



Avaliação Preliminar do Desempenho da Versão Global do Sistema de Assimilação PSAS do CPTEC/INPE Segundo as Métricas da OMM

Helena Barbieri de Azevedo¹
Luis Gustavo de Gonçalves²
Luiz Fernando Sapucci³

¹CPTEC/INPE - Centro de Previsão de Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais
Rod. Presidente Dutra, Km40 - Cachoeira Paulista (SP), Brasil – CEP 12630-000
helena.azevedo@cptec.inpe.br

²CPTEC/INPE - Centro de Previsão de Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais
Rod. Presidente Dutra, Km40 - Cachoeira Paulista (SP), Brasil – CEP 12630-000
gustavo.goncalves@cptec.inpe.br

³CPTEC/INPE - Centro de Previsão de Tempo e Estudos Climáticos/Instituto Nacional de Pesquisas Espaciais
Rod. Presidente Dutra, Km40 - Cachoeira Paulista (SP), Brasil – CEP 12630-000
luiz.sapucci@cptec.inpe.br

Abstract:

The Center for Weather Forecast and Climate Studies (CPTEC - Centro de Previsão do Tempo e Estudos Climáticos), a division of the Brazilian National Institute for Space Research, initiated a pioneer work in Brazil of implementing an operational data assimilation system. The system called the Physical-Statistical Space Assimilation System (PSAS), developed at NASA, is coupled to the General Circulation Model (GCM) of CPTEC, called GPSAS (Global-PSAS) hereafter. This system is operational since early 2004 at T126L28 configuration, changing starting in December 2006 until now to a T213L48 configuration. The knowledge of the errors associated to the analysis created from the data assimilation system as well as the numerical models errors are of great importance to that process since it makes possible to adjust the system in order to decrease the forecast errors. Hence, this work studies the analysis and forecast errors for different regions of the globe i.e. Northern and Southern Hemisphere and Tropics and the entire globe and trying to relate with the observations flow into the PSAS. The metrics used to evaluate the numerical models are from the World Meteorological Organization. The results for the period from 2007 to 2010 show that the quality of the analysis is directly related to the number of observations assimilated with bigger impact over the Southern Hemisphere. This work also shows that there was a steady improvement in the performance of the GPSAS system during the past 3 years.

Resumo:

O Centro de Previsão de Tempo e Estudos Climáticos (CPTEC), pertencente ao Instituto Nacional de Pesquisas Espaciais (INPE), iniciou um trabalho pioneiro no Brasil de implementar um sistema de assimilação de dados rodando operacionalmente. O sistema empregado é chamado *Physical-Statistical Space Assimilation System* (PSAS) desenvolvido na NASA, que acoplado ao Modelo de Circulação Geral da Atmosfera (MCGA) do CPTEC, é denominado GPSAS. Esse sistema encontra-se operacional desde o início de 2004, sendo que até dezembro de 2006 foi utilizada a versão T126L28, com resolução horizontal de 100km, e depois dessa data até o presente utiliza-se a versão T213L48. Conhecer os erros que estão associados à análise gerada pelo sistema de assimilação assim como os erros devido às limitações dos modelos de Previsão Numérica do Tempo é de grande importância nesse processo, pois, permite ajustar o sistema de maneira a diminuir os erros de previsão. Diante disso, o presente trabalho visa estudar o comportamento dos erros das análises e previsões para diferentes regiões do globo, tais como Hemisfério Norte, Hemisfério Sul, Região Tropical, e também para o globo todo, tentando relacioná-los com o fluxo de dados assimilados pelo PSAS. As métricas utilizadas para avaliar o sistema GPSAS em sua última versão, seguem os padrões exigidos pela Organização Mundial Meteorológica para avaliação de modelos numéricos. Os resultados obtidos para o período de 2007 a 2010 mostram que a qualidade das análises está diretamente relacionada com o número de dados assimilados, com impactos maiores para os dados de satélite no Hemisfério Sul. Este trabalho também demonstra uma evolução gradativa na performance do sistema de assimilação GPSAS ao longo dos últimos 3 anos.



1- Introdução

A Previsão Numérica do Tempo (PNT) é considerada como um problema de valor inicial e de condição de contorno. Dada uma estimativa do estado presente da atmosfera e as apropriadas condições de contorno é possível através de um modelo de PNT estimar o estado futuro da atmosfera. Consequentemente, uma estimativa precisa das condições iniciais pode representar uma melhoria na qualidade das previsões numéricas. Sistemas avançados de assimilação de dados têm sido estudados, desenvolvidos e aprimorados com o objetivo de construir campos de análise com alta resolução temporal e espacial ao mesmo tempo em que os erros de inicialização possam ser melhorados na modelagem atmosférica. Tais sistemas combinam redes de observações espalhadas irregularmente no tempo e espaço com previsões de curto prazo a partir de modelos numéricos para geração das condições iniciais, também conhecidas como "análises".

As primeiras tentativas da obtenção de condições iniciais ótimas para modelos numéricos foram chamadas de "análises objetivas" como o algoritmo de Cressman (Cressman, 1959) em substituição das "análises objetivas" que eram geradas manualmente através de cartas meteorológicas. As análises objetivas consistiam em interpolações tri-dimensionais das observações pontuais para a grade regular dos modelos numéricos (3DDA; *3-Dimensional Data Assimilation*). Estes métodos evoluíram posteriormente para as análises quadri-dimensionais (4DDA; *4-Dimensional Data Assimilation*) como o método de relaxação de Newton (Stauffer and Seaman, 1994) e o filtro de Kalman (Kalman, 1960). No início dos anos 60 foi também introduzida (Gandin, 1963) a interpolação estatística ou interpolação ótima (OI; *Optimal Interpolation*) que é um 3DDA método que leva em consideração a distribuição espacial da função de covariância dos erros da previsão de curto prazo do modelo numérico (FG; *First-Guess*) utilizada para a confecção da análise. Na década de 80 foram introduzidos os métodos variacionais (Le Dimet and Talagrand, 1986) onde os campos meteorológicos precisam satisfazer as equações dinâmicas do modelo numérico empregado. Lorenc (1986) demonstrou, porém, que tanto para os métodos variacionais em três ou quatro dimensões (3D-Var e 4D-Var respectivamente) a solução envolve o uso de uma função custo que precisa ser minimizada o que na prática aumenta grandemente a complexidade do processo quando comparado com outros métodos de assimilação. Contudo uma grande vantagem é que as análises variacionais são dinamicamente mais consistentes.

No Brasil, só foi possível utilizar os sistemas de assimilação de dados operacionalmente e em escala global a partir da implantação do Centro de Previsão do Tempo e Estudos Climáticos (CPTEC), uma divisão do Instituto Nacional de Pesquisas Espaciais (INPE) em meados dos anos 90. Apesar dos esforços e pesquisas no CPTEC/INPE, o primeiro sistema de assimilação de dados operacional do Brasil e América do Sul só foi implementado depois de 2000. O sistema de assimilação escolhido foi o método variacional PSAS da sigla em Inglês "*Physical-Statistical Space Assimilation System*" desenvolvido na NASA. Esse sistema acoplado ao modelo de Circulação Geral da Atmosfera (MCGA) do CPTEC é denominado GPSAS, o qual encontra-se operacional desde o início de 2004, sendo que até dezembro de 2006 foi utilizada a versão T126L28, com resolução horizontal de 100km, e depois dessa data até o presente utiliza-se a versão T213L48, com resolução horizontal de 60km no equador.

O presente trabalho apresenta a metodologia utilizada para a avaliação dos campos de FG e análises contra dados observacionais nos resultados da assimilação de dados operacional do CPTEC obtidos com a versão atualmente operacional. Na próxima seção, a metodologia que segue o padrão de avaliação da Organização Mundial de Meteorologia (OMM) é descrita em detalhes além do conjunto de observações utilizadas. São mostradas também, métricas adicionais utilizadas para o desenvolvimento deste trabalho, assim como uma breve descrição do software de avaliação utilizado. Na seção "Resultados", as estatísticas das principais variáveis meteorológicas em níveis padrão são apresentadas, seguidas de uma análise e discussão sobre a evolução do GPSAS no Brasil nos últimos anos, a qual mostrada na seção "Conclusões".

2- Dados e metodologia

Foi utilizado neste trabalho o período de dados de junho de 2007 a dezembro de 2010, os quais estão armazenados no banco de dados do CPTEC.

De acordo com Herdies et al (2008) o PSAS, tanto em sua versão global como na regional, desde sua primeira versão operacional, está apto a assimilar observações convencionais em superfície, tais como altura geopotencial, valores do conteúdo total de água precipitável e componentes do vento (esse último apenas sobre os oceanos), bem como em altitude, como perfis de altura geopotencial, componentes do vento e umidade específica, considerando-se nesse caso apenas os níveis isobáricos padrões. As fontes dos dados em superfície são as estações



convencionais, as quais fornecem dados como velocidade e direção do vento, pressão em superfície, umidade relativa e temperatura, sendo que essa última variável não é assimilada diretamente pelo sistema de assimilação PSAS. O PSAS também assimila dados vindos de plataformas e bóias oceânicas e sensores a bordo de navios, dados esses de pressão e vento. Dados de ar superior são obtidos através de radiossondagens aviões e de dados de satélites, ao quais fornecem informações de altura geopotencial componentes do vento e umidade,

Para avaliar o sistema de assimilação de dados PSAS foram utilizadas métricas da Organização Meteorológica Mundial (OMM), tais como: *RMSE* (*Root Mean Square Error*) da altura geopotencial no nível de 500hPa, *RMSVE* (*Root Mean Square Vector Error*) do vento em 250hPa e 850hPa. Outras métricas também foram utilizadas, como: *BIAS* da altura geopotencial no nível de 500hPa. Também para uma melhor análise, foram estudadas regiões, tais como Hemisfério Norte, variando de 0° a 90°, Hemisfério Sul, variando de -90° a 0°, e Região Tropical, variando de -10° a 10°, além do globo inteiro.

O *RMSE* é calculado a partir da seguinte equação abaixo:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\varphi_i - \varphi_{iobs})^2}{N}} \quad (1)$$

Onde: φ_i = valor previsto em cada ponto i ;

φ_{iobs} = valor observado em cada ponto;

N = número de pontos observados.

O *RMSVE* é calculado a partir da equação:

$$RMSVE = \sqrt{\frac{1}{N} \sum_{i=1}^N [(u_i - u_{iobs})^2 + (v_i - v_{iobs})^2]} \quad (2)$$

Onde: u_i = valor previsto da componente u do vento no ponto i ;

u_{iobs} = valor observado da componente u do vento no ponto i ;

v_i = valor previsto da componente v do vento no ponto i ;

v_{iobs} = valor observado da componente v do vento no ponto i ;

N = número de pontos observados.

O *BIAS* é calculado a partir da equação:

$$BIAS = \frac{\sum_{i=1}^N (\varphi_i - \varphi_{iobs})}{N} \quad (3)$$

Onde: φ_i = valor previsto em cada ponto i ;

φ_{iobs} = valor observado em cada ponto;

N = número de pontos observados.

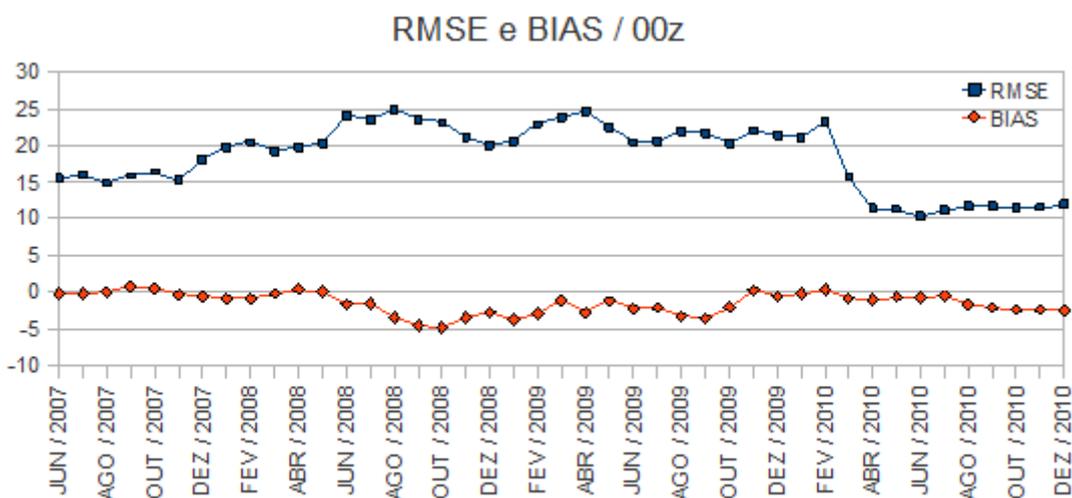
Essas estatísticas foram calculadas utilizando o software avaliador denominado *Meteostat*, o qual é uma ferramenta que foi desenvolvida dentro do Grupo de Desenvolvimento em Assimilação de Dados (GDAD) do CPTEC. Ele lê arquivos que possuem as diferenças entre observação e análise e observação e previsão, e processa estatísticas para as quatro variáveis que são assimiladas (componente u e v do vento, altura geopotencial e umidade específica). São calculadas estatísticas como *BIAS*, *RMSE*, a relação entre o *BIAS/RMSE* e número de observações, tudo em função de observação menos análise e/ou observação menos previsão. Ele



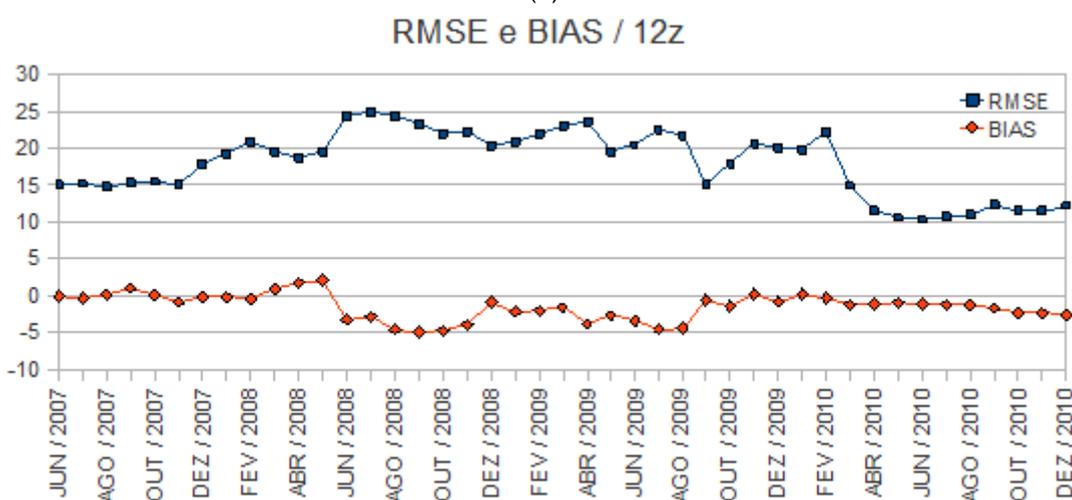
mostra esses resultados em forma de tabela ou em gráficos, gerando essas estatísticas para vinte níveis, começando da superfície até 5hPa. Com estes resultados pode-se acompanhar a evolução dos erros dentro do sistema de assimilação de dados e verificar possíveis problemas nos dados ou no modelo.

3- Resultados

Através do Meteostat foram geradas estatísticas para altura geopotencial em 500hPa, utilizando apenas dados da diferença entre a observação e a análise, para o período de junho de 2007 a dezembro de 2010. Foram feitas tabelas contendo *RMSE* e *BIAS*, para as 00Z e 12Z, a fim de gerar gráficos para uma melhor análise da evolução temporal desses erros. Os resultados desse processo são apresentados na Figura 1, onde as linhas em azul são referentes aos valores de *RMSE* e as linhas em vermelho são referentes aos valores de *BIAS*. É possível observar que para a altura geopotencial em 500hPa, de maio de 2008 a fevereiro de 2010 houve um aumento desses erros, tanto do *RMSE* como do *BIAS*. Já no último ano, a partir de março de 2010, nota-se um decréscimo acentuado do *RMSE*, tanto no horário das 00Z (Figura 1a) como no horário das 12Z (Figura 1b). Em ambos os horários verifica-se uma diminuição substancial, seguida de uma estabilização da curva do *RMSE* a partir de abril de 2010. Nota-se principalmente que quando há um aumento de um dos erros o mesmo é percebido no outro. Pode-se observar também que em dezembro de 2010 a curva do *BIAS* já está mais distante do zero, e o mesmo aumento é visto na curva do *RMSE*.



(a)



(b)

Figura 1. Variação temporal do RMSE e BIAS, da altura geopotencial em 500hPa, para os horários das 00Z (a) e 12Z (b). O eixo x representa o tempo em meses enquanto o eixo y representa a magnitude do RMSE e BIAS.



Utilizando o software GrADS (*Grid Analysis and Display System*) com dados da diferença entre a observação e o FG, foram calculados erros de *RMSE*, para a altura geopotencial em 500hPa, e *RMSVE*, para o vento em 850hPa e 250hPa, todos separados em três regiões distintas: Hemisfério Norte, Hemisfério Sul e Região Tropical. A Figura 2 mostra um exemplo, onde foi calculado o *RMSE* da altura geopotencial em 500hPa no Hemisfério Norte. O mesmo foi feito para as outras regiões tanto para as 00Z quanto para as 12Z.

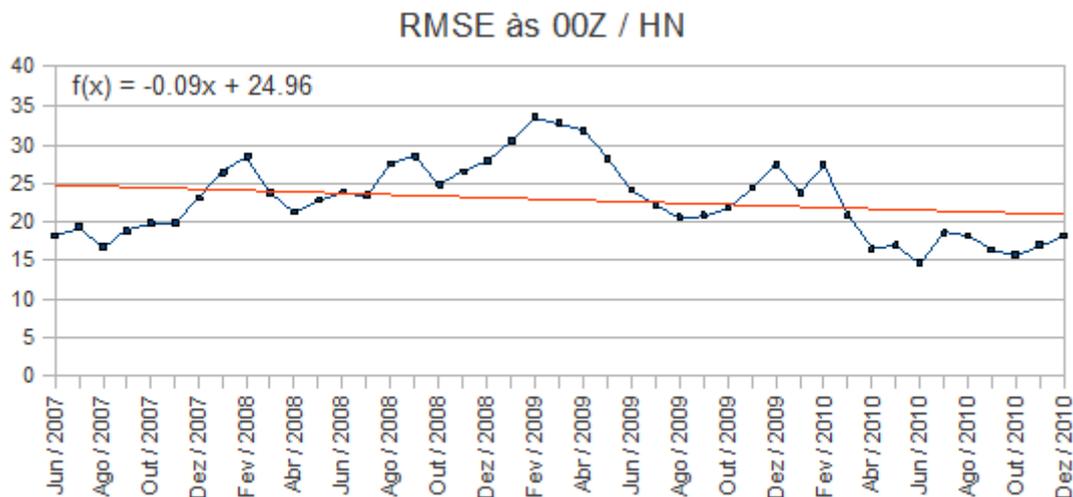


Figura 2. Variação temporal do RMSE às 00Z, da altura geopotencial em 500hPa, no Hemisfério Norte e a Linha de Tendência junto de sua equação. O eixo x representa o tempo em meses enquanto o eixo y representa a magnitude do RMSE em mgp.

Foram calculados também os coeficientes angular e linear da reta de tendência do *RMSE* e do *RMSVE*, a fim de quantificar a melhora durante todo o período estudado, os quais são mostrados na Tabela 1, ao considerar os diferentes horários avaliados (00Z e 12Z) e as regiões em estudo (HN, HS e TR) para a altura geopotencial em 500hPa (para o caso do *RMSE*), e do vento em 850hPa e vento em 250hPa (para o *RMSVE*). Nesta tabela é mostrado também a Porcentagem da Variação do Erro durante todo o período estudado, a qual mostra o aumento ou a diminuição do erro no período estudado. No caso do Hemisfério Norte, às 00Z, observa-se uma diminuição de 16% do *RMSE*, desde junho de 2007 a dezembro de 2010.

Tabela 1. Taxa Anual de Variação do *RMSE* (Geopotencial) e *RMSVE* (Vento) no período entre junho de 2007 e dezembro de 2010 para os horários das 00Z e 12Z.

00Z	RMSE			RMSVE					
	geo 500hPa			wind 850hPa			wind 250hPa		
O - F	Coef. Ang	Coef. Lin	V. Erro(%)	Coef. Ang	Coef. Lin	V. Erro(%)	Coef. Ang	Coef. Lin	V. Erro(%)
HN	-0,0947	24,961	0,16	-0,0089	4,8795	0,08	0,0057	7,2072	0,03
HS	-0,0917	27,431	0,14	-0,0381	5,9087	0,28	-0,0201	8,2959	0,10
TR	0,051	19,724	0,00	-0,0293	5,5321	0,23	0,0024	6,5328	0,02
12Z	RMSE			RMSVE					
	geo 500hPa			wind 850hPa			wind 250hPa		
O - F	Coef. Ang	Coef. Lin	V. Erro(%)	Coef. Ang	Coef. Lin	V. Erro(%)	Coef. Ang	Coef. Lin	V. Erro(%)
HN	-0,0876	24,926	0,15	-0,0044	5,0538	0,04	0,0062	7,284	0,04
HS	-0,0983	27,931	0,15	-0,0297	5,9694	0,21	-0,0152	7,7226	0,08
TR	0,0445	18,896	0,10	-0,0243	5,7835	0,18	-0,0082	6,4143	0,05

A Figura 3 mostra o número total de dados aceitos pelo sistema de assimilação de dados em função dos meses. É possível notar que o aumento no número de dados corresponde ao



mesmo período em que houve uma queda brusca nos erros calculados, mostrados na Figura 1, o que dá uma forte indicação das possíveis causas dessa melhora.

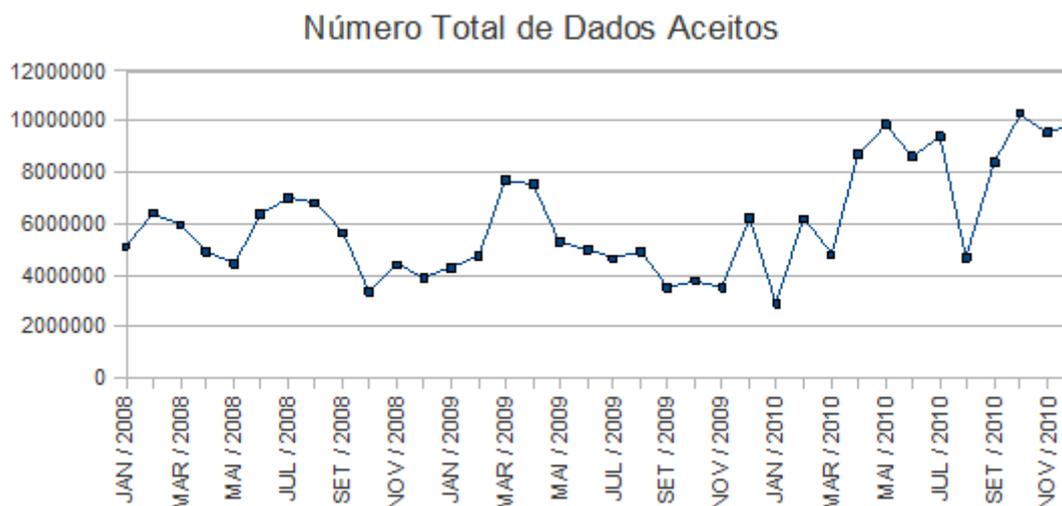
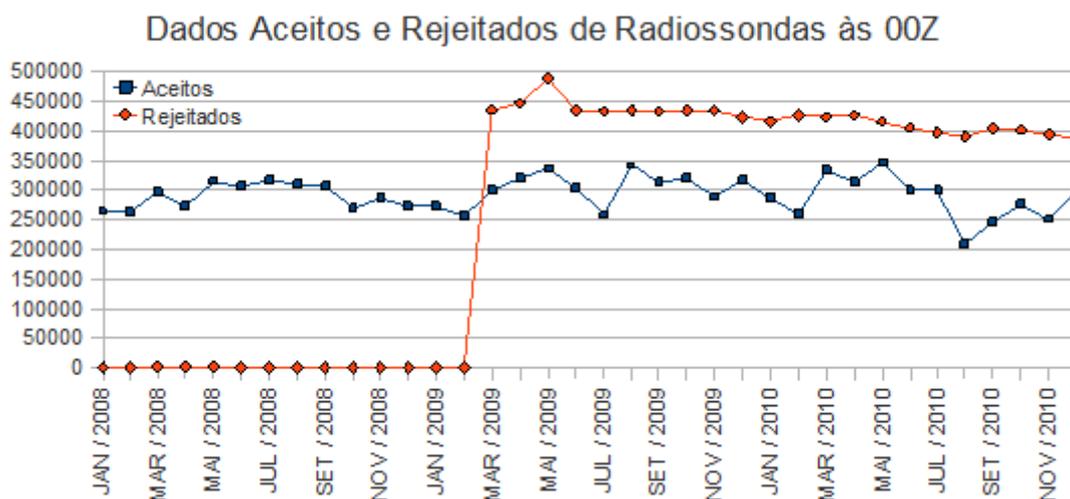


Figura 3. Número total de dados aceitos por mês. O eixo x representa o tempo em meses enquanto o eixo y representa a quantidade de dados aceitos.

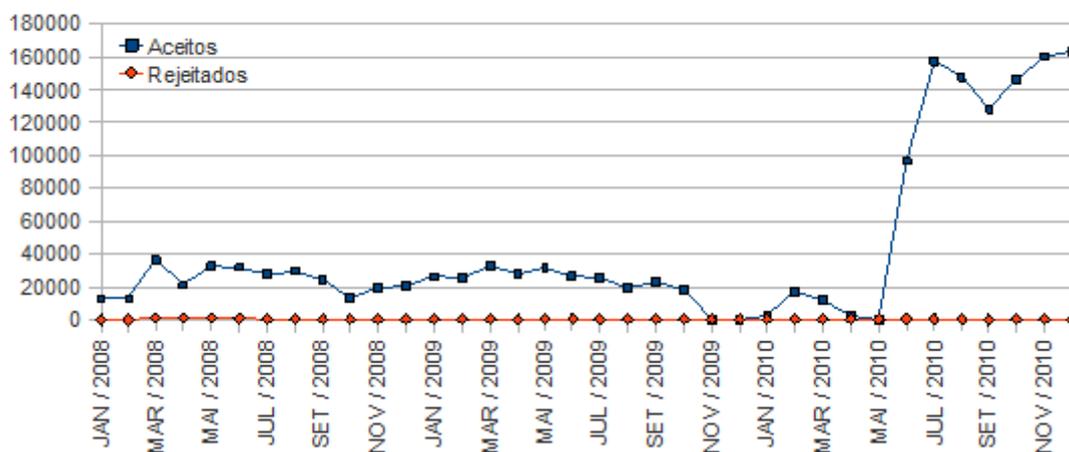
Para avaliar os diferentes tipos de dados, na figura 4a, são mostrados os totais de dados de radiossondas que foram aceitos (azul) e rejeitados (vermelho) de janeiro de 2008 até dezembro de 2010 no horário das 00Z. O aumento significativo no número de dados rejeitados pelo sistema de assimilação de dados observado a partir de março de 2009 é decorrente da mudança na metodologia de contagem dos dados rejeitados, onde passaram a ser considerados como rejeitados aqueles que não eram aceitos antes mesmo de serem submetidos ao sistema de assimilação mais os dados devidamente rejeitados. O mesmo foi observado nos dados de navios, bóias, dados dos satélites AQUA, ATOVS e GOES (CTW), os quais devem ser desconsiderados da análise. Foi observado também que para os dados oriundos do satélite Quik Scatterometer, foram assimilados apenas até agosto de 2008, momento em que o sensor parou de funcionar. Para as 12Z foi observado um aumento significativo no número de dados aceitos do satélite GOES (CTW) a partir de junho de 2010, como mostra a Figura 4b. Esse aumento pode estar relacionado com o decréscimo acentuado do RMSE e BIAS mostrados na Figura 1, o que merece investigação mais a fundo.



(a)



Dados Aceitos e Rejeitados do CTW às 12Z



(b)

Figura 4. Número total de dados aceitos por mês. Em (a) tem-se radiossondas às 00Z e em (b) tem-se o CTW às 12Z. O eixo x representa o tempo em meses enquanto o eixo y representa a quantidade de dados aceitos e rejeitados.

4- Conclusões:

Para avaliação preliminar do desempenho do GPSAS segundo as métricas da OMM, foram utilizados dados de junho de 2007 a dezembro de 2010. Foram calculados erros como *RMSE* e *BIAS* para altura geopotencial em 500hPa, e *RMSVE* do vento em 850hPa e vento em 250hPa. Esses cálculos foram efetuados para o Hemisfério Norte, Hemisfério Sul, Região Tropical e para o globo todo. Para realizar esses cálculos foram utilizados o software GrADS e o software avaliador Meteostat. Também foi analisada a quantidade de dados aceitos e rejeitados por esse sistema de assimilação de dados PSAS.

Os resultados mostraram que o *RMSE* e o *BIAS* das análises tiveram uma queda acentuada seguida de uma estabilização a partir de março de 2010 até o final do período de estudo. As causas para essa diminuição ainda são desconhecidas. Adicionais investigações estão em curso ao utilizar outras métricas e diferentes sistemas de avaliação. Há indícios que isso é decorrente do conjunto de dados assimilados pelo sistema, o que requer uma análise minuciosa os dados assimilados nesse período. Como exemplo disso foi observado um aumento significativo no número de dados aceitos do satélite GOES (CTW) às 12Z a partir de junho de 2010. Embora isso não explique a diminuição dos erros citados anteriormente, visto que essa queda ocorreu três meses antes do aumento de dados. Avaliações futuras devem ser feitas, com a finalidade de descobrir o impacto desse aumento do número de dados de CTW dos demais sistemas de observação na estabilização dos erros do sistema.

De acordo com os resultados obtidos de *RMSE* e *RMSVE* para dados da diferença entre observação e FG, pode-se perceber que grande parte deles apresentaram uma tendência de diminuição de erros. É possível observar que o Hemisfério Sul apresenta uma tendência a diminuição nos erros em todas as variáveis, e para ambos horários. Ele apresenta as 00Z uma diminuição dos erros de *RMSE*, *RMSVE* em 850hPa e *RMSVE* de 250hPa, em 14%, 28% e 10% respectivamente, e para as 12Z os erros caíram 15%, 21% e 8%. No Hemisfério Norte, os resultados mostraram que tanto para as 00Z como para as 12Z os erros de *RMSE*, para a altura geopotencial em 500hPa e *RMSVE* para vento em 850hPa tiveram uma tendência negativa, ou seja, ambos mostraram uma diminuição dos erros, porém, para o *RMSVE* do vento em 250hPa o resultado foi inverso, mostrou uma tendência positiva, o que mostra um aumento do erro em 3% às 00Z e em 4% às 12Z. Na Região Tropical observou-se que o *RMSVE* do vento em 850hPa em ambos os horários e o *RMSVE* do vento em 250hPa as 12Z, apresentaram uma queda dos erros de 23%, 18% e 5% respectivamente. O *RMSE* da altura geopotencial em 500hPa às 12Z, apresentou um aumento de 10%.

5- Referências Bibliográficas



Cressman, George P., 1959: An Operational Objective Analysis System. *Mon. Wea. Rev.*, 87, 367–374.

Gandin, L., S., 1963: Objective analysis of meteorological field. *Gidrometeorologicheskoe Izdate'stvo., Leningrad, U.S.S.R.*, 286pp.

Herdies, D. L. ; Aravéquia, J. A. ; Ferreira, S. H. S. ; Souza, R. V. A. ; Sapucci, L. F. ; Mattos, J. G. F. . A Assimilação de Dados no CPTEC/INPE. *Boletim da Sociedade Brasileira de Meteorologia*, v. 32, p. 57-64, 2008.

Jorba, O., Baldasamo J.M., 2004: Simulación de los campos de viento de La Península Ibérica y El Área geográfica de Catalunya com alta resolución espacial para distintas situaciones meteorológicas típicas. *Barcelona*

.Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering* 82 (1): 35–45. Retrieved 2008-05-03.

Le Dimet, F. X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, 38A, 97-110

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112, 1177–1194.

Souza, R. V. A. ; Herdies, D. L. ; Souza, R. A. F. ; Sapucci, L. F. ; Aravequia, J. A. ; Ferreira, S. H. S. . Tipos de observações e o controle de qualidade utilizado na assimilação de dados no CPTEC/INPE. *Boletim da Sociedade Brasileira de Meteorologia*, v. 32, p. 65-70, 2008.

Stauffer, D.R., and N.L. Seaman, 1994: Multiscale four-dimensional data assimilation. *J. Appl. Meteor.*, 33, 416-434.