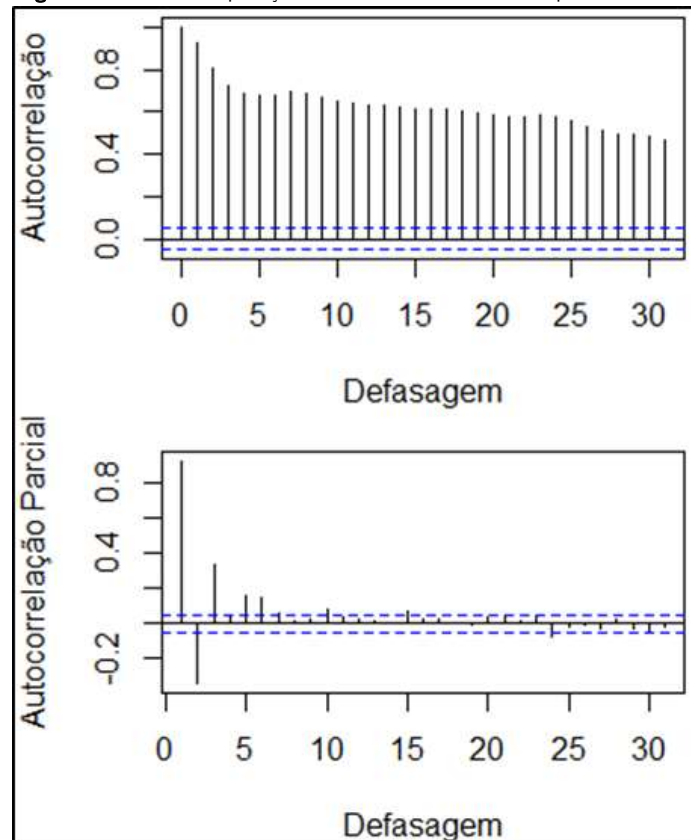
3. RESULTADOS E DISCUSSÃO

3.1 Etapa de treinamento do modelo

Para a calibração do modelo, avaliou-se em um primeiro momento, a autocorrelação dos dados de nível estático, a fim de fornecer a melhor estrutura de entrada no algoritmo. A Figura 4 apresenta a aplicação da Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) na série de dados de nível estático.

A aplicação da FAC na série histórica de nível apresentou uma alta correlação para todas as defasagens propostas. Em contrapartida, a aplicação da FACP apresentou correlação significativa até a defasagem 3, indicando que modelos até a terceira ordem de correlação são estatisticamente suficientes para explicar o fenômeno em estudo.

Figura 4 - Gráficos da aplicação da FAC e FACP na série temporal de nível estático



Deste modo, escolheu-se uma estrutura na ordem 3 do nível estático para realizar a calibração do modelo, ou seja, utilizou-se as medições dos 3 dias anteriores (h_{t-2} , h_{t-1} e h_t) do nível para compor a estrutura de entrada no SVM. Esta composição de entrada para a modelagem utiliza o valor dos 3 níveis anteriores ao dia que se deseja prever para formar a função classificadora, onde o algoritmo toma como base essas observações para estimar o próximo nível e assim sucessivamente.

Para a definição da melhor estrutura para a precipitação pluviométrica, testou-se a sequência de 1, 3, 5, 7 e 10 observações anteriores (P_{t-9} , ..., P_t), buscando a melhor correlação entre a série de dados de chuva para calibração. Os resultados dos ajustes obtidos nas diferentes estruturas testadas são apresentados na tabela 1.

Tabela 1 - Valor do coeficiente de determinação para diferentes estruturas de entrada no modelo para a calibração

Estrutura	R ²
$h_{t-2}, h_{t-1}, h_t + P_t$	0,89
$h_{t-2}, h_{t-1}, h_t + P_{t-2}, P_{t-1}, P_t$	0,90
$h_{t-2}, h_{t-1}, h_t + P_{t-4}, P_{t-3}, P_{t-2}, P_{t-1}, P_t$	0,94
$h_{t-2}, h_{t-1}, h_t + P_{t-6}, P_{t-5}, P_{t-4}, P_{t-3}, P_{t-2}, P_{t-1}, P_t$	0,89
$h_{t-2}, h_{t-1}, h_t + P_{t-9}, \dots, P_t$	0,76

Tendo em vista os resultados apresentados na Tabela 1, escolheu-se a estrutura que utiliza os últimos 3 dias de nível estático e os 5 dias anteriores de chuva. O modelo com essa estrutura apresentou o melhor ajuste para realizar as previsões dos níveis estáticos de água subterrânea para esse poço de monitoramento.

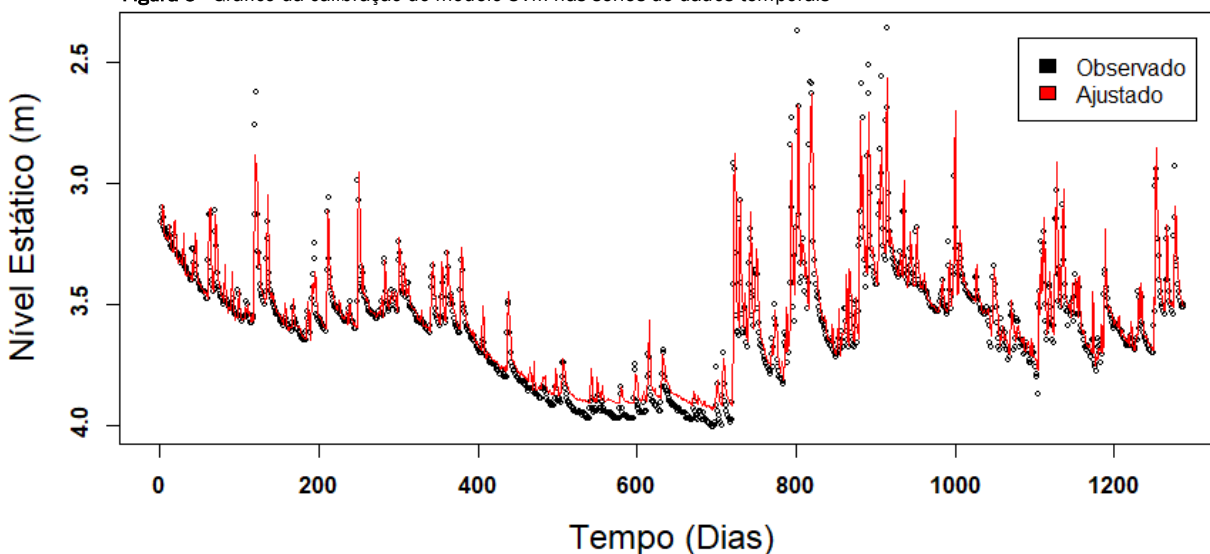
Em geral, quando se deseja obter um modelo robusto e confiável para realizar a simulação de variáveis em séries temporais, utilizam-se como *inputs* os principais aspectos responsáveis pela variabilidade do fenômeno sob estudo. No caso dos níveis de água subterrânea, a precipitação é o principal fator do ciclo hidrológico responsável pelas oscilações desta variável, uma vez que influi diretamente nas recargas aquíferas.

Alguns autores, ainda, adicionaram outras séries de dados com o propósito de melhorar as previsões da piezometria em diferentes aquíferos. No estudo de Yoon et al. (2011), a variável de nível

da maré do oceano foi adicionada no treinamento dos dados, já que essa variável pode influenciar nos níveis de água subterrânea em aquíferos costeiros. Shiri et al. (2013), optou pela adição da variável de evapotranspiração para o treinamento do modelo, obtendo uma pequena melhora para a calibração em SVM. Contudo, ambos autores salientam que as séries históricas dos níveis de água e da precipitação são as mais importantes nas simulações que utilizam o SVM, onde a melhora de performance quando utilizado outras variáveis é, muitas vezes, pequena.

Dessa forma, a precipitação pluviométrica foi considerada como a principal variável responsável pelas oscilações dos níveis de água subterrânea neste poço, procurando, também, criar um modelo parcimonioso e de aplicação prática. Sendo assim, utilizou-se a estrutura definida a partir da Tabela 1 para realizar o treinamento do modelo. Na Figura 5 apresenta-se o ajuste obtido pela técnica de SVM e os dados observados.

Figura 5 - Gráfico da calibração do modelo SVM nas séries de dados temporais



Ao analisar o gráfico da Figura 5, percebe-se que o modelo ajustou-se satisfatoriamente aos dados observados, obtendo um coeficiente de determinação de 94%, mostrando a eficiência deste procedimento para este trecho da série temporal. O valor do coeficiente permite inferir que o modelo ajustado tende a possuir um bom poder de previsão, estando em conformidade com os valores obtidos na etapa de treinamento por outros autores como Taormina et al. (2006), Yoon et al. (2011), Shirmohammadi et al. (2012) e Shiri et al. (2013).

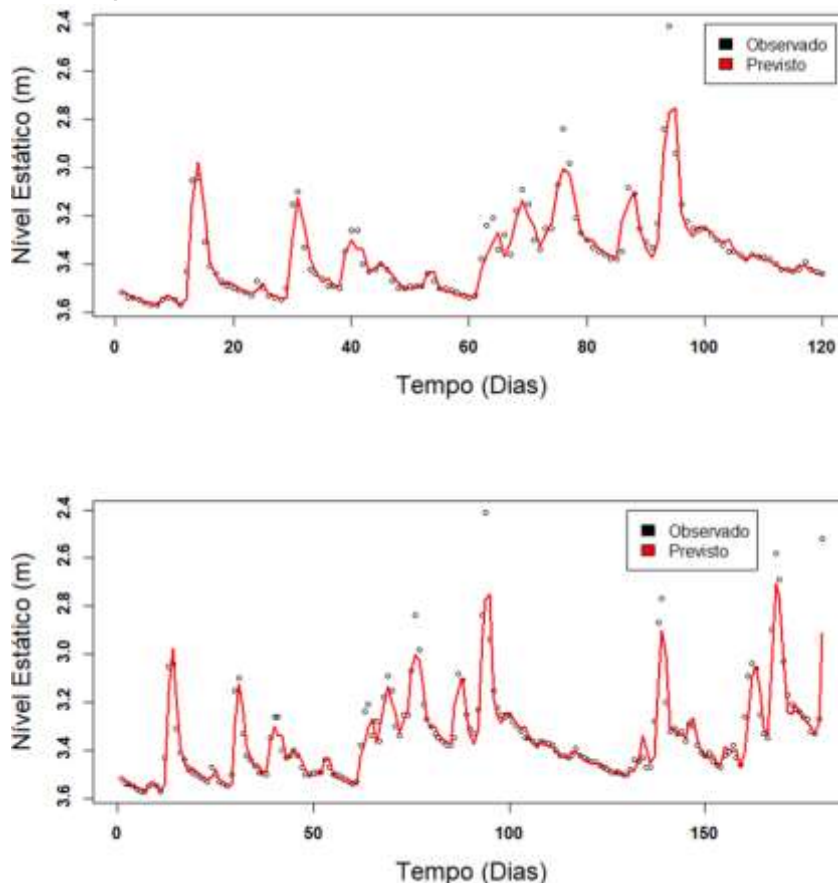
3.2. Etapa de verificação e avaliação

Uma vez gerado o modelo de regressão por meio das séries históricas de nível de água e de precipitação utilizando a metodolo-

gia de SVM, partiu-se para a criação dos horizontes de 120 e 180 dias, estimando o h_{t+1} em cada iteração. Dessa forma, a partir do modelo calibrado, plotou-se na Figura 6 as previsões realizadas e os dados observados no mesmo período.

A partir dos dados preditos pelo modelo, apresentado nos gráficos da Figura 6, nota-se que os mesmos apresentam tendências de crescimento e decaimento semelhantes ao real medido no poço de monitoramento, ou seja, possuem boa aderência à série real observada. Os coeficientes de determinação para essa fase do estudo, em ambas simulações, apresentaram correlações de 89%, evidenciando, portanto, uma boa margem de acerto pela técnica utilizada.

Figura 6 - Gráficos dos valores preditos para períodos de 120 e 180 dias.



Deve-se observar que no período de simulação nos gráficos da Figura 6, existe uma certa dificuldade de simular eventos de cheias, com níveis estáticos inferiores a 3 metros. Esse fato pode estar atribuído a baixa ocorrência dessa escala de valor na etapa de treinamento do modelo utilizado, mas que, no entanto, não influenciou significativamente na maior parte dos valores previstos na série simulada.

Esse fato traz implicações importantes para a utilização deste modelo, pois eventos de cheia são frequentemente utilizados em cálculos de recarga em aquíferos livres, podendo haver erros consideráveis quando usados para estimar recargas hídricas baseadas na previsão destas ocorrências.

Por outro lado, o modelo apresentou maior aderência nas previsões de eventos de descarga, onde ambas previsões apresentaram significativas performances, tanto pela análise visual quanto pela análise dos coeficientes de determinação. Assim, é importante validar o modelo a partir de uma análise dos resíduos, buscando-se melhores observações dos erros residuais cometidos bem como de possíveis *outliers* que são valores atípicos ou muito destoantes dos valores usuais da série de dados. Esse aspecto é importante, sobretudo, em situações onde o modelo calibrado apresenta quantidades frequentes e bem distribuídas de *outliers* que tendem a induzir erros sistemáticos na predição dos valores.

3.3. Validação e análise do erro cometido

Depois de calibrado, analisou-se os erros cometidos na etapa de treinamento por meio da relação entre os valores preditos e os

observados. Assim, apresenta-se o gráfico de resíduos (Figura 7) objetivando analisar a adequabilidade do modelo ajustado para a série de dados.

Ao observar o gráfico da Figura 7, nota-se que o ajuste do modelo ficou satisfatório para a série de dados em estudo, evidenciado pela dispersão uniforme dos pontos em torno da linha de resíduo zero. Assim, devido a distribuição aleatória dos dados, verifica-se a homoscedasticidade da série modelada.

Todavia, ocorre a presença de alguns *outliers* principalmente para níveis estáticos menores que 3,2 metros, que são em sua maioria eventos de cheia, onde a recarga do aquífero é muito rápida.

A ocorrência de valores previstos muito diferente dos valores reais observados em séries temporais de nível de água subterrânea podem advir de diversos fatores, entre eles: erros de medição, erros no equipamento, volumes de precipitação fora do normal, intensidades de precipitação muito altas, valores de nível estático não usados na etapa de treinamento do modelo, etc.

No poço de monitoramento sob estudo, inferiu-se a partir dos resíduos que os dados dos níveis de cheia (valores entre 3,2 a 2,5 metros) influenciaram na calibração do modelo, pois a recorrência destes valores é baixa no universo total de dados estudados. A análise das frequências elaborada e ilustrada na Figura 8, apresenta estatisticamente as ocorrências de faixas de valores observados de nível estático.

Figura 7 - Gráfico dos resíduos do modelo calibrado

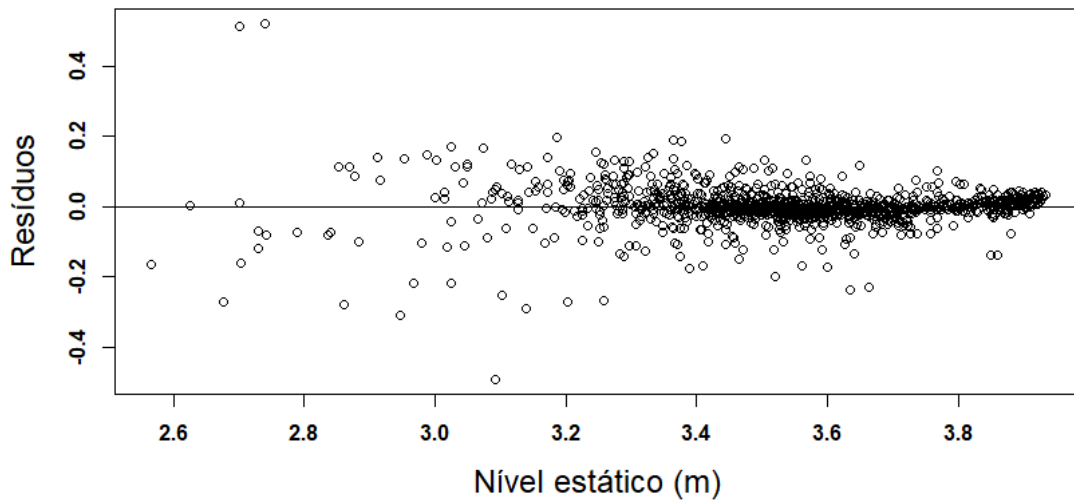
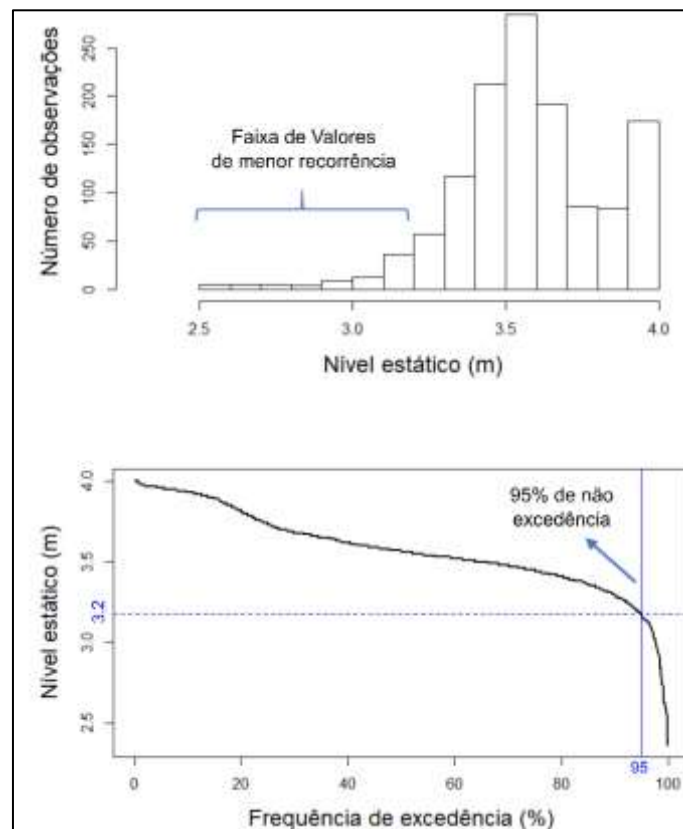


Figura 8 - Histograma de frequência dos dados calibrados e probabilidade de excedência dos níveis estáticos observados



A partir da observação dos gráficos da Figura 8, torna-se evidente a baixa ocorrência dos valores extremos de cheia, principalmente para níveis estáticos menores que 3,2 metros. O histograma de frequências demonstra que os valores entre 3,2 a 2,5 foram pouco recorrentes ao considerar a série utilizada para treinamento do modelo.

A probabilidade empírica de excedência dos valores de nível estático foi calculada (ordenada de forma decrescente e as respectivas probabilidades de excedência, estimadas pela fórmula de Weibull) (WEIBULL, 1939). O gráfico demonstra que o modelo pode ser validado, uma vez que a frequência destes eventos ocorreu em menos que 5% do tempo, ou seja, a probabilidade de excedência destes eventos é baixa frente a totalidade de dados.

Para aproximadamente 95% do tempo monitorado, os níveis estáticos foram superiores a 3,2 metros.

Desta maneira, para as séries temporais sob estudo, é provável que um período de monitoramento mais longo e com mais eventos de cheia possa oferecer resultados ainda melhores de calibragem e, conseqüentemente, de melhor previsão. Assim, a utilização de séries mais homogêneas para calibração tende a oferecer melhores ajustes nas previsões (VAPNIK, 1998).

Tendo em vista que a metodologia de SVM extrai informações da própria série de dados para criar um classificador para fazer as previsões, é possível afirmar que o modelo não captou com exa-

tidão os valores dos níveis estáticos que correspondem aos níveis mais extremos de cheia no aquífero. Contudo, isto não significa que a metodologia não é eficaz para eventos de cheia, mas sim que o período utilizado para calibração não conteve dados suficientes no intervalo entre 3,2 a 2,5 metros para que o modelo ficasse melhor calibrado para esta faixa de valores.

Em contrapartida, os eventos em que os níveis da água do poço estavam baixos foram satisfatoriamente previstos, evidenciado pelo gráfico de resíduos para os valores de níveis estáticos maiores que 3,2 metros. Do ponto de vista prático, essa característica torna o uso do modelo com base em SVM vantajoso, uma vez

que pode ser utilizado para a criação de diferentes cenários que levem em conta a possibilidade de baixa ocorrência de eventos de chuvas. Em outras palavras, o modelo ajustado com a série de dados possui melhor poder de previsão para situações onde se deseja estimar os níveis mais baixos no poço de monitoramento.

Por fim, a eficiência total do modelo, nas diferentes etapas do estudo, foi verificada por meio dos estimadores exemplificados nas equações 2, 3 e 4. A Tabela 2 ilustra os resultados alcançados.

Tabela 2 - Valores dos estimadores nas diferentes etapas da modelagem

ETAPA	R ²	CNS	EQM
Calibração	0,94	0,93	0,0036
Previsão de 120 dias	0,89	0,90	0,0033
Previsão de 180 dias	0,89	0,90	0,0046

Os valores do R² e do CNS, obtidos tanto na etapa de calibração quanto na de verificação do modelo, foram identificados como adequados na medida em que atingem em grande parte a mesma escala de valores de outros autores da área como Shirmohammadi et al. (2012), Shiri et al. (2013) e Yoon et al. (2011), permitindo constatar o sucesso da predição dos dados de nível. Os baixos valores apresentados pelo EQM mostraram que o modelo possui um baixo erro acumulado para a previsão das séries. Na medida em que se aumenta os horizontes de previsão, este erro tende a se acumular.

É importante salientar que as metodologias baseadas em autocorrelação de séries históricas induzem o conhecimento estritamente a partir dos dados utilizados para a calibração dos modelos. Valores com baixa recorrência de monitoramento ou mudanças bruscas nas escalas de valores ocasionadas por alterações climáticas constituem um ponto fraco para o uso deste tipo de abordagem.

Apesar de apresentar algumas limitações, que são inerentes a qualquer modelo teórico, o emprego da técnica de regressão/classificação baseado em SVM evidenciou algumas vantagens ao simular dados de nível estático e precipitação para fazer simulações e preenchimento de falhas em séries temporais de água subterrânea.

A metodologia permite realizar a integração de dados com outros poços ou utilizar mais estações pluviométricas no treinamento do modelo, podendo apresentar boas previsões dos níveis freáticos no contexto das bacias hidrográficas. Períodos de monitoramento mais longos ou a inserção de outras variáveis nas simulações, quando disponíveis, podem trazer aperfeiçoamentos na calibração do modelo e algum efeito aditivo na previsão.

4. CONCLUSÕES

Ao avaliar os resultados alcançados com a metodologia baseada em SVM, tanto na fase de calibração quanto na fase de previsão, observou-se um ótimo ajuste à série de dados, evidenciado pelos coeficientes de determinação que atingiram 94% na calibração e

89% na fase de previsão. Os valores alcançados pelo CNS e pelo EQM permitiram concluir que o modelo calibrado é consistente e apresenta confiabilidade para a previsão da série temporal de nível da água subterrânea, a partir da precipitação e dos níveis anteriores, com boas margens de acerto em 95% do tempo.

As previsões realizadas para os horizontes de 120 e de 180 dias, utilizando um período de monitoramento de 1288 dias para calibração, foram consideradas satisfatórias, demonstrando-se vantajosas para a simulação de dados em preenchimento de falhas ou para a simulações de cenários de recargas ou descargas para poços de monitoramento.

Em suma, a metodologia apresentou uma alternativa para estimar dados de séries temporais em escala diária do nível de água subterrânea. Este tipo de abordagem pode colaborar em análises quantitativas de recarga/descarga em aquíferos livres e na elaboração de estratégias que auxiliem na tomada de decisão e minimizem os impactos potenciais oriundos da exploração excessiva dos recursos hídricos subterrâneos.

REFERÊNCIAS

- AGÊNCIA NACIONAL DAS ÁGUAS - ANA. *Atlas Brasil: Abastecimento urbano de água. Resultados por estado*, v. 2. Brasília: DF, 2010. 92 p. Disponível em: <<http://atlas.ana.gov.br/Atlas/downloads/atlas/Resumo%20Executivo/Atlas%20Brasil%20-%20Volume%20-%20Resultados%20por%20Estado.pdf>> Acesso em: 12 de abr. de 2016.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; DE MORAES, G.; LEONARDO, J., SPAROVEK, G. Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, n.6, p. 711-728. 2013.
- ASEFA, T.; KEMBLAWSKI, M.; MCKEE, M.; KHALIL, A. Multi-time scale stream flow predictions: the support vector machines approach. *Journal of Hydrology*, v. 318, n. 1, p. 7-16. 2006.
- BEHZAD, M.; ASGHARI, K.; EAZI, M.; ALHANG, M. Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with applications*, v. 36, n. 4, p. 7624-7629, 2009.

CHANG, C.C.; LIN, C. J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, v. 2, n. 3, p. 27, 2011.

FEITOSA, E. C.; MANOEL FILHO, J.; COSTA, W. D.; FEITOSA, F. A. C.; DEMETRIO, J. G. A.; FRANÇA, H. D. Avaliação de Recursos Hídricos Subterrâneos. In: FEITOSA, F. A. C. et al. (Org.) *Hidrogeologia: conceitos e aplicações*. 3 ed. Revisada e ampliada. Rio de Janeiro: CPRM, LABHID, 2008. Cap. 7.1, p. 661-669.

HIDROWEB. Sistema de Informações Hidrológicas. Disponível em: <<http://hidroweb.ana.gov.br/>>. Acesso em: 22 de mar. de 2016.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE. *Censo 2010*. Disponível: <<http://censo2010.ibge.gov.br/>> Acesso em: 02 de mai. de 2016.

KHAN, M. S.; COULIBALY, P. Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering*, v. 11, n. 3, p. 199-205, 2006.

LIN, J.; CHENG, C.; CHAU, K. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal*, v. 51, n. 4, p. 599-612, 2006.

MACHADO, J. L. F.; FREITAS, M. A. de. *Projeto Mapa Hidrogeológico do Rio Grande do Sul*: relatório final. Porto Alegre: CPRM, 2005.

MANZIONE, R. L.; KNOTTERS, M.; HEUVELINK, G. B.; VON ASMUTH, J. R.; CAMARA, G. Transfer function-noise modeling and spatial interpolation to evaluate the risk of extreme (shallow) water-table levels in the Brazilian Cerrados. *Hydrogeology Journal*, v. 18, n. 8, p. 1927-1937, 2010.

MANZIONE, R. L.; WENDLAND, E.; TANIKAWA, D. H. Stochastic simulation of time-series models combined with geostatistics to predict water-table scenarios in a Guarani Aquifer System outcrop area, Brazil. *Hydrogeology Journal*, v. 20, n. 7, p. 1239-1249, 2012.

MITCHELL, T. M. *Machine learning*. McGraw Hill, 1997. 432 p.

PYTHON SOFTWARE FOUNDATION. *Python Language Reference, versão 3.4.5*. Disponível em: <<http://www.python.org>> Acesso em 25 de out. de 2016.

RAGHAVENDRA, S. J.; DEKA, P. C. Support vector machine applications in the field of hydrology: a review. *Applied Soft Computing*, v. 19, p. 372-386, 2014.

REDE INTEGRADA DE MONITORAMENTO DE ÁGUAS SUBTERRÂNEAS - RI-MAS. Disponível em: <<http://siagasweb.cprm.gov.br>> Acesso em 20 de jan. de 2016.

SCHÖLKOPF, B.; BURGESS, C.; VAPNIK, V. *Incorporating invariances in support vector learning machines*. In: International Conference on Artificial Neural Networks. Springer Berlin Heidelberg, p. 47-52, 1996.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. *Statistics and computing*, v. 14, n. 3, p. 199-222, 2004.

SHIRI, J.; KISI, O.; YOON, H.; LEE, K. K.; NAZEMI, A. H. Predicting groundwater level fluctuations with meteorological effect implications - A comparative study among soft computing techniques. *Computers & Geosciences*, v. 56, p. 32-44, 2013.

SHIRMOHAMMADI, B.; VAFAKHAH, M.; MOOSAVI, V.; MOGHADDAMNIA, A. Application of several data-driven techniques for predicting groundwater level. *Water Resources Management*, v. 27, n. 2, p. 419-432, 2012.

TAORMINA, R.; CHAU, K.; SETHI, R. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Engineering Applications of Artificial Intelligence*, v. 25, n. 8, p. 1670-1676, 2012.

TRIPATHI, S.; SRINIVAS, V. V.; NANJUNDIAH, R. S. Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of hydrology*, v. 330, n. 3, p. 621-640, 2006.

VAPNIK, V. *The nature of statistical Learning theory*. Springer-Verlag, New York, USA, 1995. 188 p.

VAPNIK, V. N. *Statistical learning theory*. V. 1. New York: Wiley, 1998. 732 p.

WEIBULL, W. *A statistical theory of the strength of materials*. Ingeniors Vetenskaps Akademien, 1939. 45 p.

YOON, H.; JUN, S. C.; HYUN, Y.; BAE, G. O.; LEE, K. K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *Journal of Hydrology*, v. 396, n. 1, p. 128-138. 2011.