

Comparação da variabilidade da temperatura da superfície do mar estimada pelos sensores remotos AVHRR-NOAA e MODIS-AQUA nas estações da rede Antares

João Felipe Cardoso dos Santos
Milton Kampel

Instituto Nacional de Pesquisas Espaciais – Depto. Sensoriamento Remoto (INPE/DSR)
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{joaofcs, milton}@dsr.inpe.br

Abstract. Sea surface temperature (SST) is the longest ocean's time series product observed from space and have been measured by different sensor during this time. To merge their databases is necessary, primarily, fit the time series with some model and quantify the performance of this adjust. This article aims to compare the AVHRR and MODIS databases in order to use them like a single time series. This evaluation was inserted inside the Antares Network project, a net of time series stations around the South America and Caribbean coastal. SST daytime, nighttime and the mean values of these 11 μ m band acquisitions showed low quality measurements in nighttime data while the mean between daytime and nighttime acquisitions had the better results. Was observed a positive offset from AVHRR data to MODIS data about 2°C that was corrected by the generalized linear model created to global and local adjusts. In general the local SST's models had a better performance than global one. Stations located near the Equator had higher quality in the adjust while stations apart Equator had a higher quality in the correlation. This maybe cause by the amplitude and variance differences between these stations. This first approach to reach a comparison of SST time series suggest a linear regression like a good model to adjust the SST datas for the Antares users.

Palavras-chave: time series, linear regression models, ocean temperature, adjust performance, série temporal, modelos de regressão linear, temperatura do oceano, performance do ajuste.

1. Introdução

O estudo da temperatura da superfície do mar (TSM) está vinculado, primeiramente, em compreender e quantificar incertezas e precisão das medidas (Armstrong et al. 2012). A TSM representa a maior série temporal de medidas oceanográficas obtidas por sensoriamento remoto e muitos sensores a bordo de diferentes satélites já foram utilizados na estimativa deste produto. A geração de séries temporais longas muitas vezes envolve a utilização destes diferentes sensores orbitais, o que requer um esforço de intercomparação entre diferentes fontes de dados remotos. Portanto, para manter a qualidade dos dados de TSM e dar continuidade a estas longas séries temporais é importante a comparação de diferentes produtos a serem utilizados. Este trabalho teve como objetivo comparar a TSM medida pelos sensores *Advanced Very High Resolution Radiometer* (AVHRR) a bordo dos satélites da série NOAA e *Moderate Resolution Imaging Spectroradiometer* (MODIS) a bordo do satélite *Aqua* da NASA dentro da região abordada pela rede de observações latino-americana Antares. A rede Antares, criada em 2003, tem o objetivo de analisar mudanças de longo período nos ecossistemas costeiros em torno da América do Sul e Caribe (<http://www.dsr.inpe.br/antares/>). É composta por onze estações de pesquisa, e neste trabalho foram analisadas seis delas, escolhidas por fazerem parte do último relatório do projeto NANO (2014). Foi obtido um ajuste global e outro local para as estações da rede Antares e comparados a respeito de seus desempenhos e ajustes. Estes resultados são um primeiro esforço de padronizar e aperfeiçoar uma série de medições realizadas nas estações da rede Antares.

2. Metodologia de Trabalho

Médias mensais de TSM e da qualidade destes dados foram obtidos para as estações da rede Antares (Figura 1) dos seguintes bancos de dados: 1) dados mensais das passagens diurnas e noturnas do AVHRR-NOAA derivadas do projeto Pathfinder V5 (Kilpatrick et al. 2001), entre o período de janeiro de 1985 e dezembro de 2009. 2) dados mensais das passagens diurnas e noturnas do sensor MODIS-AQUA, nível de processamento 3 a bordo do satélite *Aqua*, entre o período de julho de 2002 e setembro de 2014. A série de TSM destes sensores coexiste entre o período de julho de 2002 a dezembro de 2009 e este foi o período utilizado neste trabalho.

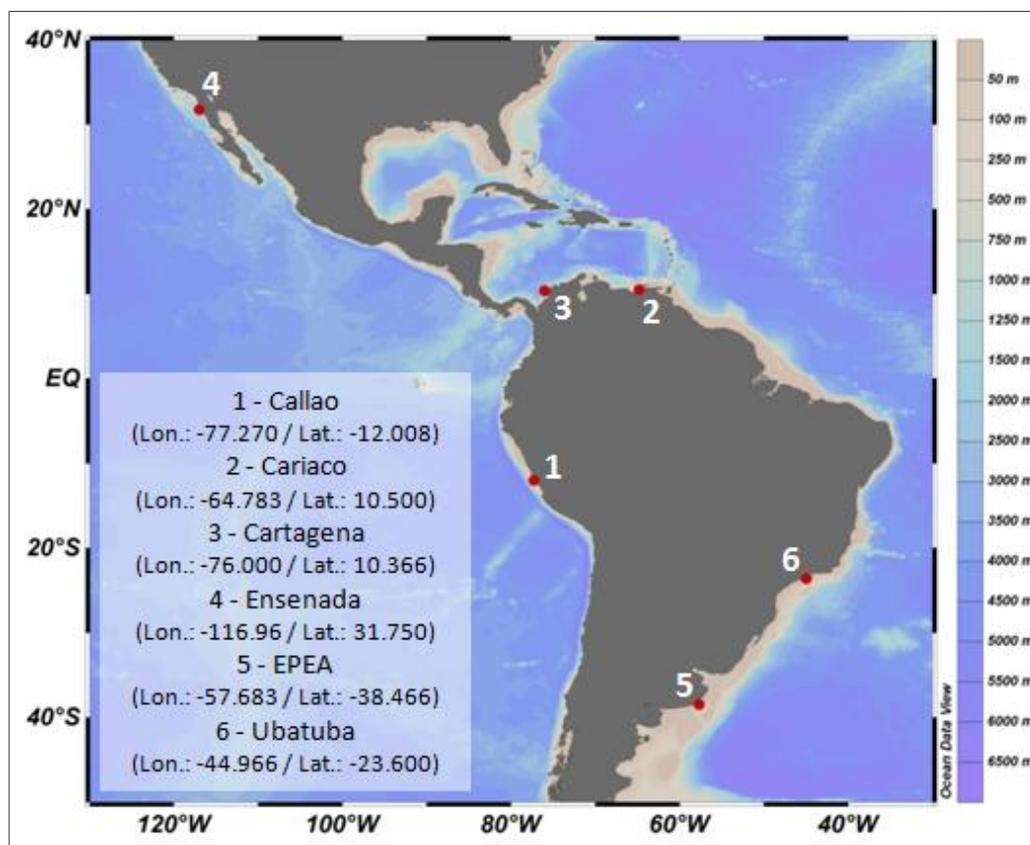


Figura 1: Localização das estações escolhidas da rede sul-americana do projeto Antares-ChloroGIN..

Os dados foram obtidos no formato HDF4, com resolução espacial de 4 km. As medidas, em cada estação, foram extraídas utilizando o software SeaDAS versão 7.1. e passaram por um critério de qualidade sugerido por Armstrong et al., (2012) onde foram descartadas as de qualidade inferior a 50%, como mostra a Tabela 1.

Tabela 1: Critério para descartar ou utilizar os dados de TSM dos sensores AVHRR-NOAA e MODIS-AQUA.

Sensor	Qualidade Inferior a 50% (Descartado)				Qualidade Superior a 50% (Utilizado)			
	0	1	2	3	4	5	6	7
AVHRR								
MODIS	Ruim		Questionável		Bom		Melhor	

Com os dados de qualidade superior a 50% foram gerados dois novos bancos de dados definidos como: AVHRR_(média) e MODIS_(média), que são os valores médios entre as medidas

diurnas e noturnas dos respectivos sensores. Os dados de TSM do AVHRR_{(dia),(noite),(média)} e do MODIS_{(dia),(noite),(média)} foram correlacionados em cada estação pré-definida da Figura 1. Com a série de dados com melhor correlação foi gerado um modelo de ajuste linear entre o AVHRR e MODIS para cada estação Antares, aplicando o modelo linear generalizado, do inglês *Generalized Linear Model* (GLM) no programa estatístico *R stats* versão 3.1.1. O desempenho de cada modelo foi avaliado pelo diagnóstico de independência de Durbin-Watson, pelo coeficiente de determinação (r^2) e correlação (r), pela raiz quadrada do erro médio quadrático (*Root mean square error*) [RMSE] e pelo teste de normalidade de Shapiro-Wilk. Analogamente foi obtido um modelo de ajuste global, utilizando os dados co-existent entre o AVHRR e MODIS de maior correlação sobre as seis estações concomitantemente. O desempenho do modelo global foi avaliado sobre os mesmos testes descritos anteriormente e aplicados no software PAST versão 3.

3. Resultados e Discussão

O banco de dados do AVHRR contém 300 meses de TSM. Destes, a estação de Callao, no Perú, apresenta a menor quantidade de dados com qualidade superior a 50%, totalizando 246, 128 e 256 medidas diurnas, noturnas e médias, respectivamente. As demais estações tiveram entre 84% e 100% dos seus dados utilizados, contemplando a quase totalidade da série temporal sem perda de informações (Tabela 2). É usual utilizar de uma janela de pixels para evitar a perda de informação, mas neste caso, como as estações estão próximas a região costeira, uma janela espacial pode aumentar a cobertura amostral e diminuir a eficácia do valor que se deseja. Por isso foi utilizado apenas o pixel correspondente a coordenada nominal. Com exceção das medidas obtidas em Callao durante a passagem noturna, os dados do MODIS apresentaram uma maior cobertura temporal (>96%). As demais estações medidas pelo MODIS, na passagem noturna, também foram as que apresentaram maior viés, fato que pode estar relacionado à faixa de banda de 11 μ m utilizada neste trabalho ao invés da banda de 4 μ m. Porém, como não existem medidas de TSM diurnas na banda de 4 μ m foi feita toda a análise com base na banda de 11 μ m, mantendo o padrão de correspondência com o AVHRR que obtém suas medidas de bandas na faixa de 11 μ m (Martin, 2004).

Tabela 2: Número de meses com medidas consideradas de qualidade superior a 50%.

	Número de medidas com qualidade superior à 50%					
	AVHRR			MODIS		
	(dia)	(noite)	(média)	(dia)	(noite)	(média)
Callao	246	128	256	143	105	145
Cariaco	279	295	300	147	147	147
Cartagena	256	274	294	147	145	147
Ensenada	294	256	298	147	143	147
EPEA	254	258	275	147	147	147
Ubatuba	275	277	290	147	146	147

Com as medidas superiores a 50% de qualidade, foi feita uma análise de correlação entre os bancos de dados AVHRR_{(dia),(noite),(média)} com o MODIS_{(dia),(noite),(média)}. Os valores do AVHRR_(média) e MODIS_(média) apresentaram os melhores coeficientes de correlação (Tabela 3). O pior resultado foi observado na estação de Cartagena, com um r de 86,49%, devido a pequena variância (1.129°C) da TSM ao longo da série (Tabela 4); fazendo com que pequenas diferenças entre os dados resultem em grandes diferenças e, conseqüentemente, em uma menor correlação.

Tabela 3: Coeficiente de correlação entre as estações da rede Antares e as passagens diurnas e noturnas e a média da TSM dos sensores AVHRR e MODIS.

Estação		r	Estação		r
Callao	AVHRR _(dia) x MODIS _(dia)	0.9148	Cariaco	AVHRR _(dia) x MODIS _(dia)	0.9536
	AVHRR _(noite) x MODIS _(noite)	0.9008		AVHRR _(noite) x MODIS _(noite)	0.9637
	AVHRR _(média) x MODIS _(média)	0.921		AVHRR _(média) x MODIS _(média)	0.9767
Cartagena	AVHRR _(dia) x MODIS _(dia)	0.8611	Ensenada	AVHRR _(dia) x MODIS _(dia)	0.9709
	AVHRR _(noite) x MODIS _(noite)	0.7124		AVHRR _(noite) x MODIS _(noite)	0.958
	AVHRR _(média) x MODIS _(média)	0.8649		AVHRR _(média) x MODIS _(média)	0.9753
EPEA	AVHRR _(dia) x MODIS _(dia)	0.9709	Ubatuba	AVHRR _(dia) x MODIS _(dia)	0.9022
	AVHRR _(noite) x MODIS _(noite)	0.958		AVHRR _(noite) x MODIS _(noite)	0.9276
	AVHRR _(média) x MODIS _(média)	0.9753		AVHRR _(média) x MODIS _(média)	0.9558

Weeks et al. (2008) justificam o uso da climatologia do AVHRR como referência para inferir a anomalia do MODIS baseado na premissa de que ambos os dados têm suas TSM corrigidas entre os valores de pele e de balde. Além disso, Martin (2004) demonstra que a medida na banda de 11 μ m do MODIS é a abordagem histórica implementada na série temporal do AVHRR. Por isso, o uso do AVHRR_(média) e MODIS_(média) para gerar o modelo de regressão e a série temporal acoplada, neste estudo de caso, foi considerado apropriado.

Os dois bancos de dados estão defasados (*offset*) em ΔT entre 2 e 3°C dependendo da estação (Tabela 4) e que pode ser observado em todas as estações na Figura 2. Valores mais elevados na TSM do AVHRR também foram observados por Weeks et al., (2008); onde somente dados noturnos foram utilizados.

Tabela 4: Tabela descritiva da TSM medida por AVHRR_(média) e MODIS_(média).

	Callao		Cariaco		Cartagena	
	AVHRR _(média)	MODIS _(média)	AVHRR _(média)	MODIS _(média)	AVHRR _(média)	MODIS _(média)
N amostral	69	89	90	90	88	90
Mínimo	17.625	12.068	26.813	23.446	29.700	26.605
Máximo	25.425	22.272	32.625	29.792	34.275	30.372
Média	21.232	17.410	29.593	26.425	31.772	28.672
Variância	2.946	3.198	2.416	2.692	1.129	0.804
Desv. Padrão	1.716	1.788	1.554	1.641	1.062	0.896
Mediana	21.225	17.054	29.475	26.292	31.931	28.878
	Ensenada		EPEA		Ubatuba	
	AVHRR _(média)	MODIS _(média)	AVHRR _(média)	MODIS _(média)	AVHRR _(média)	MODIS _(média)
N amostral	88	90	88	90	87	90
Mínimo	16.988	13.769	12.225	9.303	22.313	19.549
Máximo	25.238	21.759	24.150	21.217	31.088	27.963
Média	20.720	17.325	18.165	15.194	27.033	23.823
Variância	5.494	4.749	15.149	14.313	3.730	4.060
Desv. Padrão	2.344	2.179	3.892	3.783	1.931	2.015
Mediana	20.381	17.091	17.456	14.522	26.925	23.580

Uma possível explicação para este *offset* pode se dar ao fato do sensor MODIS-AQUA apresentar uma correção atmosférica ruim, que não consegue retirar por completo pixels com nuvens, que acabam influenciando na visada do sensor e sendo computado com a TSM.

As maiores temperaturas estão nas estações ao longo do Equador, que são Cariaco (Venezuela) e Cartagena (Colômbia) com TSM média de 29°C e 31°C, respectivamente. Estas também são as estações com menores variâncias de TSM ao longo da série (~2,5 em Cariaco e 1,0 em Cartagena) e, portanto, uma sazonalidade menos expressiva. A estação EPEA, na Argentina, é a que tem a maior variância (15,149°C) ao longo do ano e uma sazonalidade mais marcante. Os valores mínimos de TSM são observados na estação EPEA. Estas estações com maiores discrepâncias entre si serão observadas em relação ao ajuste do GLM

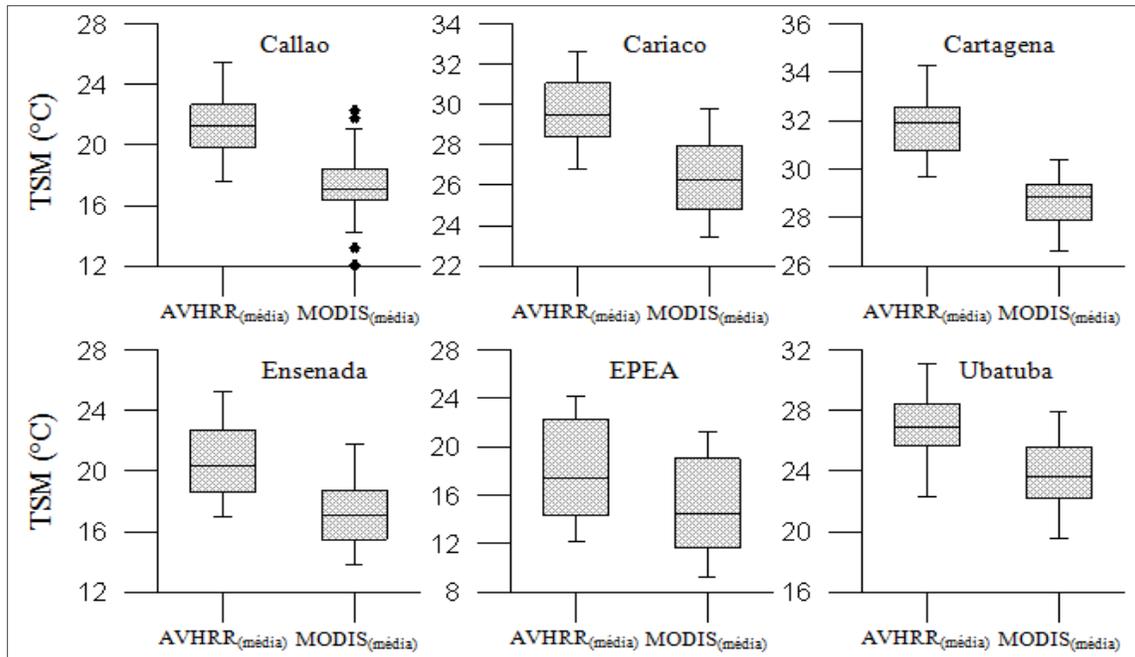


Figura 2: *Boxplot* dos dados mensais da TSM nas estações da rede Antares entre julho de 2002 e dezembro de 2009.

Os dados do MODIS_(média), mesmo sendo pré-avaliados pelo critério de qualidade, apresentaram valores *outliers* na estação de Callao (Figura 2). Estes *outliers* foram retirados, pois ao serem gerados os modelos de regressão e feita a análise dos resíduos, estes pontos invalidaram o modelo. Portanto, além das medidas de qualidade inferior a 50%, os meses de ago/2002 e fev/2008 do MODIS_(média) foram excluídos do modelo. Estes *outliers* também foram retirados da série temporal e substituídos pelos valores do AVHRR_(média) na série temporal acoplada (jan/85 a set/14).

A Figura 3 ilustra o modelo de GLM global gerado para todas as estações da rede Antares. Ainda sobre a Figura 3 pode-se observar em: a) a variação do *offset* na TSM do AVHRR_(média) e MODIS_(média) foi corrigida com o modelo global AVHRR_(ajustado); em c) os resíduos seguem uma distribuição não linear e variam entre -2,4 e 2,4; e) foi obtido um coeficiente de correlação e determinação de ~99%. O resultado do teste de Durbin-Watson foi 1,7, o que sugere uma independência entre os dados utilizados e ajustados (valores próximos a 2 são mais independentes). O teste de Shapiro-Wilk aplicado sobre os resíduos retornou um valor de ~0,96 com um coeficiente de correlação, Figura 3c), de ~98% com a normal. Resíduos com distribuição normal devem apresentar valores de Shapiro-Wilk próximos a 1, neste caso o GLM global tende a uma distribuição normal.

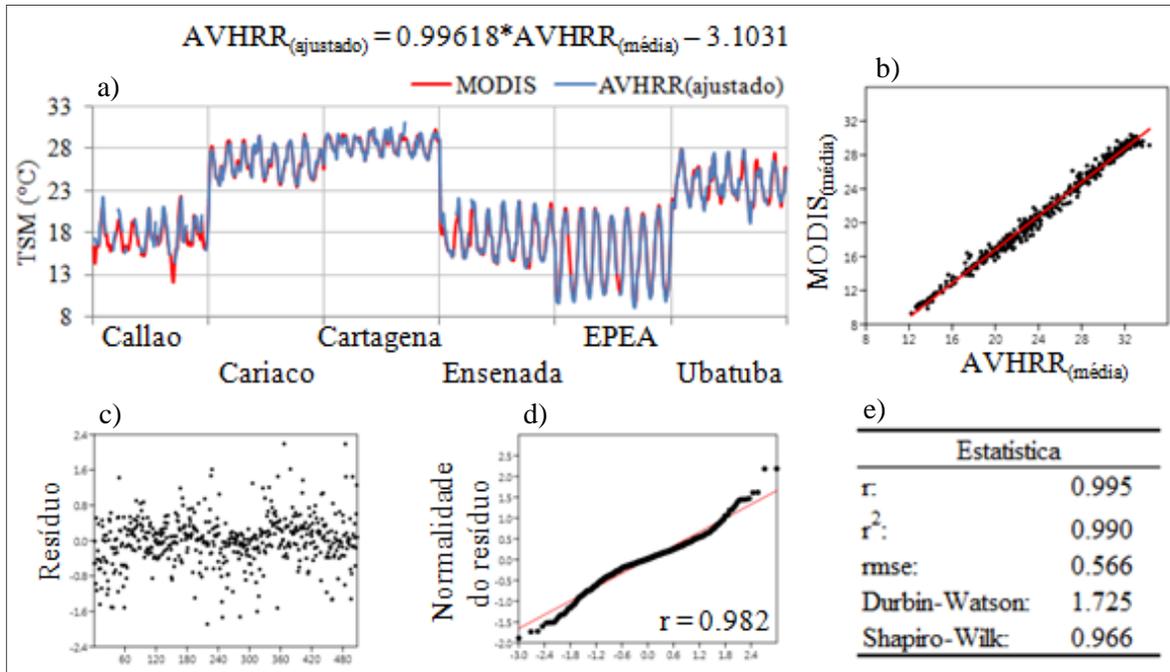


Figura 3: Função do GLM global e a análise de ajuste do modelo.

A Figura 4 representa a série temporal pelo modelo GLM local e global da estação Callao no Perú. Podemos observar que ambos os modelos foram capazes de ajustar as séries temporais sobre os dados do sensor MODIS_(média) mesmo tendo suas constantes (β_0) e seus coeficientes (β_1) de regressão bem distintos. Podemos observar ainda que o modelo local apresentou um desempenho melhor em relação ao modelo global, principalmente no valor do rmse de 0,543 para o GLM local e 0,759 para o global.

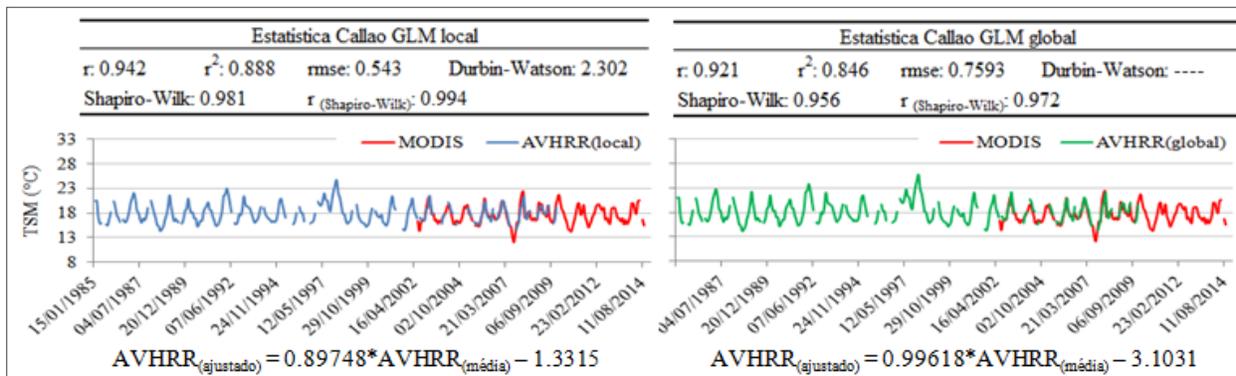


Figura 4: Desempenho do modelo local e global de ajuste da TSM na estação Callao, no Perú.

Analogamente à estação Callao, foram plotadas as séries temporais completas das demais estações da rede Antares. Semelhante ao observado na Figura 4, os modelos locais e o global de todas as estações foram capazes de ajustar o *offset* entre os dados satisfatoriamente. O resultado do ajuste está descrito na Tabela 5.

Tabela 5: Desempenho dos modelos locais e global de ajuste da TSM nas estações da rede Antares.

	Estatística GLM local			Estatística GLM global		
Equação →	$\text{Cariaco}_{(\text{ajustado})} = 1.031 * \text{Cariaco}_{(\text{média})} - 4.084$			$\text{Cariaco}_{(\text{ajustado})} = 0.996 * \text{Cariaco}_{(\text{média})} - 3.103$		
Cariaco	r: 0.977	r ² : 0.954	rmse: 0.350	r: 0.977	r ² : 0.954	rmse: 0.3574
	Shapiro-Wilk: 0.986 r (Shapiro-Wilk): 0.993			Shapiro-Wilk: 0.989 r (Shapiro-Wilk): 0.994		
Equação →	$\text{Cartagena}_{(\text{ajustado})} = 0.778 * \text{Cartagena}_{(\text{média})} - 3.933$			$\text{Cartagena}_{(\text{ajustado})} = 0.996 * \text{Cartagena}_{(\text{média})} - 3.103$		
Cartagena	r: 0.894	r ² : 0.799	rmse: 0.403	r: 0.865	r ² : 0.748	rmse: 0.540
	Shapiro-Wilk: 0.962 r (Shapiro-Wilk): 0.984			Shapiro-Wilk: 0.959 r (Shapiro-Wilk): 0.974		
Equação →	$\text{Ensenada}_{(\text{ajustado})} = 0.916 * \text{Ensenada}_{(\text{média})} - 1.653$			$\text{Ensenada}_{(\text{ajustado})} = 0.996 * \text{Ensenada}_{(\text{média})} - 3.103$		
Ensenada	r: 0.975	r ² : 0.951	rmse: 0.484	r: 0.975	r ² : 0.951	rmse: 0.559
	Shapiro-Wilk: 0.916 r (Shapiro-Wilk): 0.953			Shapiro-Wilk: 0.9024 r (Shapiro-Wilk): 0.9463		
Equação →	$\text{EPEA}_{(\text{ajustado})} = 0.978 * \text{EPEA}_{(\text{média})} - 2.548$			$\text{EPEA}_{(\text{ajustado})} = 0.996 * \text{EPEA}_{(\text{média})} - 3.103$		
EPEA	r: 0.993	r ² : 0.982	rmse: 0.437	r: 0.991	r ² : 0.982	rmse: 0.570
	Shapiro-Wilk: 0.967 r (Shapiro-Wilk): 0.982			Shapiro-Wilk: 0.964 r (Shapiro-Wilk): 0.9418		
Equação →	$\text{Ubatuba}_{(\text{ajustado})} = 0.999 * \text{Ubatuba}_{(\text{média})} - 3.178$			$\text{Ubatuba}_{(\text{ajustado})} = 0.996 * \text{Ubatuba}_{(\text{média})} - 3.103$		
Ubatuba	r: 0.962	r ² : 0.926	rmse: 0.5467	r: 0.956	r ² : 0.914	rmse: 0.5921
	Shapiro-Wilk: 0.966 r (Shapiro-Wilk): 0.982			Shapiro-Wilk: 0.952 r (Shapiro-Wilk): 0.972		

Os modelos locais tiveram melhor ou igual desempenho quando comparados ao modelo global em cada estação. Cariaco, na Venezuela, foi a estação com o melhor desempenho entre os modelos, com um rmse igual a 0,35. Embora Cariaco tenha apresentado o melhor desempenho, EPEA, na Argentina, foi a estação com o melhor resultado do coeficiente de determinação (r²) e correlação (r).

A Figura 2 mostra a diferença na amplitude e na variância da TSM nestas estações. A menor amplitude de TSM e maior variância, como discutido anteriormente, faz com que as estações de Cariaco e Cartagena consigam um desempenho melhor do modelo, mas não a melhor correlação. Isso porque, como a variância é alta e a análise de desempenho é, basicamente, a diferença entre o valor estimado e o medido, quanto maior a variância maior será a diferença empregada no cálculo do desempenho. Estações com maiores amplitudes de TSM e menores variâncias apresentam uma melhor correlação entre os sensores AVHRR e MODIS e um pior desempenho do GLM.

O alto índice de correlação (r) sugere que as estações de EPEA e Ensenada (México) sofram menos com as variações de curto período (diárias) na TSM. Dessa forma o padrão temporal se mantém estável por mais tempo na superfície do mar sendo analisado pelos sensores, ao longo do mês, com maior verossimilhança. Reynolds et al. (2007) observa essa mudança diária na TSM principalmente no Golfo do México durante o inverno em medidas do AVHRR aplicadas a um algoritmo de interpolação ótima – *Optimum Interpolation* (OI).

O teste de Shapiro-Wilk indica a normalidade dos resíduos (melhor quanto mais próximo de 1). Todos os modelos tendem a uma distribuição normal de seus resíduos, mas somente as estações de Callao e Cariaco têm um nível de significância superior a 5%. Essa informação pode ser útil na avaliação das anomalias da TSM, pois esta utiliza da climatologia sobre o critério de que os dados são independentes e normalmente distribuídos.

4. Conclusões

Os sensores AVHRR e MODIS mostraram-se eficazes para interpolação de séries temporais de longo período. A correlação dos bancos de dados mostrou que a média entre as passagens diurnas e noturnas foi mais representativa para esta finalidade, tendo em vista que novas abordagens devem ser discutidas para justificar o uso da média. A abordagem da técnica GLM foi eficaz, pois se pode observar que os dados estão altamente correlacionados e apenas defasados em um valor constante. Esta constante variou até 2 vezes entre as medidas dos sensores, mas tanto o modelo de ajuste global quanto o local solucionaram este efeito de *offset*. Além da constante de regressão variar consideravelmente entre o GLM local e global, as análises mostraram um melhor desempenho dos modelos locais. Possivelmente, modelos zonais possam ser empregados, pois foi observado que as estações mais próximas ao Equador respondem de maneira semelhante às variações na TSM medidas pelos sensores, da mesma forma que estações mais distantes também se assemelham. Outra forma de determinação de padrões de modelos regionais pode ser obtido através dos processos que regem cada região (exemplo: ressurgência ou ENSO), porém dados auxiliares de vento, índices climáticos e outras variáveis devem ser considerados nesse tipo de abordagem (Casey, 2002). Esta avaliação pode ajudar a encontrar regiões semelhantes de TSM e que possam servir de auxílio nas inferências das medidas das estações da rede Antares. O exercício de avaliação dos sensores AVHRR e MODIS para uso na rede Antares é apenas o primeiro passo para se poder interpretar os processos meteorológico-oceanoográficos que ocorrem na América do Sul e Caribe e esforços para distribuir e informar qualquer avanço neste sentido tem sido feito, como a exemplo deste trabalho.

Agradecimentos

Ao CNPq pela bolsa concedida, ao IOUSP pela participação no projeto Antares Ubatuba.

Referências Bibliográficas

- Armstrong, E.M.; Wagner, G.; Vazquez-Cuervo, J.; Chin, T.M. Comparisons of regional satellite sea surface temperature gradients derived from MODIS and AVHRR sensors. **International Journal of Remote Sensing**, v. 33, n. 21, p. 6639-6651, 2012.
- Casey, K.S. Daytime vs nighttime AVHRR sea surface temperature data: a report regarding Wellington et al. (2001). **Bulletin of Marine Science**, v. 70, n. 1, p. 169-175, 2002.
- Kilpatrick, K.A.; Podestá, G.P.; Evans R. Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database, **J. Geophys. Res.**, v. 106(C5), p. 9179-9197.
- Martin S. **An introduction to ocean remote sensing**. Cambridge University Press, 2004. 476 p.
- NANO. Variability in phytoplankton pigments at the Antares-ChloroGIN time-series-stations. **Report from the Latin American Regional Project 2013-2014**. 2014. 96p.
- Narayanan, M.; Vasan, D.T.; Bharadwaj, A.K.; Thanabalan, P.; Dhileeban, N. Comparison and validation of “sea surface temperature (SST)” using MODIS and AVHRR sensor data. **International Journal of Remote Sensing & Geoscience**, v. 2, i. 3, p. 1-7, 2013.
- Reynolds, R.W.; Smith, T.M.; Liu, C.; Chelton, D.B.; Casey, K.S.; Schlax, M.G. Daily High-Resolution-Blended Analyses for Sea Surface Temperature. **J. Climate**, v. 20, p. 5473-5496. 2007.
- Weeks, S.J.; Anthony, K.R.N.; Bakun, A.; Feldman, G.C.; Hoegh-Guldberg, O. Improved predictions of coral bleaching using seasonal baselines and higher spatial resolution. **Limnology & Oceanography**, v. 53, p. 1369-1375, 2008.