

## MODELAGEM DO ATRASO ZENITAL TROPOSFÉRICO A PARTIR DE REGRESSÃO LINEAR MÚLTIPLA

HAROLDO ANTONIO MARQUES  
LUIZ FERNANDO SAPUCCI  
JOÃO FRANCISCO GALERA MONICO

Universidade Estadual Paulista - Unesp  
Faculdade de Ciências e Tecnologia - FCT  
Departamento de Cartografia, Presidente Prudente - SP  
{haroldoh2o@gmail.com, galera@fct.unesp.br}

Instituto Nacional de Pesquisas Espaciais,  
Centro de Previsão de Tempo e Estudos Climáticos  
Divisão de Desenvolvimento e Modelagem, Cachoeira Paulista – SP  
{sapucci@cptec.inpe.br}

---

**ABSTRACT** - GPS satellite observations are subject to several error sources as those involved in the signal propagation through the terrestrial atmosphere. For geodetic positioning the atmosphere is usually divided into ionosphere and troposphere. While the ionosphere effects are minimized by a linear combination of the observables from dual frequency receivers, the main effect of the troposphere, i.e. Zenital Tropospheric Delay - ZTD, can be minimized by empirical models. The two most used models are the Hopfield and the Saastamoinen, which was developed based mainly on data collected in the North Hemisphere. In this paper it will be presented a multiple regression analyze of meteorological variables measured on terrestrial surface aiming to determine the best variables for the ZTD determination in Brazil. For this analyze the reference ZTD values were calculated using data from radiosondes launched in Brazil. An initial analyze was carried out for the radiosondes launched in Curitiba city, and the results showed that the methodology is adequate for investigating about the contribution of each meteorological variable in the ZTD modeling.

---

### 1 INTRODUÇÃO

As observações coletadas a partir dos satélites GPS estão sujeitas as mais diversas fontes de erros, dentre as quais se destacam os erros envolvidos com a propagação do sinal através da atmosfera terrestre. Para os geodestas, a atmosfera é composta pela troposfera e pela ionosfera. A ionosfera é uma camada da atmosfera compreendida entre aproximadamente 50 e 1500 km acima da Terra e é caracterizada pela presença de elétrons livres (LEICK, 1995). A troposfera é a camada compreendida entre a superfície terrestre até aproximadamente 50 km de altura. Ela é formada por partículas neutras e a maior concentração de gases encontra-se até uma altura de 12 km, composta por nitrogênio, oxigênio, dióxido de carbono, argônio, vapor d'água, entre outros, contendo aproximadamente 70% da massa total da atmosfera (SAPUCCI, 2001, 2005). As ondas eletromagnéticas ao se propagarem em tais camadas, sofrem diferentes influências, as quais são minimizadas de diferentes formas.

No ajustamento de observáveis GPS para fins de posicionamento geodésico, valores iniciais do principal efeito da troposfera (Atraso Zenital Troposférico – *Zenital*

*Tropospheric Delay* – ZTD), geralmente, são calculados a partir de modelos matemáticos empíricos. Os modelos mais utilizados para a realização dessa tarefa são o de Hopfield e o de Saastamoinen, os quais foram desenvolvidos com base em dados coletados em sua maioria no hemisfério Norte (SEEBER, 2003). Outras estratégias podem ser adotadas tal como a estimativa do ZTD a partir do ajustamento pelo Método dos Mínimos Quadrados (MMQ). Porém, existem diversas aplicações que empregam apenas modelos empíricos. Dessa forma, modelos adequados para aplicações sobre o território brasileiro fazem-se necessários.

Nesse trabalho, foi realizada uma análise de diversas variáveis meteorológicas medidas na superfície terrestre para verificar sua contribuição na modelagem do ZTD. Tanto os valores do ZTD quanto os valores de superfície foram obtidos a partir das observações realizadas por diversas radiosondas lançadas sobre a cidade de Curitiba. Nessa análise, foram empregados modelos de regressão linear múltipla.

## 2 MODELO DE REGRESSÃO LINEAR

A análise por regressão é a metodologia estatística para predição de valores de uma ou mais variáveis respostas (dependentes) a partir de uma coleção de valores de variáveis explicativas (independentes). Essa análise também pode ser utilizada para verificar os efeitos das variáveis predictoras na variável resposta. O modelo de regressão linear, com uma simples variável resposta  $Y$ , relacionado com as variáveis predictoras  $z_1, z_2, \dots, z_r$ , é dado pela equação:

$$Y = \beta_0 + \beta_1 z_1 + \dots + \beta_r z_r + \varepsilon \quad (1)$$

Na expressão (1), a componente  $\varepsilon$  representa os resíduos devido aos efeitos aleatórios e sistemáticos. O termo “linear” refere-se ao fato da média ser uma função linear dos parâmetros incógnitos  $\beta_0, \beta_1, \dots, \beta_r$ . As variáveis predictoras podem ou não entrar no modelo como termos de primeira ordem. Com  $n$  observações independentes em  $Y$  e valores associados de  $z_i$ , o modelo completo se torna:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 z_{11} + \beta_2 z_{12} + \dots + \beta_r z_{1r} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 z_{21} + \beta_2 z_{22} + \dots + \beta_r z_{2r} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 z_{n1} + \beta_2 z_{n2} + \dots + \beta_r z_{nr} + \varepsilon_n \end{aligned} \quad (2)$$

A determinação dos parâmetros desconhecidos do modelo é feita a partir de alguma técnica que minimize a soma dos quadrados dos resíduos, o que pode ser feito pelo MMQ (JOHNSON; WICHERN, 1992).

### 2.1 Avaliação do modelo

A avaliação do modelo de regressão consiste na verificação do quanto o modelo de regressão linear é representativo. Então, utiliza-se a seguinte expressão:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad (3)$$

onde, a componente do lado esquerdo da igualdade representa o desvio das observações com relação à média, denominado de Soma Quadrática Total (SQT). Já a primeira componente do lado direito da igualdade, representa o desvio do valor observado em relação ao valor ajustado, ou seja, a Soma Quadrática dos Resíduos (SQRes), enquanto que a segunda componente do lado direito da igualdade é o desvio do valor ajustado com relação à média, a qual trata-se da Soma Quadrática da Regressão (SQReg). Dessa forma, utilizam-se alguns coeficientes para a análise estatística do modelo escolhido:

- Coeficiente de determinação  $R^2$ : representa o lucro relativo introduzido pelo modelo, ou seja, a porcentagem da variabilidade dos valores

observados  $Y_i$ 's explicada pelo modelo de regressão. Quanto maior o valor de  $R^2$ , melhor o modelo ajusta os dados. A equação de  $R^2$  é dada por:

$$R^2 = \frac{SQReg}{SQT} \quad (4)$$

- $R^2$  ajustado: leva em conta o número de predictoras no modelo e é útil para comparar modelos com diferentes números de variáveis predictoras. A fórmula é dada por:

$$R^2_{ajust} = 1 - \frac{SQRes}{SQT/gl} \cdot \frac{n-1}{n-p} \quad (5)$$

Na expressão (5),  $n$  representa o número de observações,  $p$  é o número de variáveis regressoras e  $gl$  representa os graus de liberdade.

Uma maneira de realizar a escolha de forma automática do subconjunto de variáveis mais adequadas para o modelo de regressão, é utilizar o método denominado *Stepwise* “passo a passo” ou o método *Forward* “Passo a Frente”, cuja descrição mais detalhada pode ser encontrada em (JOHNSON; WICHERN, 1992).

## 3 RESULTADOS

As variáveis, suas respectivas unidades e denominações envolvidas na análise, bem como as componentes do ZTD estão definidas na Tabela 1:

Tabela 1 – Variáveis utilizadas na análise por regressão.

Variável	Denominação	Unidade
ZTD	Atraso zenital troposférico	(m)
DZHR	Atraso zenital troposférico hidrostático	(m)
DZWR	Atraso zenital troposférico úmido	(m)
PresR	Pressão atmosférica	(hPa)
PWs	Pressão parcial do vapor d'água	(hPa)
PHs	Pressão parcial hidrostática	(hPa)
Psatu	Pressão de saturação	(hPa)
Tempe	Temperatura	(°C)
Tvirtual	Temperatura virtual	(K)
URela	Umidade relativa	(%)
RM	Razão de mistura	(g/kg)
TPO	Temperatura do ponto de orvalho	(°C)
TPoten	Temperatura Potencial	(K)
TPEquiv	Temperatura Potencial equivalente	(K)
DVento	Direção do vento	(°)
VVento	Velocidade do vento	(m/s <sup>2</sup> )
Zonal	Componente zonal do vento	(m/s <sup>2</sup> )
Merid	Componente meridional do vento	(m/s <sup>2</sup> )

A princípio foi realizada uma análise para identificar quais variáveis observadas possuam distribuição normal. Em seguida foi calculada a correlação entre as variáveis, de forma que foi possível verificar quais estão mais correlacionadas com o ZTD. Todas as análises foram realizadas no *software* MiniTab.

As variáveis Psatu, URela, DVento, Zonal e Merid, não foram aceitas no teste de normalidade e, conseqüentemente não foram utilizadas na análise inicial. O teste de normalidade utilizado foi o de *Ryan-Joiner*, o qual é baseado na correlação.

### 3.1 Análise de correlação e agrupamento das variáveis

A análise de correlação mostra que a pressão atmosférica (PresR) apresenta uma correlação de 100% com a componente hidrostática do Atraso Zenital Troposférico (DZHR) e, para a pressão parcial hidrostática (PHS) essa correlação foi de 80%.

No caso da componente úmida (DZWR), a variável que apresenta a mais forte correlação é a temperatura do ponto de orvalho (TPO) com valor de 70,9%, seguida pela pressão parcial do vapor d'água (PWs) e razão de mistura (RM) com valores de 69,8% e 69,6%, respectivamente. Além disso, há diversos outros grupos de variáveis que apresentam forte correlação entre si. Dessa forma, optou-se por fazer um agrupamento em cluster, cujas variáveis que se agrupam com as componentes úmida e hidrostática do ZTD são apresentadas na Tabela 2:

Tabela 2 – Variáveis agrupadas com DZHR e DZWR.

COMPONENTES DO ZTD	VARIÁVEIS AGRUPADAS
DZHR	PresR PHS
DZWR	PWs Tempe Tvirtual RM TPO TPoten TPEquiv

O agrupamento apresentado na Tabela 2 mostra que as variáveis agrupadas, provavelmente, são as que devem fazer parte do modelo de regressão, visto que elas apresentam mais forte correlação do que as outras variáveis. Além disso, verifica-se que a modelagem da componente úmida requer mais variáveis que a modelagem da componente hidrostática.

### 3.2 Modelagem do ZTD a partir de Regressão

Foram gerados modelos de regressão para as componentes DZWR e DZHR, usando apenas uma variável de cada vez. A intenção é analisar a influência de cada variável na variabilidade total do modelo. Em seguida, foram adicionadas novas variáveis para compor os modelos de regressão múltipla do DZWR e DZHR.

A Tabela 3 apresenta os modelos de regressão gerados para a componente hidrostática usando somente uma variável em cada modelo (regressão simples), além dos valores de R<sup>2</sup> e R<sup>2</sup> ajustado.

Tabela 3 - Modelo de regressão linear para DZHR.

Modelo de Regressão Linear (1 variável)	R <sup>2</sup> (%)	R <sup>2</sup> ajustado (%)
DZHR = -0,00204 + 0,00228 PresR	100,0	100,0
DZHR = 0,920 + 0,00130 PHS	63,8	63,1
DZHR = 2,27 - 0,000611 TPoten	5,6	3,8
DZHR = 2,15 - 0,000182 TPEquiv	4,0	2,2
DZHR = 2,09 - 0,000619 RM	2,7	0,8
DZHR = 2,09 - 0,000378 PWs	2,0	0,1
DZHR = 2,08 + 0,000812 VVento	2,4	0,5
DZHR = 2,09 - 0,000456 TPO	2,1	0,2
DZHR = 2,18 - 0,000339 Tvirtual	2,0	0,1
DZHR = 2,09 - 0,000361 Tempe	1,8	0,0

A partir da Tabela 3, verifica-se que o primeiro modelo, correspondente à variável pressão atmosférica PresR, explica 100% da variabilidade total do modelo de regressão linear para a componente hidrostática DZHR. Além disso, verifica-se uma grande diferença na percentual de explicação desse modelo com relação aos outros da Tabela 3. Dessa forma, não há necessidade de acrescentar novas variáveis. Logo, o modelo de regressão para DZHR é dado por:

$$DZHR = -0,00204 + 0,00228 \text{ PresR.} \quad (6)$$

A Tabela 4 apresenta os modelos gerados para a componente hidrostática usando somente uma variável em cada modelo e os valores de R<sup>2</sup> e R<sup>2</sup> ajustado.

Tabela 4 – Modelo de regressão linear para DZWR.

Modelo de Regressão Linear (1 variável)	R <sup>2</sup> (%)	R <sup>2</sup> ajustado (%)
DZWR = -0,190 + 0,0214 TPO	58,3	57,3
DZWR = -0,176 + 0,0179 PWs	56,9	56,1
DZWR = -0,167 + 0,0251 RM	56,5	55,7
DZWR = -1,79 + 0,00584 TPEquiv	52,7	51,8
DZWR = -3,78 + 0,0134 Tvirtual	39,2	38
DZWR = -0,101 + 0,0142 Tempe	35,9	34,7
DZWR = -3,88 + 0,0135 TPoten	35,3	34,1
DZWR = 6,70 - 0,00731 PHS	25,6	24,2
DZWR = 1,39 - 0,00135 PresR	0,4	0,0
DZWR = 0,159 + 0,00066 VVento	0,0	0,0

A variável que explica a maior porcentagem (58,3%) da variabilidade total do modelo de regressão para a componente úmida DZWR é a temperatura do ponto de orvalho TPO, como mostrado na Tabela 4, seguida pela pressão parcial do vapor d'água PWS e a razão de mistura RM com valores de R<sup>2</sup> iguais a 56,9% e 56,5%, respectivamente. Essas três variáveis são as mais correlacionadas com DZWR. Dessa forma, as variáveis foram acrescentadas uma de cada vez em cada modelo para a componente úmida e, com o auxílio do método *Stepwise* o modelo escolhido para DZWR foi:

$$DZWR = -0,228 + 0,0685 \text{ TPO} - 0,0564 \text{ RM.} \quad (7)$$

As variáveis TPO e RM no modelo para DZWR da expressão (7), explicam 58,30% da variabilidade total

( $R^2$ ) do modelo, além disso, o valor de  $R^2$  ajustado foi de 59,89%.

Usando os modelos (6) e (7), foram calculados os valores do ZTD e comparados com aqueles advindos do cálculo das observações de altitude pelas radiossondas, de forma que se obtiveram os resíduos do modelo ajustado. Dessa forma, o histograma de resíduos, cujos valores estão em metros, é apresentado na Figura 1.

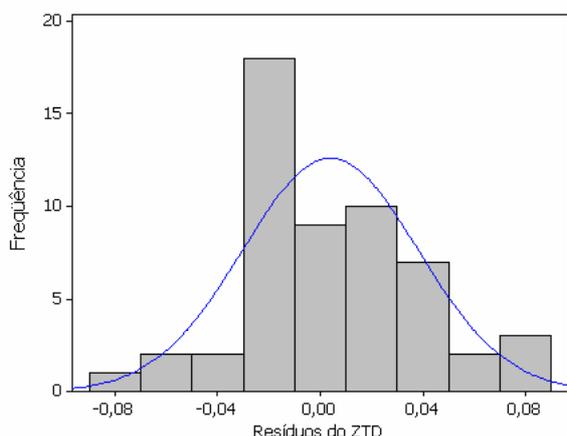


Figura 1 – Histograma dos resíduos do ZTD.

A análise de normalidade a partir do teste *Ryan-Joiner* no *software* MiniTab, indicaram que os resíduos seguem a distribuição normal, o que pode ser confirmado na Figura 1. Os resíduos ficaram num intervalo de mais ou menos 0,008 m, cuja média teve valor de 0,004 m, o desvio padrão foi de 0,035 m e o Erro Médio Quadrático (EMQ) foi de 0,035 m.

#### 4 CONCLUSÕES

Esse artigo apresentou a análise de variáveis meteorológicas medidas na superfície terrestre com o objetivo de encontrar as mais adequadas para a modelagem do ZTD, de forma que seja possível encontrar modelos mais apropriados para aplicações no território brasileiro. Para a realização dessa tarefa, foi feito um estudo a partir de regressão linear múltipla de algumas variáveis, onde se verificou que a componente hidrostática DZHR pode ser modelada apenas por valores da pressão atmosférica medidos nas estações terrestres. Essa variável explicou 100% da variabilidade do modelo.

Com relação à componente úmida DZWR, verifica-se que essa é mais difícil de ser modelada. A escolha do modelo de regressão foi feita com o auxílio do método *Stepwise* e as variáveis mais adequadas foram a temperatura do ponto de orvalho TPO e a razão de mistura RM, as quais explicaram 58,30% da variabilidade total dessa componente. Porém, há outras variáveis que apresentam correlação (em torno de 60% e 70%) com DZWR na mesma magnitude que a correlação de RM, como é o caso da pressão parcial do vapor d'água (PWS), da temperatura potencial equivalente (TPE) e da pressão de saturação (PWS). O modelo de regressão obtido foi

comparado com os valores advindos do cálculo das observações de altitude pelas radiossondas e os resíduos do modelo ajustado apresentaram um EMQ de 0,035m, valor esse que motiva a dar continuidade a essa pesquisa.

Essa análise envolveu somente uma estação na cidade de Curitiba – PR. É necessário levar em consideração que as condições atmosféricas variam em função do local e da estação do ano, principalmente, no caso do Brasil que possui dimensões continentais. Dessa forma, pretende-se realizar essa análise para outras estações em futuros trabalhos. Os resultados obtidos com essa análise ao considerar as diversas regiões climáticas do Brasil podem trazer uma contribuição significativa para a modelagem empírica do ZTD sobre o território brasileiro.

#### REFERÊNCIAS

JOHNSON, A. B. O.; WICHERN D. W. “**Applied Multivariate Statistical Analysis**”. New Jersey, 1992

LEICK, A. **GPS satellite surveying**. New York: John Wiley & Sons, 1995. 560p.

MINITAB. MINITAB QUALITY COMPANION. 2005: **MINITAB – Statistical Software**. Disponível em: <<http://www.minitab.com>>. Acesso em: mai. 2007.

SAPUCCI, L. F. **Estimativa do vapor d'água atmosférico e avaliação da modelagem do atraso zenital troposférico utilizando GPS**. 2001. 167 f. Dissertação (Mestrado em Ciências Cartográficas) - Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente.

SAPUCCI, L. F. **Estimativas do IWV utilizando receptores GPS em bases terrestres no Brasil: Sinergia entre a Geodésia e a Meteorologia**. 2005. Tese (Doutorado em Ciências Cartográficas) - Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista, Presidente Prudente.

SEEBER, G. **Satellite Geodesy: Foundations, Methods, and Applications**. 2<sup>nd</sup>. New York: Walter de Gruyter, 2003. 589p.