

Regressão linear simples na validação de modelos numéricos hidrodinâmicos

Simple linear regression in the validation of hydrodynamic numerical models

Camila Caroline Guedes Viana^{*}
Regina Serrão Lanzillotti^{**}

A modelagem numérica hidrodinâmica é uma ferramenta amplamente utilizada devido à sua importância no estudo hidrodinâmico de regiões costeiras e oceânicas, entretanto, para que seja fidedigna, a sua calibração e validação são fundamentais. O objetivo deste artigo é evidenciar a necessidade de se realizarem análises estatísticas inferenciais que possibilitem a aferição da aderência do modelo hidrodinâmico aos dados observados em campo. Para tanto, valoriza-se o método de regressão linear simples, a ênfase para realizar-se a ANOVA, a inferência para adequabilidade e a representação gráfica do estudo dos resíduos.

Palavras-chave: Modelagem numérica hidrodinâmica. MOHID. Validação. Regressão linear simples. ANOVA.

Hydrodynamic numerical modeling is a widely used tool because of its importance in the hydrodynamic study of coastal and oceanic regions. However, for these to be reliable, calibration and validation is essential. The objective of this paper is to highlight the need to perform inferential statistical analyses that allow the determination of the adherence of the hydrodynamic model to the data observed in the field. Therefore, the simple linear regression method, the emphasis to perform the ANOVA, the inference for adequacy and the graphical representation of the study of the residues are valued.

Keywords: Hydrodynamic numerical modeling. MOHID. Validation. Simple linear regression. ANOVA.

1 Introdução

Um modelo é uma representação simplificada de algo real, realizado em diferentes escalas e graus de refinamento, possibilitando a visão panorâmica de um sistema, para compreendê-lo melhor, permitindo uma análise sinótica do ambiente. A escolha tem grande influência na resolução de uma situação, pois a modelagem necessita adequar-se ao sistema (ASSAD et al, 2009).

O uso de modelos numéricos computacionais consiste numa abordagem facilitadora para o entendimento da dinâmica oceanográfica, por isso seus resultados devem ser comparados a dados observacionais tendo uma validação consistente.

Restringindo-se ao modelo hidrodinâmico utilizado neste artigo, este corresponde ao MOHID, desenvolvido pelo Instituto Superior Técnico de Portugal e HIDROMOD Ltda. É programado em ANSI FORTRAN95, recorrendo à programação orientada por objetos, ferramenta que possibilita o uso de vários modelos encaixados em simultâneo, sendo a comunicação ocorrente

^{*} Graduada em Oceanografia (UERJ). Mestranda em Oceanografia pela Universidade do Estado do Rio de Janeiro (UERJ) - Rio de Janeiro/RJ - Brasil. E-mail: cgv.camila@gmail.com.

^{**} Doutora em Engenharia de Transportes (PET-COPPE-UFRJ). Professora na Graduação e Pós-graduação IME/UERJ, professora associada no PGCM/UERJ - Rio de Janeiro/RJ - Brasil.

em um único sentido, com isso, as condições de contorno são fornecidas do modelo geral aos submodelos. Esse procedimento permite estudar com detalhe áreas de pequenas dimensões com uma malha muito fina (MOHID, 2011; MIRANDA et al., 2000).

O sistema MOHID é constituído por mais de 40 módulos capazes de simular a hidrodinâmica, os fenômenos de dispersão, com abordagens lagrangiana e euleriana, a qualidade da água e o transporte de sedimentos (MOHID, 2011).

O módulo hidrodinâmico, presente neste trabalho, calcula as equações primitivas do movimento em três dimensões, assumindo as aproximações de Boussinesq, Reynolds e o equilíbrio hidrostático (MARETEC, 2013; NEVES et al., 2001); sendo as equações discretizadas espacialmente pelo método numérico de volumes finitos (CHIPPADA et al., 1998; MARTINS et al., 2001), o que propicia a utilização de um sistema genérico de coordenadas verticais (sigma, cartesiana, lagrangiana, “espaçamento fixo” e harmônica) permitindo ao usuário selecioná-las, dependendo das características da área de estudo, e dos processos de interesse (GONÇALVES, 2007). No plano horizontal, o grid é ortogonal Cartesiano, com apenas um grau de liberdade, em uma grade C de Arakawa (MOHID, 2009).

A discretização temporal é feita a partir do método ADI (*Alternate Direction Implicit*). Esse algoritmo computa alternadamente uma componente da velocidade horizontal implicitamente, enquanto a outra é calculada explicitamente, permitindo preservar as vantagens da estabilidade (métodos implícitos) sem que haja a desvantagem de elevados custos computacionais e erros associados. Possibilita a adoção de passos de tempo relativamente grandes (MARETEC, 2013; MOHID, 2012).

A análise estatística de resultados também é utilizada para verificar a adequação dos modelos adotados em previsões. Várias premissas são importantes na análise estatística, sendo que uma delas é a escolha adequada da modelagem adotada e dos testes pertinentes para verificação da fidedignidade dos resultados, pois um equívoco pode levar a conclusões inconsistentes.

A adoção do modelo implica, inicialmente, verificar a distribuição de probabilidade em função da distribuição de frequência dos dados; se converge para a Curva Normal Padrão, a distribuição deve ser considerada simétrica e mesocúrtica, pois se percebe que os dados se concentram em torno da média e se dispersam simetricamente a partir desse ponto central (Figura 1).

Em função do perfil observado, serão utilizados os resultados de elevação da superfície livre extraídos do MOHID, que sofrerão a análise estatística de regressão linear simples para validar a aderência de modelos numéricos hidrodinâmicos aos dados levantados em campo. Caso isso ocorra, há viabilidade no prosseguimento da implementação do modelo hidrodinâmico.

A relevância deste artigo fundamenta-se na escassez de análises estatísticas inferenciais realizadas em modelos hidrodinâmicos. Contudo, a ocorrência de uma boa representação do cenário de simulação ambiental e conseqüentemente a previsão fidedigna, somente ocorrerão caso as validações sejam aceitáveis. Uma vez que isso ocorra, há permissão no uso em estudos hidrodinâmicos, diagnósticos e prognósticos e/ou valores de entrada em outro modelo numérico hidrodinâmico complementar.

2 Metodologia

Em modelagem numérica oceanográfica tem-se observado a utilização de modelo de aderência (ajustamento) para validar o modelo hidrodinâmico em função da raiz do erro quadrático médio (RMSE -*Root-Mean Square Error*). Outro indicador é a raiz do erro quadrático médio relativo (NRMSE), consubstanciado na razão do RMSE em relação à amplitude máxima de variação dos dados observados (normalização). Acresce o cálculo tanto do coeficiente de correlação linear (r), quanto o *skill* (d).

Esse procedimento, quando aplicado sobre a série de amplitude de maré, indica o ajustamento à elevação e o deslocamento de fase entre a onda de maré modelada e a observada, respectivamente. Quanto ao *skill*, possibilita a interpretação da concordância entre os valores observados e modelados, de ambos os parâmetros oceanográficos (amplitude e fase) (WILLMOTT, 1982). As expressões inerentes aos modelos de aderência nesses estudos encontram-se a seguir:

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=0}^n (^{n}mod - ^{n}obs)^2} \quad (1)$$

$$NMRSE = \frac{RMSE}{(\max(^{n}obs) - \min(^{n}obs))} * 100\% \quad (2)$$

$$r = \frac{\sum_{i=0}^n ((^{n}mod - \mu(^{n}mod)) * (^{n}obs - \mu(^{n}obs)))}{\sqrt{\sum_{i=0}^n (^{n}mod - \mu(^{n}mod))^2 * \sum_{i=0}^n (^{n}obs - \mu(^{n}obs))^2}} \quad (3)$$

$$d = 1 - \frac{\sum_{i=0}^n (^{n}mod - ^{n}obs)^2}{\sum_{i=0}^n (|(^{n}mod - \mu(^{n}obs))| + |(^{n}obs - \mu(^{n}obs))|)^2} \quad (4)$$

Onde, n é o tamanho da amostra e as variáveis assumem valores tais quais,

^{n}mod = elevação da superfície livre obtida do modelo

^{n}obs = elevação da superfície livre observada

A proposta deste artigo é complementar o procedimento descrito anteriormente agregando testes estatísticos que aferem a aderência da distribuição de frequência à distribuição de probabilidade (SIEGEL 1956). A regressão linear simples estabelece uma relação funcional entre duas variáveis minimizando o erro de estimativa e a análise da variância verifica a adequabilidade do modelo inferencial estatístico (MORETTIN; BUSSAB, 2013).

2.1 Teste de normalidade dos dados

Os testes de normalidade são usados para determinar se as variáveis se ajustam à distribuição de probabilidade Normal e, entre os utilizados, destaca-se o teste de Jarque-Bera, que utiliza a assimetria e curtose (FERREIRA, 2006).

$$JB = n \left[\frac{A^2}{6} + \frac{(C-3)^2}{24} \right] \quad (5)$$

Onde:

n : tamanho da amostra

A : coeficiente de assimetria da amostra

C : coeficiente de curtose da amostra

A estatística teste JB segue uma distribuição qui-quadrado com 2 graus de liberdade, $X^2_{\alpha,2}$, cuja hipótese nula (H_0) corresponde à normalidade dos dados inerentes às variáveis.

Se $JB > X^2_{\alpha,2}$, rejeita-se a H_0 .

2.2 Regressão linear simples

Consiste na obtenção da função linear que explique a variação da variável dependente (Y , resposta) pela variação da variável independente (X , explicativa), sendo os parâmetros estimados pelo Método dos Mínimos Quadrados. Esse método consiste em minimizar a soma dos quadrados das distâncias entre os valores estimados e observados. Dessa forma estabelece-se uma relação funcional entre essas variáveis minimizando o erro de estimativa.

$$Y = a + bX + \varepsilon \quad (6)$$

Onde:

Y = valor observado para a variável dependente.

X = valor observado para a variável independente.

a = constante de regressão, representa o intercepto da reta com o eixo Y .

b = coeficiente de regressão, representa a variação de Y em função da variável X , coeficiente angular da reta, tangente da linha regressiva.

ε = erro associado à distância entre o valor observado (Y) e o correspondente ponto na curva, do modelo proposto, para X .

2.3 Coeficiente de correlação linear de Pearson (r)

O coeficiente de Pearson (r) exprime grau de intensidade da associação linear existente entre as variáveis (MORETTIN; BUSSAB, 2013).

$$r = \frac{\sum_{i=0}^n [(mod - \mu(mod)) \cdot (obs - \mu(obs))]}{\sqrt{\sum_{i=0}^n (mod - \mu(mod))^2 \cdot \sum_{i=0}^n (obs - \mu(obs))^2}} \quad (7)$$

Onde:

mod = variável obtida do modelo

obs = variável observada em campo

2.4 Grau de adequação do ajustamento (r^2)

O grau de adequação (r^2) indica a qualidade do ajustamento, isto é, a proporção da variação explicada pela regressão. Esse índice varia no intervalo fechado de 0 a 1, sendo que os valores próximos de 1 (um) são indicativos da adequação do modelo hidrodinâmico. O método consubstancia-se na equação:

$$SQT = SQE + SQR \quad (8)$$

Onde:

SQT = Soma dos Quadrados Total (SQT)

SQE = Soma dos Quadrados Explicada (SQE)

SQR = Soma dos Quadrados dos Resíduos (SQR)

$$r^2 = \frac{SQE}{SQT} = 1 - \frac{SQR}{SQT} \quad (9)$$

2.5 Análise de variância (ANOVA)

A ANOVA consiste de uma tabela que verifica a viabilidade de se utilizar o modelo linear regressivo, ou seja, avaliar se o modelo linear postulado é adequado (Tabela 1). A decisão da adequação do modelo proposto é verificada pela distribuição de probabilidade F de Snedecor, visto que a estatística-teste corresponde à razão entre a variância da regressão e a obtida pelos resíduos.

| 187 |

Tabela 1: Tabela ANOVA

Fonte de Variação	Soma dos Quadrados	Graus de Liberdade (DF)	Média Quadrática	F de Snedecor
Regressão	SQRegressão	k-1	$QMRegressão = \frac{SQRegressão}{K - 1}$	$\frac{QMRegressão}{QMResíduo}$
Resíduo, ou independente da Regressão	SQResíduo	(n-2)-(k-1)	$QMResíduo = Se^2 = \frac{SQResíduo}{(n - 2) - (k - 1)}$	
Total	SQTotal	n-2	$QMTotal = \frac{SQTotal}{n-2}$	

Onde:

K: número de parâmetros da função regressiva

Hipótese nula, H_0 : Coeficiente angular (b) = 0

Se $F_{\text{calculado}} < F_{\text{tabelado}}$, aceita-se a hipótese nula, sendo a estatística teste, o F de Snedecor.

2.6 Histograma de resíduos

O histograma permite visualizar se os resíduos aderem à curva normal padrão. Caso isso ocorra, considera-se que a distribuição dos erros tem caráter aleatório, logo o modelo de regressão linear pode ser considerado adequado.

2.7 Fluxograma da regressão linear simples na validação de modelos numéricos hidrodinâmicos

O fluxograma a seguir apresenta as etapas para validar modelos hidrodinâmicos segundo a inferência da regressão linear simples.

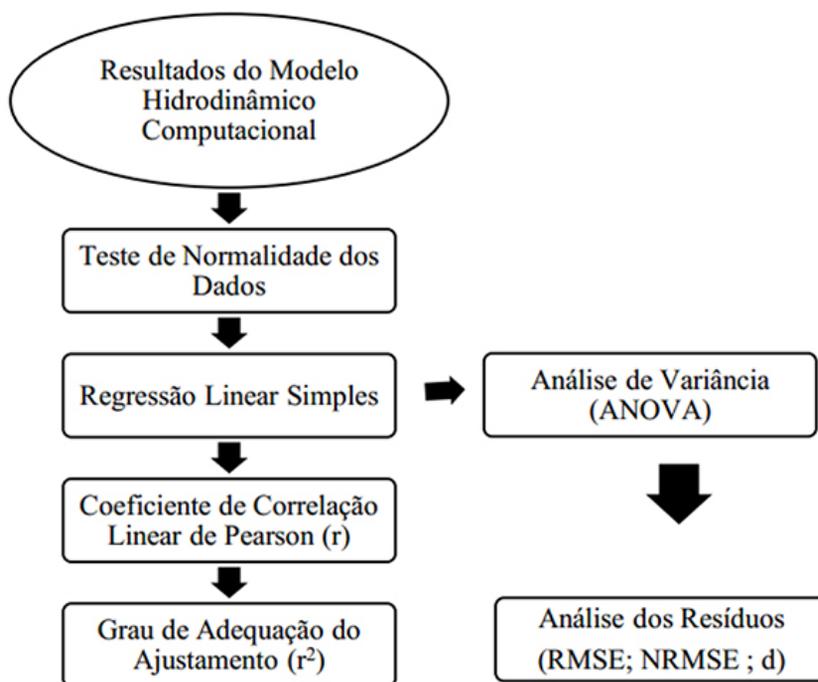


Figura 1. Fluxograma das etapas para validar modelos hidrodinâmicos

3 Aplicação e discussão

A faixa de interesse compreende a região metropolitana do estado do Rio de Janeiro e engloba toda a área da Baía de Guanabara (BG), entre 22°41' e 22° 58' S e 43°02' e 43°18' W, e a região da Plataforma Continental Interna adjacente desde Ponta Negra (22° 57' S e 042° 41' W) a leste até a Ilha Rasa de Guaratiba (23°04'S e 043° 32' W) a oeste, limitada a sul pela latitude 23°18'S (Figura 2).

O domínio do modelo possui duas grades aninhadas, ambas com apenas um nível vertical. A grade externa se estende aproximadamente 40 km adiante, considerando o fim da BG como sendo a região da Fortaleza de Santa Cruz. A grade interna se estende 8 km afora.

As equações são resolvidas em duas grades regulares. A grade externa possui resolução espacial de aproximadamente 200 metros (0.002°) com 325 linhas e 428 colunas (Figura 2). A grade interna (aninhada) possui resolução espacial de aproximadamente 50 metros (0.0005°), possuindo 645 linhas e 505 colunas (Figura 2).

O arquivo batimétrico resulta da junção de um arquivo de digitalização de cartas náuticas e folhas de bordo, além de cartas náuticas (1506 e 1620) digitalizadas no software QGIS 2.10.1 (anteriormente denominado Quantum GIS).

O modelo foi inicializado do repouso, tendo como forçamento a maré do FES 2004, com imposição da elevação, amplitude e fase de 14 constituintes harmônicas.

Para avaliação dos resultados, serão utilizados dados de 15 estações, apresentadas e nomeadas na Figura 3, onde serão realizadas comparações das elevações da superfície do mar, entre os valores resultantes do modelo — resultantes de um modelo 2D forçado com maré, a partir de constantes extraídas do FES 2004, com a grade externa (200 m) —, e a previsão realizada com o PACMARÉ (FRANCO, 1992), a partir de componentes harmônicas da Sales et al. (2000).

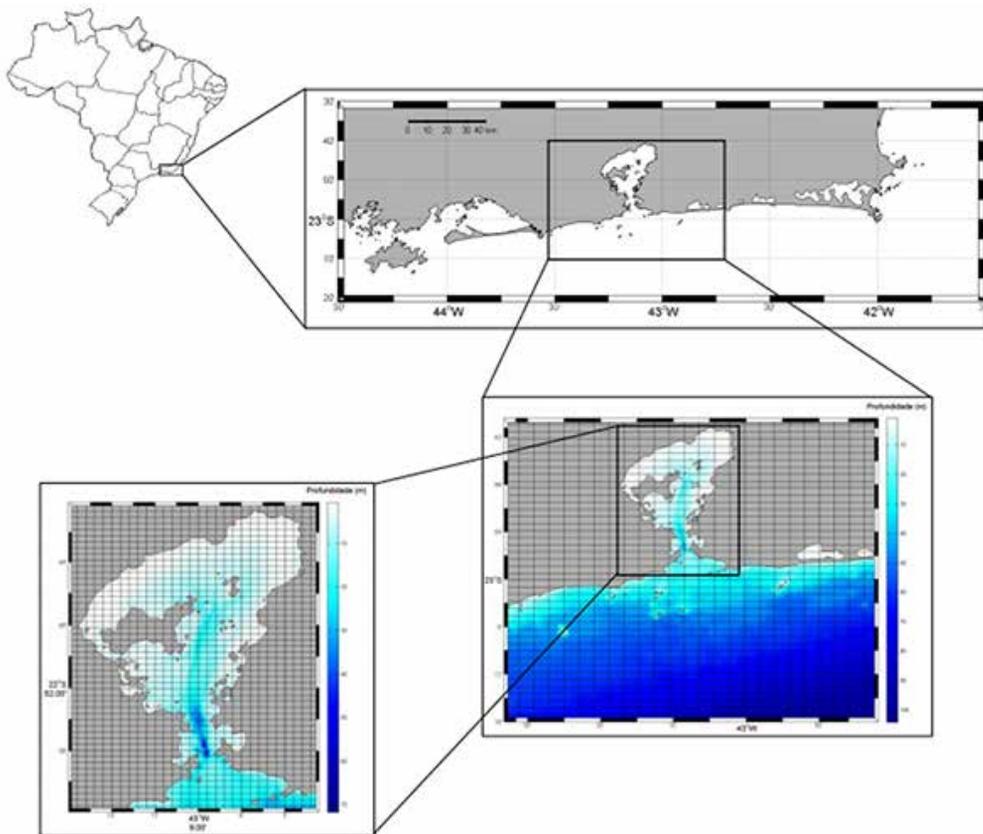


Figura 2: Localização da região de interesse, no estado do Rio de Janeiro. Bem como o destaque do mapa batimétrico da Baía de Guanabara para as grades externa e interna

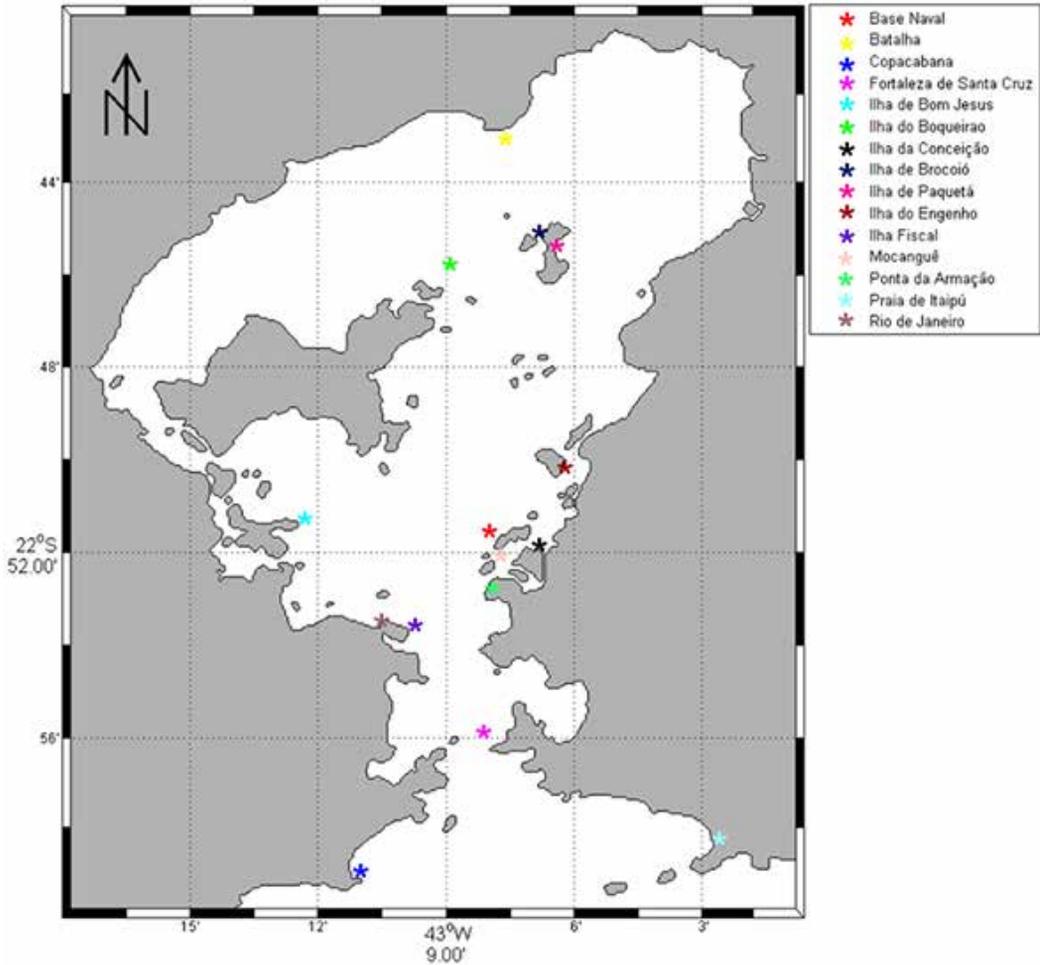


Figura 3: Localização dos pontos de avaliação da elevação da superfície livre do modelo

3.1 Análise dos dados

A Tabela 2 apresenta as medidas estatísticas de posição (média) e dispersão (desvio padrão), além dos coeficientes de assimetria e curtose para cada estação de observação. Esses valores permitiram realizar o teste de normalidade de JB.

Os valores da média tanto para o modelo hidrodinâmico quanto para os dados observados em campo estão bastante próximos a zero, medidas condizentes, já que a elevação da superfície livre varia em torno do valor (nível) de referência zero. Uma vez que as médias foram apresentadas com duas decimais, isso leva a crer na nulidade da medida, o que não deve ser interpretado dessa forma, visto que os desvios padrões variaram de 0,29 a 0,35. O valor máximo das médias foi 0,08, Ponta da Armação, o que não deve ser considerado como algo atípico, já que essa variação não é significativa no contexto dessa validação. A mesma aferição reporta-se à Ilha de Bom Jesus, que apresentou média -0,01.

A regressão linear deve ser aplicada quando a distribuição dos valores em apreço segue a distribuição Gaussiana (Normal Padrão). Neste estudo todas as estações apresentaram aderência a essa distribuição de probabilidade, viabilizando o método proposto. A estatística do teste JB impõe o cálculo dos coeficientes de assimetria e curtose, estes explicitados na Tabela 2, organizada de maneira a apresentar primeiramente os pontos mais internos à BG, seguindo aos mais externos.

Tabela 2: Medidas estatísticas para comparação entre os resultados do MOHID e as previsões do PAC-MARÉ para as estações maregráficas

Estações	Média		Desvio Padrão		Assimetria		Curtose		Teste JB
	MOHID	Observado	MOHID	Observado	MOHID	Observado	MOHID	Observado	
Batalha	0,00	0,04	0,35	0,34	-0,07	0,14	2,36	2,35	Dist. Normal
Ilha de Brocoió	0,00	0,01	0,35	0,34	-0,07	0,17	2,36	2,34	Dist. Normal
Ilha de Paquetá	0,00	0,00	0,34	0,32	-0,07	-0,15	2,36	2,36	Dist. Normal
Ilha do Boqueirão	0,00	0,03	0,34	0,34	-0,07	0,14	2,36	2,43	Dist. Normal
Ilha do Engenho	0,00	0,00	0,33	0,31	-0,07	0,07	2,36	2,46	Dist. Normal
Ilha de Bom Jesus	0,00	-0,01	0,32	0,31	-0,08	0,02	2,36	2,20	Dist. Normal
Base Naval	0,00	0,00	0,32	0,30	-0,08	0,06	2,36	2,32	Dist. Normal
Ilha da Conceição	0,00	0,00	0,32	0,35	-0,08	-0,07	2,36	2,23	Dist. Normal
Mocanguê	0,00	0,00	0,32	0,32	-0,08	0,13	2,35	2,50	Dist. Normal
Rio de Janeiro	0,00	0,05	0,31	0,30	-0,08	-0,01	2,36	2,35	Dist. Normal
Ponta da Armação	0,00	0,08	0,31	0,31	-0,09	-0,03	2,36	2,50	Dist. Normal
Ilha Fiscal	0,00	0,03	0,32	0,29	-0,08	-0,08	2,34	2,28	Dist. Normal
Fortaleza de Santa Cruz	-0,01	0,00	0,29	0,29	-0,10	-0,02	2,36	2,53	Dist. Normal
Praia de Itaipu	0,00	0,00	0,27	0,26	-0,14	-0,05	2,39	2,40	Dist. Normal
Copacabana	0,00	0,00	0,27	0,27	-0,14	-0,15	2,39	2,35	Dist. Normal

A relação de funcionalidade entre o MOHID e os dados coletados foi obtida pelo coeficiente de correlação (r), sendo que a regressão linear simples permitiu estabelecer a funcionalidade inferencial dessa relação. Nesta fase optou-se por tratar os dados segundo o cotidiano adotado nas áreas afins à oceanografia para validar modelos hidrodinâmicos. As medidas aferidas neste sentido metodológico encontram-se na Tabela 3, organizada de maneira a apresentar primeiramente os pontos mais internos à BG, seguindo os mais externos.

Em todas as estações ocorreram coeficientes angulares positivos, o que mostra a concordância na tendência crescente entre o resultado do MOHID e os dados coletados. Inclusive o mais expressivo é verificado para a Ponta da Armação, localizado próximo à entrada da BG, contrapondo à Praia de Itaipu, que mostrou menor inclinação linear, concordância aquém das

demais estações. Da observação dos resultados constantes nessa tabela, pode-se concluir que as estações que se encontram em regiões rasas possuem piores concordâncias, ainda assim, pode-se considerar que, em relação aos estudos recentes CARVALHO (2011), HELINE (2009), MARIANO (2014), RIBEIRO (2013), os resultados verificados são bem adequados.

A concordância praticamente perfeita, quando o coeficiente linear está próximo a zero, pode ser avaliada com precisão de duas decimais para as estações Ilha de Paquetá, Ilha do Engenho, Base Naval, Ilha da Conceição, Mocanguê, Fortaleza de Santa Cruz, Praia de Itaipu e Copacabana.

A validação do ajuste linear para o MOHID e os dados coletados devem ser visualizados pelo coeficiente de determinação (r^2). As estações situadas em regiões de menor profundidade apresentaram menores aderências, explicitadas pela Ilha da Conceição, Paquetá e Praia de Itaipu, valores de 0,76, 0,83 e 0,74, respectivamente. As encontradas em regiões de maior profundidade, coeficiente de 0,96 obtido para Copacabana, Rio de Janeiro e Ponta da Armação, embora nas demais localidades a variação esteja no intervalo entre 0,90 a 0,94. Essa apreciação conduz a aferir que a regressão linear simples foi uma função que representou a aderência.

Sob o enfoque tradicional, o RMSE apresentou valores baixos, entre 0,19 e 0,06, em todas as estações, sendo que os maiores valores foram verificados nas estações Ilha da Conceição, Ilha de Paquetá e Praia de Itaipu. Os valores referentes ao *skill* foram bastante altos, superiores a 0,84 (84%), aferindo que os erros apresentados pelo modelo podem ser considerados desprezíveis.

Tabela 3: Medidas aferidas segundo o cotidiano adotado nas áreas afins à oceanografia para validar modelos hidrodinâmicos

Estações	Regressão Linear	r	r ²	RMSE	NRMSE	Skill
Batalha	$Y = 0,9099x + 0,0421$	0,93	0,86	0,12	7,0	0,94
Ilha de Brocoió	$Y = 0,9494x + 0,0256$	0,96	0,92	0,10	7,2	0,95
Ilha de Paquetá	$Y = 0,8577x + 0,0008$	0,91	0,83	0,15	9,0	0,89
Ilha do Boqueirão	$Y = 0,9414x + 0,0314$	0,97	0,94	0,10	6,8	0,96
Ilha do Engenho	$Y = 0,9154x + 0,0020$	0,96	0,92	0,09	6,2	0,96
Ilha de Bom Jesus	$Y = 0,9286x - 0,0094$	0,95	0,90	0,07	5,1	0,97
Base Naval	$Y = 0,9210x + 0,0028$	0,97	0,94	0,06	4,7	0,98
Ilha da Conceição	$Y = 0,9411x + 0,0026$	0,87	0,76	0,19	11,4	0,84
Mocanguê	$Y = 0,9512 + 0,0029$	0,95	0,90	0,10	6,4	0,94
Rio de Janeiro	$Y = 0,9340x + 0,0489$	0,98	0,96	0,07	5,6	0,97
Ponta da Armação	$Y = 0,9666x + 0,0802$	0,98	0,96	0,08	5,6	0,97
Ilha Fiscal	$Y = 0,8870x + 0,0352$	0,96	0,92	0,09	7,3	0,96
Fortaleza de Santa Cruz	$Y = 0,9641x + 0,0054$	0,96	0,92	0,08	6,5	0,95
Praia de Itaipu	$Y = 0,8281x - 0,0001$	0,86	0,74	0,14	11,7	0,85
Copacabana	$Y = 0,9619x + 0,0003$	0,98	0,96	0,06	4,7	0,98

Os coeficientes de correlação (r) exibiram valores elevados, resultado favorável à interação entre os resultados do MOHID e o dado coletado, pois o mais expressivo, $r = 0,98$, foi verificado

nas estações Rio de Janeiro, Ponta da Armação e Copacabana, enquanto os menores, $r = 0,86$ e $r = 0,87$, para as estações Praia de Itaipu e Ilha da Conceição, nessa ordem.

Finalmente, coube realizar a ANOVA, cujas medidas obtidas para o cálculo da estatística teste F de Snedecor encontram-se na Tabela 4, com p_{valor} correspondente a 0,0000. O risco probabilístico da não aderência entre as duas medições pode ser considerado praticamente nulo, inclusive a estatística observada F de Snedecor foi mais significativa nas estações de águas de menor profundidade (Ilha da Conceição e Praia de Itaipu). Nas estações mais profundas, essa estatística-teste mostrou-se mais predominante nas estações Fortaleza de Santa Cruz e Ilha Fiscal.

Tabela 4: Tabela ANOVA aplicada às elevações (continua)

Fonte de Variação	Soma dos Quadrados	DF	Média Quadrática	F Snedecor	P _{Valor}
<i>BATALHA</i>					
X1	135,29	1	135,29	8902,2278	0,0000
Erro	20,41	1343	0,0152		
Total	155,7	1347	0,1156		
<i>ILHA DE BROCOIÓ</i>					
X1	145,56	1	145,56	16016,9668	0,0000
Erro	12,21	1343	0,0091		
Total	157,765	1347	0,1171		
<i>ILHA DE PAQUETÁ</i>					
X1	116,88	1	116,88	6615,1056	0,0000
Erro	23,73	1343	0,0177		
Total	140,61	1347	0,1044		
<i>ILHA DO BOQUEIRÃO</i>					
X1	141,33	1	141,33	18392,0727	0,0000
Erro	10,32	1343	0,0077		
Total	151,65	1347	0,1126		
<i>ILHA DO ENGENHO</i>					
X1	121,21	1	121,21	16486,063	0,0000
Erro	98,74	1343	0,0735		
Total	219,95	1347	0,1633		
<i>ILHA DE BOM JESUS</i>					
X1	118,97	1	118,97	13680	0,0000
Erro	11,68	1343	11,6800		
Total	130,65	1347	0,0970		
<i>BASE NAVAL</i>					
X1	116,93	1	116,93	2402,0220	0,0000
Erro	65,38	1343	0,0487		
Total	182,31	1347	0,1353		

Tabela 4: Tabela ANOVA aplicada às elevações (conclusão)

Fonte de Variação	Soma dos Quadrados	DF	Média Quadrática	F Snedecor	P _{Valor}
<i>ILHA DA CONCEIÇÃO</i>					
X1	123,49	1	123,49	4061,6935	0,0000
Erro	40,83	1343	0,0304		
Total	164,32	1347	0,1220		
<i>MOCANGUÊ</i>					
X1	122,94	1	122,94	11348,4377	0,0000
Erro	14,55	1343	0,0108		
Total	137,49	1347	0,1021		
<i>RIO DE JANEIRO</i>					
X1	113,92	1	113,98	2951,4719	0,0000
Erro	51,86	1343	0,0386		
Total	165,78	1347	0,1231		
<i>PONTA DA ARMAÇÃO</i>					
X1	120,7	1	120,7	2609,5932	0,0000
Erro	62	1343	0,0463		
Total	182,7	1347	0,1357		
<i>ILHA FISCAL</i>					
X1	105,94	1	105,94	1479,6828	0,0000
Erro	96,15	1343	0,0716		
Total	202,09	1347	0,1500		
<i>FORTALEZA DE SANTA CRUZ</i>					
X1	106,01	1	106,01	1487,7780	0,0000
Erro	95,69	1343	0,0713		
Total	201,70	1347	0,1497		
<i>ITAIPU</i>					
X1	68,19	1	68,187	3867,3568	0,0000
Erro	24	1343	0,0176		
Total	92,19	1347	0,0682		
<i>COPACABANA</i>					
X1	93,73	1	93,734	4036,1910	0,0000
Erro	31,19	1343	0,0232		
Total	124,92	1347	0,0927		

A visualização da distribuição dos resíduos convergiu para a distribuição normal padrão, simétrica e mesocúrtica (figuras 4 e 5). Neste artigo, a exemplificação foi feita para as estações externa e interna à BG (Batalha e Fortaleza de Santa Cruz), consideradas pelos autores como relevantes a estudos futuros.

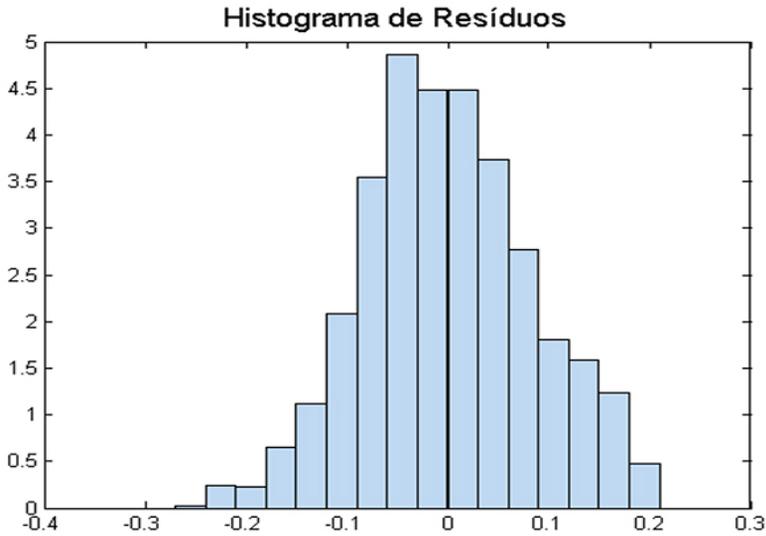


Figura 4: Distribuição dos resíduos para a estação Fortaleza de Santa Cruz

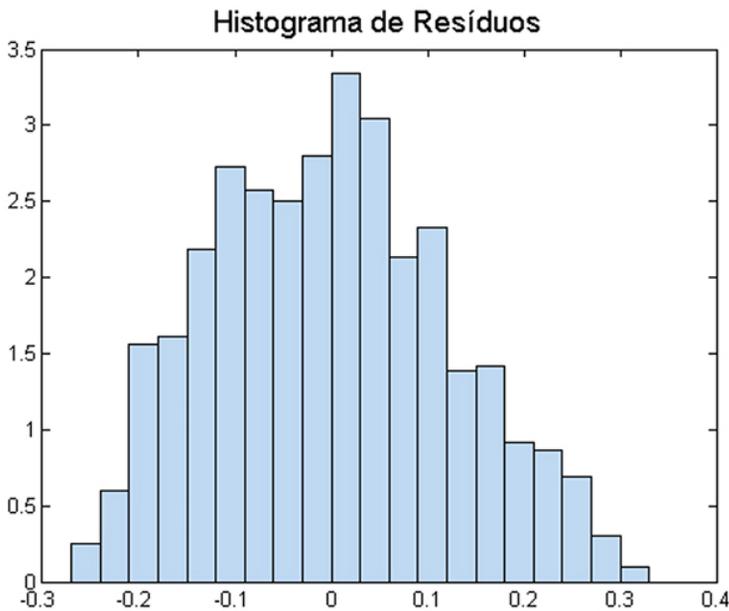


Figura 5: Distribuição dos resíduos para a estação Batalha

4 Conclusão

A utilização de métodos estatísticos é julgada como uma etapa primordial no que tange à utilização e desenvolvimento de modelos numéricos hidrodinâmicos. Assim sendo, é de extrema significância verificar a adequação da saída do modelo (MOHID) aos dados coletados ou ainda previstos, desde que sejam representativos do que se deseja estudar.

Contudo, a aderência somente poderá ser comprovada fazendo-se uso de metodologias estatísticas inferenciais adequadas ao estudo, pois a escolha incorreta poderá levar a conclusões infundadas. Na grande maioria das estações, os coeficientes, tanto de correlação quanto de ajustamento, evidenciaram um elevado grau de aderência entre as saídas do modelo e os dados coletados.

Somente não foram evidenciadas nas estações de Ilha da Conceição e Praia de Itaipu, entretanto esperam-se menores ajustes em áreas extremamente rasas e confinadas, como é o caso dessas duas localidades, uma vez que um modelo hidrodinâmico necessita de arquivos batimétricos de resolução espacial refinada. Apesar disso, essas estações apresentaram coeficientes de correlação e ajustamento aceitáveis quando comparados aos resultados obtidos por CARVALHO (2011), HELINE (2009), MARIANO (2014), RIBEIRO (2013).

Portanto, os grandes empecilhos à adequabilidade do modelo a dados coletados está na ausência de robustez dos dados batimétricos e o refinamento na resolução espacial do modelo.

Outro aspecto analítico obtido a partir da análise de variância (ANOVA) é a comprovação da adequabilidade do modelo linear regressivo aplicado a todas as estações, uma vez que F de Snedecor calculado foi menor que o tabelado, conduzindo a aceitação da hipótese nula.

Por conseguinte, é possível afirmar que essa análise possui refinamento e consistência, dada a realização da análise de variância, proporcionando um grau de refinamento que o coeficiente de correlação e o grau de ajustamento não retornam.

Em todas as estações, a distribuição de resíduos apresentou aderência à curva normal mostrada nos histogramas, porém, quando avaliados os resíduos segundo o NRMSE e *skill*, apenas obtêm-se os índices numéricos, carece a visualização. Ressalta-se que a ocorrência de aderência entre os resultados do modelo hidrodinâmico e os dados coletados, possibilita a viabilidade no prosseguimento do desenvolvimento desse modelo.

Considera-se como acurada essa proposta metodológica, uma vez que ela valoriza a representação gráfica, a inferência para adequabilidade e a ênfase para realizar-se a ANOVA, método ainda não popularizado entre os pesquisadores das áreas afins da oceanografia física. A relevância deste artigo deve ser evidenciada porque contribui para minimizar a escassez de análises estatísticas inferenciais realizadas em modelos hidrodinâmicos.

Referências

ASSAD, L. P. F.; MANO, M. F.; DECCO, H. T.; JUNIOR, A. R.T; LANDAU, L. *Noções básicas de modelagem hidrodinâmica computacional e de dispersão de poluentes*. Rio de Janeiro: COPPE/UFRJ, 2009. 126p.

BLUMBERG, Alan F.; KANTHA, Lakshmi H. Open boundary condition for circulation models. *Journal of Hydraulic Engineering*, v. 111, n. 2, p. 237-255, 1985.

CARVALHO, J. S. *Avaliação de um modelo numérico 2D para o estudo da circulação da Baía de Guanabara, RJ*. 2011. Monografia (Bacharelado em Oceanografia) – Faculdade de Oceanografia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2011.

CHIPPADA S.; DAWSON C. N.; MARTINEZ, M. L.; WHEELER, M. F. A Godunov-type finite volume method for the system of shallow water equations. *Computer Methods in Applied Mechanics and Engineering*, vol. 151, Issues 1–2, p. 105-129, 15 January 1998.

FERREIRA, M. R. P. Análise da sensibilidade dos testes de normalidade de Jarque-Bera e Lilliefors em modelos de Regressão Linear. *Rev. Mat. Estat.*, São Paulo, v. 24, n. 4, p. 89-98, 2006.

FRANCO, A. S. *Marés-programas para previsão e análise*. Cotia – SP, 1992.

GONÇALVES, N. D. F. *Método de Volumes Finitos em Grades Não-Estruturadas*. Dissertação (Mestrado em Engenharia Matemática) – Faculdade de Ciências da Universidade do Porto, Departamento de Matemática Aplicada/Departamento de Matemática Pura, Cidade do Porto, 2007.

HELINE, A. O. *Aplicação do Modelo Numérico Hidrodinâmico Delft3D na Baía de Guanabara*. Trabalho de conclusão de curso (Graduação) – Centro Universitário Monte Serrat, Curso de Oceanografia, Santos-SP, 2009.

MARETEC. *Descrição do MOHID*. Instituto Superior Técnico da Universidade Técnica de Lisboa. 2013.

MARIANO, L. S. *Estimativas do tempo de residência na Baía de Guanabara a partir de um modelo numérico*. Monografia (Bacharelado em Oceanografia) – Faculdade de Oceanografia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

MARTINS, F.; LEITÃO, P.; SILVA, A.; NEVES, R. 3D modelling in the Sado estuary using a new generic vertical discretization approach. *Oceanologica Acta*, vol. 24 – Supplement, p. 51-62, 2001.

| 197 |

MIRANDA, R.; BRAUNSCHWEIG, F.; LEITÃO, P.; MARTINS, F.; SANTOS, A. MOHID 2000: A coastal integrated object oriented model. In: BREBBIA, C. A. & BLAIN, W. R. (Editors). *Hydraulic Engineering Software VIII*. WIT Press, 2000. p. 393-401.

MOHID. *Hydrodynamic and transport model user guide*. Instituto Superior Técnico, Technical University of Lisbon. 2009.

MOHID. *MOHID Description*. MARETEC - Instituto Superior Técnico da Universidade Técnica de Lisboa. 2011.

MOHID. *MOHID - Descrição do MOHID*. MARETEC - Instituto Superior Técnico da Universidade Técnica de Lisboa. 2012.

MORETTIN, P. A.; BUSSAB, W. O. *Estatística Básica*. 8. ed. São Paulo: Editora Saraiva, 2013.

NEVES, R.; SILVA, A.; BRAUNSCHWEIG, F.; LEITÃO, P.C.; WANDEL, R.; PINA, P.; PITA, G. Hidroinformática e optimização da gestão de ecossistemas aquáticos. In: A HIDROINFORMÁTICA EM PORTUGAL (Seminário), 2001, Lisboa, Portugal. *Atas...* Lisboa : Laboratório Nacional de Engenharia Civil, 2001. 1 CD-ROM.

RIBEIRO, L. G. *A dinâmica da maré na Baía de Sepetiba/RJ a partir da implementação de um modelo hidrodinâmico*. 2013. Monografia (Bacharelado em Oceanografia) – Faculdade de Oceanografia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2013.

SALES, F. J. P.; BENTES, F. C. M.; SANTOS, J. A. (Eds.). *Catálogo de estações maregráficas brasileiras*. Rio de Janeiro: FEMAR, 2000. 280p.

SIEGEL, S. *Nonparametric Statistics for the Behavioral Sciences*. New York, Toronto, London: McGraw-Hill Book Company, Inc., 1956.

WILLMOTT, C. J. Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, v. 63, n. 11, p. 1309-1313, 1982.