

Modelagem de dados por regressão múltipla para previsão do consumo de água em Umuarama

Data modeling by multiple regression for forecasting the water consumption in Umuarama

Vinicius Vianna

Graduando do curso de Engenharia de Produção da Pontifícia Universidade Católica do Paraná (PUCPR). Estagiário na área de Desenvolvimento Operacional da Companhia de Saneamento do Paraná – Sanepar. Curitiba, PR [Brasil]
vinicius_vinna@hotmail.com

Marcelo Dalcul Depexe

Engenheiro Civil pela Universidade Federal de Santa Maria – UFSM, Mestre em Engenharia de Produção pela Universidade Federal de Santa Catarina – UFSC. Atua na área de Desenvolvimento Operacional da Companhia de Saneamento do Paraná – Sanepar. Curitiba, PR [Brasil]
mdepexe@sanepar.com.br

Resumo

Neste trabalho, objetivou-se encontrar uma equação capaz de prever o consumo de água (volume micromedido) da cidade de Umuarama. Utilizando diferentes variáveis, dados obtidos de séries temporais distintas, conhecimentos estatísticos e práticos, dois modelos matemáticos foram propostos, por meio de um aplicativo de regressão múltipla. Com tais modelos, e atentando para as características próprias do estudo da variação do volume consumido de água, foi possível simular previsões e, com os seus resultados, fazer análises comparativas. Assim, analisaram-se a precisão e os erros obtidos. Ao fim do estudo, chegou-se a uma equação capaz de prever o volume consumido de água de Umuarama com erros aceitáveis.

Palavras-chave: Consumo de água. Previsão. Regressão múltipla.

Abstract

This study aimed to find an equation to predict the water consumption of the city of Umuarama. Using a number of variables, data obtained from distinct time series, and statistical and practical knowledge, we proposed two mathematical models, using an application for multiple regression analysis. Using such models and observing the specific procedures for studying variation in the volume of water consumption, we were able to simulate forecasts and make comparative analyses with them. Thus, we analyzed the resulting accuracy and errors. At the end of the study, we found an equation capable of predicting the volume of water consumed in Umuarama with acceptable levels of error.

Key words: Forecast. Multiple regression. Water consumption.

1 Introdução

O processo de previsão de demanda do consumo de água é uma atividade crítica e desafiadora. Segundo Trautwein Junior (2004), o consumo de água varia, até mesmo, ao longo de um dia e está sujeito a problemas operacionais de diversas naturezas, como quedas de energia, acidentes, quebras em componentes do sistema.

Estes vários aspectos sugerem dificuldades na realização de boas previsões e, por isso, levaram diversos pesquisadores a buscarem formas de estimar o consumo de água de uma população e estudar os vários fatores envolvidos nesse tipo de estudo. Conforme apresentado por Arbués et al. (2003), os estudos realizados com estes objetivos possuem dois enfoques, a saber: a determinação dos melhores métodos e a determinação das variáveis que apresentam a melhor correlação com o consumo de água. Ressalta-se que diferentes métodos podem apresentar diferentes resultados para as mesmas análises. Também é importante analisar as variáveis e as séries históricas de maneira adequada.

A influência dos aspectos climáticos, tais como temperaturas máxima e mínima, no consumo de água de uma população, é algo que chama a atenção quando se estuda esse tema. Griffin e Chang (1990) salientam a importância da análise de fatores climáticos, como temperatura e incidência de chuvas, para projetar a demanda de água, além da influência da variação da tarifa dos serviços de água e esgoto.

Zhou et al. (2000) também destaca que outros fatores, além dos climáticos, também podem influenciar ao longo do tempo, como aspectos socioeconômicos relacionados a atividades industriais e de outras instituições.

Destaca-se que existem estudos que se propõem a investigar o comportamento do consumo em países e cidades específicas com resultados,

por vezes, diferentes. Schleich e Hillenbrand (2009) realizam um estudo sobre os fatores determinantes da demanda de água na Alemanha. Além de fatores climáticos, como temperatura, precipitação, os autores avaliaram a influência de aspectos sociais e econômicos, como o preço, a idade média da população, a quantidade de moradores por residência e a renda média. Com relação aos fatores climáticos, os autores salientam a influência das chuvas na variação da demanda de água, mas apontam pequena relação entre temperatura e consumo. Já Slavíková et al. (2013) apresentam conclusão contrária ao analisar os dados de duas cidades da República Tcheca. Os resultados demonstram relação positiva entre temperatura e consumo de água, porém não foi identificada a mesma situação ao analisarem-se os dados de precipitação.

Yasar, Bilgili e Simsek (2012) aplicam um modelo de regressão não linear para previsão da demanda de água na cidade de Adana, na Turquia. As variáveis independentes analisadas são o preço médio mensal da tarifa, o número de clientes da companhia de saneamento (ligações de água), temperatura atmosférica, umidade relativa do ar, precipitação, radiação solar, insolação, velocidade do vento e pressão atmosférica. Ao analisar os dados, os autores concluem que o total de clientes e a temperatura são as variáveis mais importantes para a previsão de demanda da água.

Assim, é possível prever algumas das dificuldades e variações presentes neste estudo e também mirar variáveis que possuem mais probabilidade, em teoria, de correlacionar-se de forma adequada com o consumo.

Resultados diferentes para pesquisas semelhantes revelam o quão específico o padrão de consumo de cada localidade pode ser. Neste artigo, o objetivo será desenvolver um modelo matemático que explique a variação do consumo de água na cidade de Umuarama (Paraná), com base na análise

de dados históricos e na utilização de ferramentas estatísticas. Espera-se determinar uma equação que seja adequada à cidade em questão e apresente estimativas de consumo com erros aceitáveis.

2 Materiais e métodos

2.1 Umuarama – série histórica

A cidade de Umuarama, localizada no Noroeste paranaense, com população urbana de 93.455, conforme Censo Demográfico 2010 (IBGE, 2011), foi o alvo do estudo. Para a realização dos trabalhos, foram utilizados dados mensais técnicos e climáticos do período entre janeiro de 1978 e outubro de 2011. Os dados técnicos foram obtidos por meio do banco de dados da Companhia de Saneamento do Paraná (Sanepar), que possui a concessão para operação do sistema de abastecimento de água do município. Os dados analisados se referem ao número de ligações ativas de água, número de economias residenciais, número de economias totais e volume micromedido (m^3).

Já os dados climáticos utilizados foram fornecidos pelo Instituto Agrônomo do Paraná (IAPAR), com informações sobre temperatura máxima ($^{\circ}C$), temperatura média ($^{\circ}C$), temperatura mínima ($^{\circ}C$), precipitação (mm), umidade relativa do ar (%) e evaporação (mm). Os dados referentes à temperatura são médias mensais das temperaturas máximas e das mínimas, e também uma média geral.

Para estimar o consumo de água na cidade foi utilizado o indicador volume micromedido, que é a somatória do volume registrado em todos os hidrômetros.

2.2 Dados

Os dados foram organizados em planilhas eletrônicas. Os títulos de cada coluna e suas siglas estão listados a seguir.

- Ligações Ativas de Água (Lig);
- Economias Residenciais (EcRes);
- Economias Totais (EcTot);
- Volume Micromedido (VM);
- Temperatura Máxima (Tmax);
- Temperatura Média (Tmed);
- Temperatura Mínima (Tmin);
- Precipitação (Prec);
- Umidade Relativa do Ar (Umid);
- Evaporação (evap).

A tabela com os dados básicos para a realização do estudo apresenta o padrão representado na Tabela 1, na qual constam os dados do ano de 1978.

Esses dados foram trabalhados matematicamente para preencher lacunas existentes no banco de dados. Mediante essas operações também foi possível criar novos indicadores. Um dos que se revelaram mais importantes durante o estudo, foi o Coeficiente Sazonal (CS). Este indicador inclui o efeito da sazonalidade nos modelos e já foi usado em outros estudos com abordagens semelhantes, como, por exemplo, sendo calculado por meio da média móvel (SOUZA; REIS, 2000). Aqui, ele foi calculado pelo quociente entre o volume micromedido do mês pela média do volume micromedido do ano, como visto na Equação 1.

$$CS = \frac{VM_{mês}}{VM_{ano}} \quad (1)$$

2.3 Regressão múltipla

As análises estatísticas foram realizadas com o auxílio de um aplicativo de Regressão Múltipla denominado BER – Análise de Regressão (SAMOHYL, 2005). Desenvolvido por Robert Wayne Samohyl, Éder Daniel Corvalão e Rubson Rocha, este aplicativo foi fundamental no estudo

Tabela 1: dados do ano de 1978 utilizados no estudo

1978/ Mês	dias	Lig água	Econ res água	Econ total água	medido	Temp max	Temp med	Temp min	Precipit (mm)	Umidade relativa (%)	Evaporação Piche (mm)
		Lig	EcRes	EcTot	VM	Tmax	Tmed	Tmin	Prec	Umid	evap
Jan.	31	4563	5077	6187	121955	31,89	26,15	21,55	92,3	64,5	5,26
Fev.	28	4663	5181	6298	137869	33,43	26,63	21,33	102	55,14	7,16
Mar.	31	4785	5314	6455	124590	31,15	25,55	21,12	131,9	64,97	5,32
Abr.	30	4861	5401	6549	125659	29,11	22,08	16,5	78	47,73	6,37
Mai.	31	4902	5455	6604	128507	25,61	18,91	13,84	50,9	51,02	5,84
Jun.	30	5073	5665	6819	120016	24,86	18,6	14,23	42,2	57,97	5,32
Jul.	31	5267	5908	7071	101791	26,16	20,15	15,69	184,1	63,54	4,55
Ago.	31	5302	5951	7111	107914	23,95	17,76	13,04	58,2	59,41	8,01
Set.	30	5422	6076	7245	94227	25,53	20,52	16,24	158,9	65,92	4,72
Out.	31	5596	6302	7473	122337	30,12	23,96	18,92	96,1	56,07	7,04
Nov.	30	5605	6383	7561	149414	29,24	23,91	19,52	184,2	64,47	4,9
Dez.	31	6153	7079	8301	148737	30,77	25,25	20,73	170,1	63,07	7,42

dos dados, apresentando diversos indicadores e até interpretações para os valores obtidos.

Entre os critérios, métodos e conceitos estatísticos oferecidos pelo aplicativo e utilizados no estudo, alguns dos mais importantes para o desenvolvimento do trabalho foram relacionados e descritos brevemente a seguir.

- R^2

O coeficiente de determinação (R^2) mede a redução proporcionada na variação total na variável predita com o uso do conjunto de variáveis preditoras (SAMOHYL, 2005). Este indicador mede a capacidade do modelo de estimar os valores de Y . Variando de 0 até 1, quanto maior seu valor, mais confiável o modelo.

- Estatística Durbin-Watson

Testa a hipótese que não há autocorrelação defasada presente nos resíduos. O valor computado do teste Durbin-Watson é comparado com valores correspondentes da Tabela F. Interpretação: se não há autocorrelação, a distribuição é simétrica perto do 2, seu valor médio (SAMOHYL, 2005). Os valores variam de 0 a 4. Quando o valor

é muito baixo, indica correlação positiva entre os resíduos, já quando é muito alto, mostra correlação negativa.

- Box-Pierce e Ljung-Box

Indicadores que apontam a possibilidade dos resíduos não serem autocorrelacionados. São capazes de mostrar se a autocorrelação amostral é significativa.

- Teste RESET

Do inglês: RESET – *Regression Specification Error Test*. Teste de especificação do modelo de regressão (SAMOHYL, 2005). É um teste de diagnóstico que indica problemas na modelagem (RAO; TILT; LESTER, 2012).

- Teste Breush-Pagan

Para variância constante do resíduo do modelo, assume que os resíduos são independentes e distribuídos normalmente (SAMOHYL, 2005).

- Teste Jarque-Bera

Testa a normalidade da distribuição dos dados. O valor desse índice deve ser menor que 5,99

para apontar normalidade dos dados, com $\alpha = 0,05$ (JARQUE; BERA, 1987). Esse teste só é válido para séries estacionárias.

- Estacionariedade

De acordo com Andrade (2004), o processo será francamente estacionário se o valor esperado em qualquer momento não depender do tempo, a variância dos seus valores, seja qual for o período, for constante e finita e a covariância entre diferentes observações depender do intervalo entre essas observações, mas não do período.

Dados que possuem tendência de crescimento ou diminuição ao longo do tempo são ditos não estacionários. Ressalta-se a importância dessa característica, pois muitos dos testes estatísticos foram formulados pressupondo a estacionariedade das séries.

3 Análise dos dados

3.1 Análise da evolução histórica

Com as séries históricas devidamente inseridas numa planilha eletrônica, foi possível começar a estudá-las mais detalhadamente. Para possibilitar uma visão geral do desenvolvimento e comportamento das séries, algumas análises gráficas foram realizadas. Elas foram muito úteis, pois permitiram uma avaliação geral e inicial de forma simplificada.

Primeiramente, verificou-se um gráfico de linhas comparando as evoluções do número de ligações, das economias residenciais e das economias totais, conforme o Gráfico 1. Esse gráfico comprovou a tese de que esses indicadores crescem numa taxa semelhante. É possível notar a tendência de crescimento, ou seja, a não estacionariedade inerente ao estudo do processo de consumo de água.

Também é possível identificar comportamentos não esperados em alguns períodos de tempo. Deve-se lembrar de que muitos desses dados, prin-

cipalmente os mais antigos, foram catalogados manualmente, o que aumenta a suspeita sobre alguns possíveis dados errôneos.

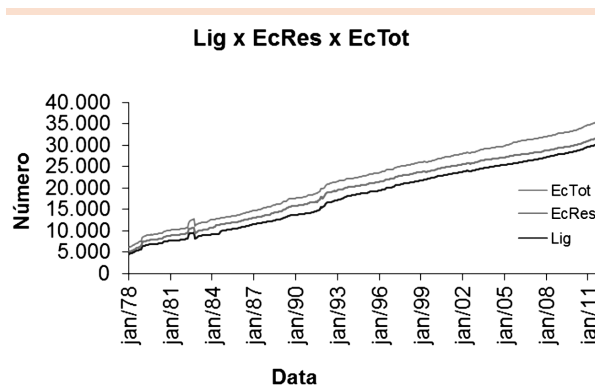


Gráfico 1: Evolução do número de ligações, economias residenciais e economias totais

O volume micromedido, indicador-chave no estudo, também foi avaliado de forma a entender seu comportamento. O Gráfico 2 apresenta a evolução deste volume entre janeiro de 1978 e outubro de 2011.

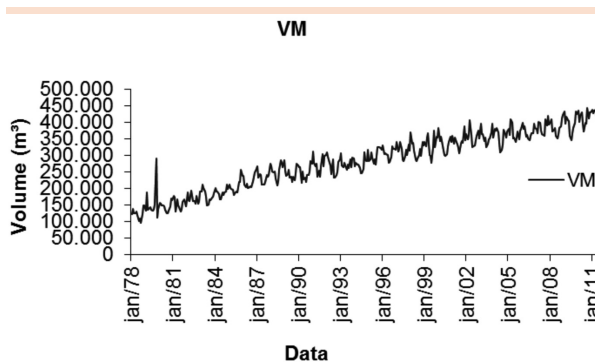


Gráfico 2: Evolução histórica do volume micromedido

Nota-se que o volume micromedido de água em Umuarama cresceu ao longo do tempo. Essa análise visual já é suficiente para afirmar que esta série de dados não é estacionária. Outra possível inferência diz respeito ao comportamento sazonal desse consumo hídrico. Isso se deve às oscilações da demanda, em função de mudanças de hábitos de consumo ao longo do ano. Verifica-se que nos

meses mais quentes há tendência de aumento dos volumes, enquanto nos meses de inverno ocorre redução do consumo, o que indica que pode haver influência de fatores climáticos.

Segundo Falkenberg (2005), o consumo aumenta conforme aumenta a temperatura. Para testar se essa afirmação se aplica a este estudo, inicialmente observaram-se as séries históricas das temperaturas máxima, média e mínima, entre 1995 e 2010.

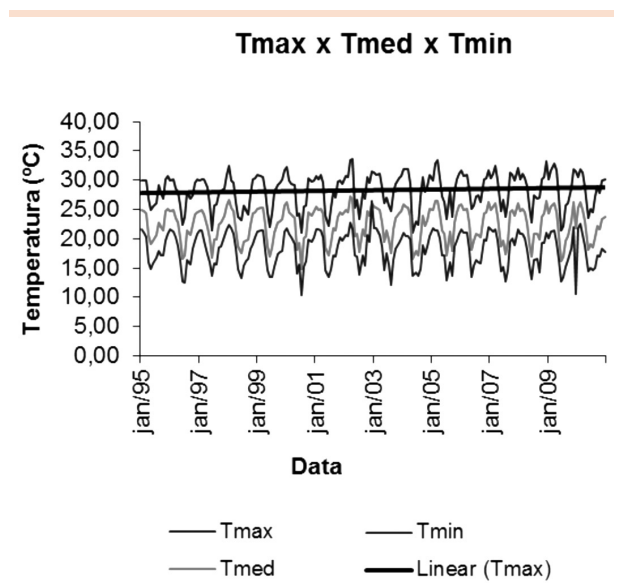


Gráfico 3: Comparação do desenvolvimento das temperaturas máxima, média e mínima, entre 1995 e 2009

O Gráfico 3 revela que as temperaturas variam de modo semelhante ao longo do ano. Outra observação importante vem junto com a linha de tendência da temperatura máxima, que mostra uma tendência de crescimento muito suave. Com o objetivo de unir as duas análises, foi elaborado um gráfico para comparação da variação da temperatura máxima média mensal no período entre janeiro de 2000 e janeiro de 2010 e a variação do volume micromedido. O período de tempo para essa análise é menor para facilitar a análise visual das séries. A escolha da temperatura máxima se deu

devido a sua maior influência na variação do consumo. O Gráfico 4 possui dois eixos Y, um para o volume micromedido, e outro, para a temperatura máxima, apresentando os seguintes dados.

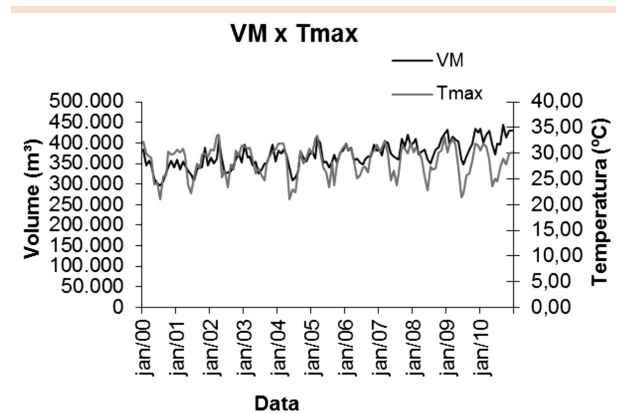


Gráfico 4: Comparação entre o volume micromedido e a temperatura máxima entre 2000 e 2010

A partir desse gráfico é possível confirmar a tese de que o consumo de água varia com a temperatura, observando-se um consumo maior nos meses de verão. Ressalta-se que isso ocorre mesmo tendo a série do volume micromedido uma tendência positiva de crescimento, enquanto a de temperatura máxima possui maiores indícios de estacionariedade. Ao analisar os dados, percebe-se que fevereiro apresenta valores inferiores ao esperado para um mês de calor. Isso ocorre devido a uma característica especial desse mês – seu número reduzido de dias. Esse fato, numa análise em que se consideram dados mensais, também exerce influência, pois há diferença no número de dias entre as leituras de consumo realizadas entre meses consecutivos.

Essas análises preliminares permitiram o avanço dos estudos. Assim, foi possível iniciar as primeiras tentativas de regressão com o auxílio do aplicativo BER.

3.2 Modelagem dos dados

Nas análises iniciais, utilizaram-se dados do período entre janeiro de 1978 e dezembro de

2010. Neste estudo, definiu-se o volume micromedido como a variável dependente Y, pois o objetivo central é prever o consumo indicado por ela.

Para a primeira modelagem, buscou-se usar variáveis que, teoricamente, possuíssem grande influência na estimação do consumo. A temperatura máxima foi escolhida por ser um fator externo; e o número de ligações, por ser um fator interno e controlado. Por meio do aplicativo BER, utilizaram-se essas variáveis e optou-se pela inclusão de uma constante na equação. A seguir, a Equação 2 representa o modelo proposto pelo aplicativo; e os resultados dos testes estatísticos são mostrados na Tabela 2.

O Gráfico 5 é um comparativo entre o volume micromedido estimado e o volume micromedido real pelo modelo proposto, que possibilita uma análise visual dos dados.

A partir desses dados, é possível observar algumas condições próprias desse tipo de estudo. Ao analisar os resultados numéricos, encontram-se vários índices que indicariam a inviabilidade do modelo. O alto valor do teste Jarque-Bera aponta a não normalidade; o teste Durbin-Watson, a correlação negativa entre os resíduos e os valores de

Q atestam que os resíduos não são autocorrelacionados. Porém, observa-se uma grande semelhança no desenvolvimento das curvas no Gráfico 5, o que leva a acreditar na validade do modelo.

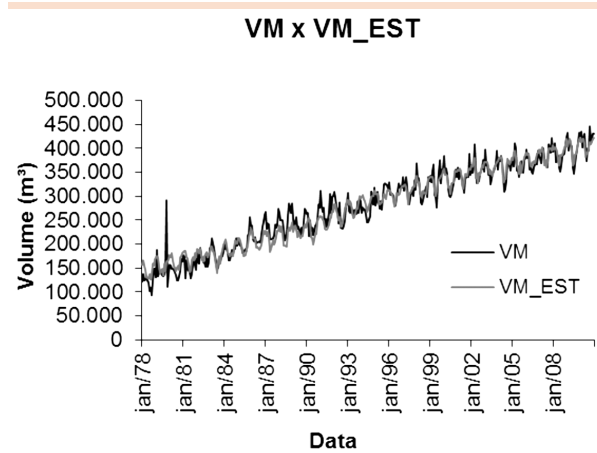


Gráfico 5: Comparação entre os valores reais do volume micromedido e das previsões realizadas pelo Modelo 1, entre 1978 e 2008

Um aspecto importante e indispensável no entendimento dos resultados é a não estacionariedade do volume micromedido. Ressalta-se, novamente, que muitos dos testes estatísticos partem da premissa da estacionariedade. Portanto, este fator deve ser considerado ao analisar os resultados obtidos.

$$VM = -50762,5954425333 + 4950,23413722687 \times Tmax + 10,967624933245 \times Lig$$

(2)

Tabela 2: Resultados das análises estatísticas do primeiro modelo proposto

Fonte	GL	SQ	QM	F	Valor de p		
Regressão	2	2.701.821.456.173,56	1.350.910.728.086,78	3.464,84	0,0000		
Resíduo	393	153.227.417.669,43	389.891.648,01				
TOTAL	395	2.855.048.873.843,00					
R2 =	0,9463			s(resíduo)	19745,6742	DW resíduo	1,3472
R2ajust. =	0,9461			EAM	15302,1810	Q	623,4762
SIC=	8,6073			EPAM	0,0643	Q*	643,3785
AIC=	8,6028			PRESS	155.575.146.581,95	RESET	28,5883
HQ=	8,5939			PRESS-AP	0,0650	BP=	29,6738
				U de Theil	0,8051	JB(resíduo)	282,81

Para avaliar de maneira mais profunda os resultados do modelo, realizou-se uma análise de seus erros. Inicialmente, foi calculada a diferença entre o volume micromedido estimado e o micromedido real para, em seguida, dividir essa diferença pelo volume micromedido real. Essa forma de avaliar os erros é denominada *Mean Absolute Percentage Error* (MAPE). O primeiro passo é calcular esse erro percentual para cada uma das previsões e, após isso, calcular a média aritmética de todos os erros. A fórmula utilizada aqui é mostrada na Equação 3.

$$MAPE = \frac{VM_{estimado} - VM_{real}}{VM_{real}} \times 100 \quad (3)$$

Assim, foi obtida uma sequência que indica a variação do erro ao longo da série. O Gráfico 6 mostra o comportamento do erro percentual das previsões feitas de 1978 até 2010.

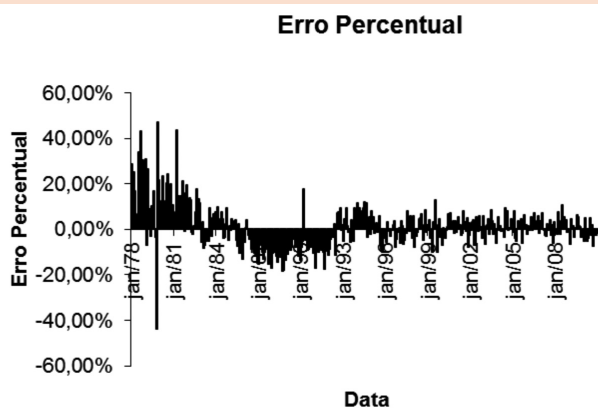


Gráfico 6: Erros percentuais das previsões do Modelo 1 entre 1978 e 2008

Analisando os resultados do cálculo do MAPE, observa-se que quase metade das previsões (47,2754%) estão distantes pelo menos cinco pontos percentuais para mais ou para menos da meta de erro igual a zero. Também é possível verificar que, apesar da média apresentar um valor

interessante (1,0088%), em algumas previsões o erro foi quase de 50%.

Para testar o modelo numa situação mais próxima da real, a equação foi utilizada de modo a prever o consumo no ano de 2011. Apesar da disponibilidade de alguns dados desse ano, eles não foram usados na modelagem. Assim, foi possível simular o comportamento do modelo numa situação de previsão normal. Para isso, foram utilizadas as temperaturas máximas do ano anterior e, por ser um dado mais facilmente previsto, as ligações de 2011. Os testes anteriores foram refeitos, os valores do volume micromedido real e do volume micromedido estimado foram comparados (como visto no Gráfico 7), e a análise da variação do erro percentual foi representada no Gráfico 8.

Observam-se erros aceitáveis e uma tendência do modelo em subestimar os volumes. Destacam-se o valor percentual médio dos erros (-2,12%) e os dois valores extremos (1,73% e -5,61%).

Ao analisar o Gráfico 6, em que o MAPE foi calculado desde 1978, é observado um comportamento variável da curva. Nos primeiros anos, existem mais e maiores erros. Contudo, esses erros começam a reduzir e variar mais próximos de zero num passado mais recente. Essa inferência possibilitou buscar-se um modelo baseado em dados obtidos nos últimos anos. Analisando os resultados, identificou-se um comportamento mais adequado a partir de 1996. Por isso, a próxima etapa do estudo contou com um banco de dados menor. Os dados utilizados são do período entre janeiro de 1996 e dezembro de 2010.

Com o objetivo de aumentar a precisão da previsão nessa nova modelagem, foi incluído também o coeficiente sazonal de cada mês do ano anterior. Assim, a Equação 4 foi encontrada, e os resultados das análises estatísticas do modelo são mostrados na Tabela 3.

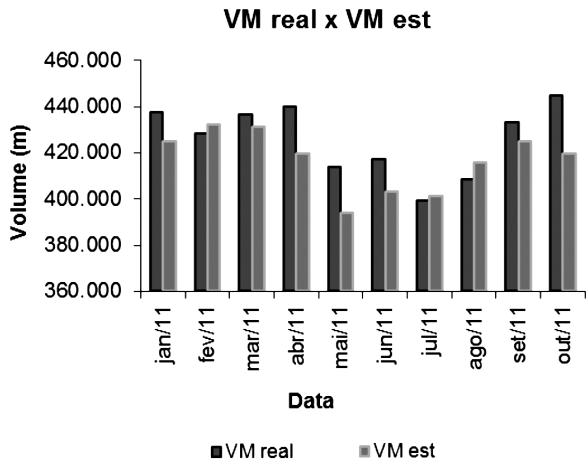


Gráfico 7: Comparação entre o volume real e o micromedido para as previsões do Modelo 1 no ano de 2011

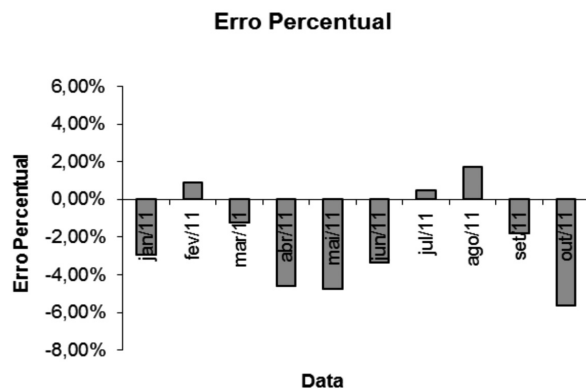


Gráfico 8: Análise dos erros para das previsões do Modelo 1 para 2011

Dois valores chamam muita atenção nesses resultados: o valor muito alto do R² (0,9824), e o valor muito baixo do teste Durbin-Watson (0,7956). Esses dois fatores, quando combinados, são um dos grandes indicadores de regressões espúrias. Segundo Chiarella e Gao (2002), uma regressão espúria é a que tende a aceitar uma relação falsa, ou rejeitar uma verdadeira, por causa de sistemas falhos de regressão. Comum em séries não estacionárias, ela pode ocorrer devido a alguns fatores, como inclusão de variáveis insignificantes para o estudo ou omissão de variáveis importantes.

Especificamente neste trabalho, encontram-se subsídios teóricos e conhecimentos provenientes da prática para aceitar como verdadeira a relação entre ligações, temperatura máxima e volume micromedido. Já a variável coeficiente sazonal introduziu a sazonalidade na modelagem.

Existem recursos para transformar os dados de modo a adicionar estacionariedade nas séries. Um deles é a diferenciação dos dados até remover-se a tendência. Entretanto, esse método excluiria o crescimento da curva do volume micromedido, o que acabaria por invalidar a previsão aqui proposta.

$$VM = 147940,905267986 + 6,15228892112577 \times 10^{-9} \times Lig^3 + 9,51963773512075 \times 10^{-3} \times Tmax^4 + 106621,460965388 \times CS^3$$

(4)

Tabela 3: Resultados das análises estatísticas do segundo modelo proposto

Fonte	GL	SQ	QM	F	Valor de p
Regressão	3	250.923.227.494,12	83.641.075.831,37	3.277,90	0,0000
Resíduo	176	4.490.932.049,52	25.516.659,37		
TOTAL	179	255.414.159.543,64			
R2 =	0,9824			s(resíduo)	5051,4017
R2ajust. =	0,9821			EAM	4036,2631
SIC=	7,4472			EPAM	0,0113
AIC=	7,4415			PRESS	4.720.590.931,29
HQ=	7,4128			PRESS-AP	0,0116
				U de Theil	0,2137

DW resíduo	0,7956
Q	203,4083
Q*	210,8722
RESET	0,5536
BP=	2,4782
JB(resíduo)	0,30

Considerando essas implicações teóricas, testes com maior característica prática foram realizados. Assim como no primeiro modelo, testes gráficos foram feitos. Primeiramente, um comparativo entre os o volume micromedido real e o volume micromedido estimado entre os anos de 1996 e 2010, como mostrado no Gráfico 9.

A análise gráfica apresenta um resultado ótimo, com a curva do VM_EST, por muitas vezes, se sobrepondo sobre a curva do VM. Isso indica grande precisão na previsão. Dando sequência às análises, verifica-se o comportamento do MAPE ao longo da série no Gráfico 10.

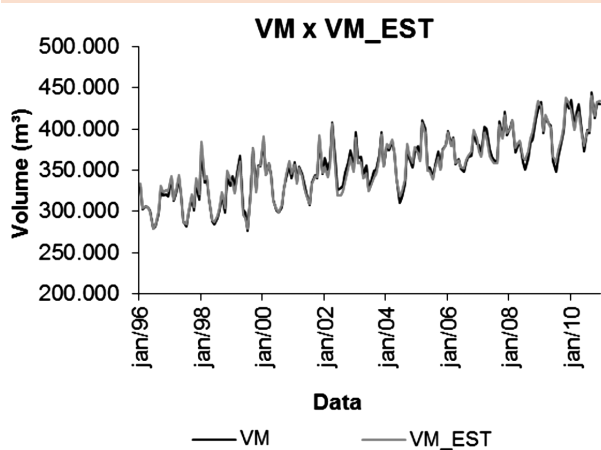


Gráfico 9: Comparação entre os valores reais do volume micromedido e das previsões realizadas pelo Modelo 1 entre 1996 e 2010

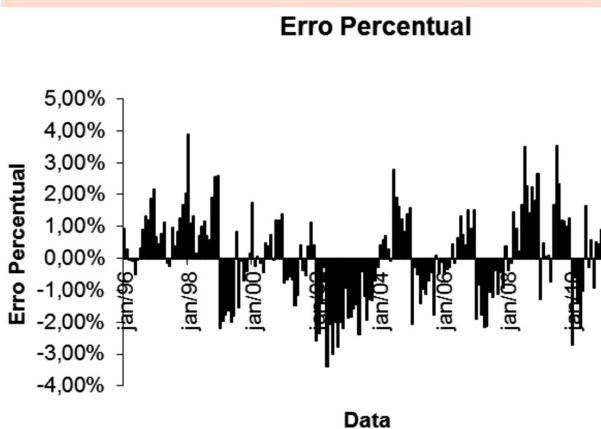


Gráfico 10: Erros percentuais das previsões do Modelo 2 entre 1996 e 2010

Confirmando as conclusões da análise gráfica, o estudo dos erros revela grande ganho de precisão no modelo após a mudança do período estudado e a inclusão do coeficiente sazonal. Com mais de 85% dos dados concentrados na região de menor erro e nenhuma previsão apresentando erro de mais de cinco pontos percentuais, observou-se melhora significativa na modelagem. A média dos erros ficou próxima a zero (0,0206%), e os valores máximo e mínimo não ultrapassaram quatro pontos percentuais para mais e para menos.

Com o intuito de simular o modelo com dados não utilizados em sua modelagem, o teste de previsão para o ano de 2011 foi repetido.

Num primeiro momento, a análise do Gráfico 11 parece revelar ganho de precisão. Nos meses iniciais, as curvas apresentam comportamento semelhante, ocorrendo alguns erros maiores no fim da série. Um gráfico com o desenvolvimento do erro percentual foi montado e está representado no Gráfico 12.

O ganho de precisão na previsão dos dados de 2011 foi mais modesto se comparado com o de precisão na série completa. Destaca-se que a curva percentual do erro aproximou mais do zero neste novo modelo, assim como a média (-0,1%); entretanto, a variação foi maior (entre -5,0% e 5,9%).

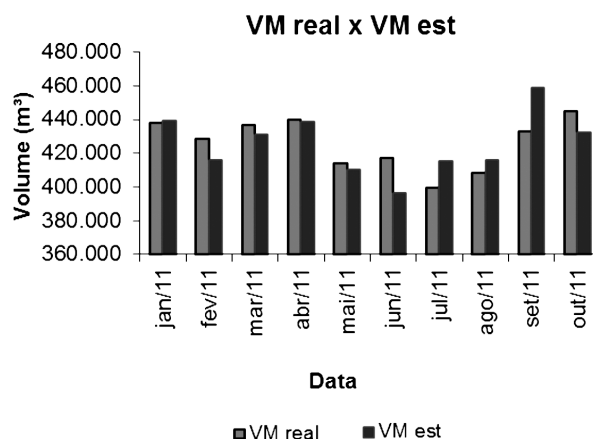


Gráfico 11: Comparação entre o volume real e o micromedido para as previsões do Modelo 2 no ano de 2011

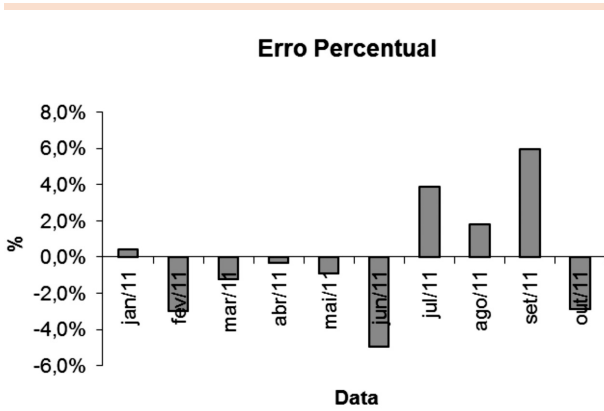


Gráfico 12: Análise dos erros para das previsões do Modelo 2 para 2011

Com os resultados dos dois modelos, foi possível fazer uma comparação entre eles. Para isso, um gráfico comparando o desenvolvimento real do volume micromedido, das estimativas propostas pelo primeiro modelo e das estimativas propostas pelo segundo modelo foi montado. O resultado é mostrado no Gráfico 13.

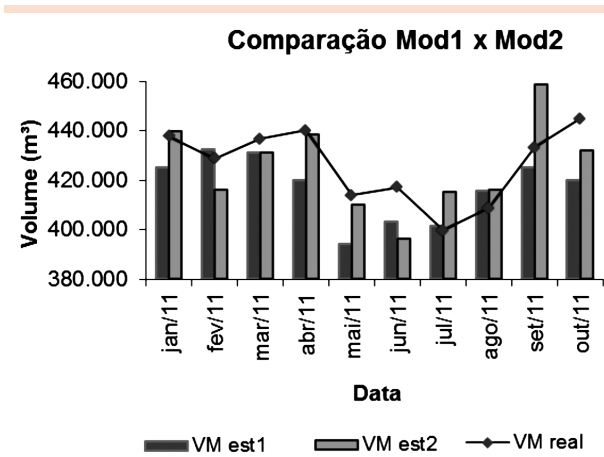


Gráfico 13: Comparação entra resultados do Modelo 1, do Modelo 2 e o volume micromedido real

É possível visualizar que ambas as estimativas apresentaram maiores e menores níveis de acerto ao longo da série. Aprofundando a análise dos resultados, a Tabela 4 apresenta os erros percentuais das previsões de 2011 em ordem crescente e lado a lado. Além disso, também é possível

comparar as médias aritméticas simples, médias aritméticas sem os valores extremos e os valores máximos e mínimos de cada série.

Tabela 4: Primeira análise comparativa entre os resultados dos modelos

Mod1 (%)	Mod2 (%)
-5,61	-4,97
-4,77	-2,95
-4,59	-2,86
-3,34	-1,23
-2,93	-0,93
-1,82	-0,32
-1,23	0,41
0,46	1,82
0,90	3,87
1,73	5,93
Média	
-2,12	-0,12
Média sem extremos	
-2,16	-0,27
Máximo	
1,73	5,93
Mínimo	
-5,61	-4,97

A partir dessa tabela é possível afirmar que, em média, o Modelo 2 é mais preciso. Essa informação é confirmada pelo cálculo da média sem extremos, que evita possíveis compensações entre os valores positivos e negativos, o que mascararia o resultado.

A fim de comparar a variação dos erros de mais uma forma, de modo a garantir uma análise correta, a tabela anterior é rerepresentada na Tabela 5, considerando, porém, somente o módulo dos erros percentuais.

Esta nova tabela dá ainda mais subsídios para confirmar a superioridade da previsão da segunda modelagem. Com médias e com o valor mínimo mais interessantes, esse modelo revela ter um bom nível de precisão e resultados compatíveis com os objetivos do estudo.

Tabela 5: Primeira análise comparativa entre os resultados dos modelos

Mod1 (%)	Mod2 (%)
0,46	0,32
0,90	0,41
1,23	0,93
1,73	1,23
1,82	1,82
2,93	2,86
3,34	2,95
4,59	3,87
4,77	4,97
5,61	5,93
Média	
2,74	2,53
Média sem extremos	
2,66	2,38
Máximo	
5,61	5,93
Mínimo	
0,46	0,32

3.3 Análise de sensibilidade

Com o Modelo 2 em mãos, foi possível realizar análise de sensibilidade a fim de descobrir como variações na temperatura máxima, nas ligações e no coeficiente sazonal afetam a previsão. Para tanto, foram realizadas alterações percentuais para mais e para menos em cada uma das variáveis, uma a uma. Esta mesma análise foi utilizada de forma semelhante por Wyatt (2010).

Cada variável passou por alterações de $\pm 5\%$, $\pm 10\%$ e $\pm 25\%$, enquanto as outras eram mantidas constantes. Estas variações foram determinadas com o intuito de testar a influência das variações de diferentes grandezas. Com essa variação feita, a variável era reintroduzida na equação e novas previsões para o volume micromedido eram realizadas.

Os resultados indicaram que a equação é mais sensível à variação positiva no caso do coeficiente sazonal e da temperatura máxima, enquanto a

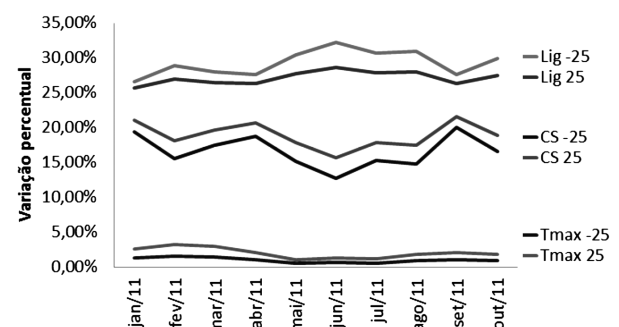
variação negativa das ligações é mais influente. Também é possível inferir que, em sequência, as variações que mais afetaram os resultados foram as ligações, o coeficiente sazonal e, por fim, a temperatura máxima. O coeficiente sazonal apresentou maior diferença média na variação, quando comparando a mudança para mais e para menos. A Tabela 6 indica o módulo da variação percentual em 25% para mais e para menos em cada uma das variáveis. O período adotado foi entre janeiro e outubro de 2011.

Os dados da tabela anterior serviram de base para o Gráfico 14. Nele é possível comparar a influência da variação nas previsões ao longo do tempo.

Tabela 6: Resultados e comparação dos testes da análise de sensibilidade

Data 2011	Variação percentual em módulo (%)					
	Lig -25	Lig 25	CS -25	CS 25	Tmax -25	Tmax 25
Jan	26,67	25,77	19,37	21,11	1,29	2,62
Feb	28,91	27,00	15,57	18,17	1,62	3,26
Mar	27,98	26,49	17,50	19,71	1,47	2,97
Abr	27,67	26,32	18,73	20,64	1,04	2,12
Mai	30,47	27,80	15,21	17,87	0,49	1,01
Jun	32,20	28,65	12,78	15,74	0,66	1,36
Jul	30,67	27,90	15,28	17,93	0,57	1,19
Ago	30,92	28,03	14,82	17,55	0,87	1,79
Set	27,63	26,30	20,03	21,57	1,01	2,06
Out	29,99	27,56	16,54	18,96	0,90	1,85

Variação percentual em módulo

**Gráfico 14: Comparação entre os resultados das variações**

Ressalta-se que as ligações cresce de maneira quase linear e dificilmente se reduz e crescerá até uma saturação do sistema. Já o coeficiente sazonal varia de acordo com o volume micromedido. A análise da temperatura máxima é a mais significativa, pois essa é a variável que, na prática, possui maior tendência de variabilidade negativa e positiva.

4 Conclusão

O estudo revelou particularidades do comportamento do consumo de água. A busca por resultados incluiu utilização de diversas variáveis, séries históricas de tamanhos diferentes e análises gráficas de resultados e erros. Nesse ponto dificuldades foram encontradas. Uma delas foi a quantidade de variáveis que, teoricamente, podem influenciar no processo de previsão de demanda e na escolha das variáveis mais adequadas. Ressalta-se que, na prática, algumas variáveis não contribuíram de forma efetiva com o aumento da precisão dos modelos da pesquisa. Outra grande dificuldade foi lidar com a não estacionariedade do volume micromedido, uma característica estatística inerente ao estudo da demanda de água, o que dificultou a análise dos resultados dos testes estatísticos e exigiu outras formas de análise.

Desse modo, os conhecimentos práticos foram fundamentais. Reconhecer como verdadeira a relação entre o volume micromedido, ligações e temperatura possibilitaram a aceitação de modelos. Outro aspecto importante foi considerar a sazonalidade do consumo. Os resultados demonstram a influência das temperaturas máximas no consumo de água. Os modelos elaborados apresentaram erros baixos, quando comparados aos dados reais, sugerindo possíveis utilizações práticas.

Os resultados também indicam que os modelos elaborados com base em dados mais recentes apresentam melhores resultados. Portanto, dados muito antigos podem não representar a realidade atual. Isso implica na necessidade de atualização dos modelos, pois pode ocorrer uma perda de precisão em função de alterações na população, nos hábitos de consumo e em outras características locais que não estão sob controle.

A análise de sensibilidade revelou a influência de cada uma das variáveis sobre o modelo. Foi possível ponderar e observar o efeito de variações nos resultados das previsões, assim compreendendo melhor os fatores-chave que determinam a qualidade do modelo. Para futuros estudos, sugere-se o uso de outros métodos, como a suavização exponencial.

Referências

- ANDRADE, J. S. *Apontamentos de econometria aplicada*. 2004. Disponível em: <<http://www4.fe.uc.pt/mwg-internal/de5fs23hu73ds/progress?id=08SmPoztaX>>. Acesso em: 12 mar. 2013.
- ARBUÉS, F.; GARCÍA-VALIÑAS, M. Á.; MARTÍNEZ-ESPIÑEIRA, R. Estimation of residential water demand: a state-of-the-art review. *Journal of Socio-Economics*, v. 32, n. 1, p. 81-102, 2003.
- CHIARELLA, C.; GAO, S. *Type I spurious regression in econometrics*. Working Paper 114. School of Finance and Economics, University of Technology, Sydney. abr., 2002. Disponível em: <www.business.uts.edu.au/finance>. Acesso em: 12 mar. 2013.
- FALKENBERG, A. V. *Previsão de consumo urbano de água em curto prazo*. 2005. 89 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia)–Universidade Federal do Paraná, Curitiba, 2005.
- GRIFFIN, R.C.; CHANG, C. Pretest analyses of water demand in thirty communities. *Water Resources Research*, v. 26, n. 10, p. 2251-2255, 1990.
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Sinopse do Censo Demográfico 2010*. Rio de Janeiro: IBGE, 2011. 261 p.
- JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. *International Statistical Review*, v. 55, n. 2, p. 163-172, 1987.



RAO, K. K.; TILT, C. A.; LESTER, L. H. Corporate governance and environmental reporting: an Australian study. *Corporate Governance*, v. 12, n. 2, p. 143-163, 2012.

SAMOHYL, R. W.; ROCHA, R.; CORVALAO, E. BER – Análise de Regressão. Software: Regressão Múltipla BER, 2005. Disponível em: <<http://qualimetria.wordpress.com/2012/12/09/previsao/>>. Acesso em: 12 mar. 2013.

SCHLEICH, J.; HILLENBRAND, T. Determinants of residential water demand in Germany. *Ecological Economics*, v. 68, n. 6, p. 1756-1769, 2009.

SLAVÍKOVÁ, L. et al. Impacts of climate variables on residential water consumption in the Czech Republic. *Water Resources Management*, v. 27, n. 2, p. 365-379, 2013.

SOUZA, J. M. G. de; REIS, J. N. P. A evolução dos custos da cesta básica na cidade de Fortaleza no período de 1985 a 1997. *Revista Econômica do Nordeste*, Fortaleza, v. 31, n. 1 p. 66-82, jan.-mar. 2000.

TRAUTWEIN JUNIOR, B. Avaliação de métodos para previsão de consumo de água para curtíssimo prazo: um estudo de caso em empresa de saneamento. 2004. 107 p. Dissertação (Mestrado em Engenharia de Produção e Sistemas)–Pontifícia Universidade Católica do Paraná, Curitiba, 2004.

WYATT, A. S. Non-Revenue Water: financial model for optimal management in developing countries. RTI Press publication n°. MR-0018-1006. *Research Triangle Park*. Jun., 2010.

YASAR, A.; BILGILI, M.; SIMSEK, E. Water demand forecasting based on stepwise multiple nonlinear regression analysis. *Arabian Journal for Science and Engineering*, v. 37, n. 8, p. 2333-2341, 2012.

ZHOU, S. L. et al. Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*, v. 236, n. 3, p.153-164, 2000.

Recebido em 14 mar. 2012 / aprovado em 19 abr. 2013

Para referenciar este texto

VIANNA, V.; DEPEXE, M. D. Modelagem de dados por regressão múltipla para previsão do consumo de água em Umuarama. *Exacta – EP*, São Paulo, v. 11, n. 1, p. 33-46, 2013.