

Reanálise regional do CPTEC de 2000-2009 com resolução de 20 km sobre a América do Sul

Luiz Fernando Sapucci, Dirceu Luis Herdies, João Gerd Zell Mattos, Luis Gustavo Gonçalves de Gonçalves, Julio Pablo R. Fernández, Simone S. Tomita Lima, José Antonio Aravequia.

Instituto de Pesquisas Espaciais - INPE, Centro de previsão de Tempo e Estudos Climáticos – CPTEC. Rodovia Presidente Dutra, Km. 40 – Cachoeira Paulista – SP, Brasil.
luiz.sapucci@cptec.inpe.br

ABSTRACT: The so called re-analyses are processes with the objective of creating optimal atmospheric states for specific times where discontinuities resulting from eventual problems with data and deficiencies in the model/data assimilation systems are minimized. Several methodologies to create Reanalysis fields have been employed in the past few years in main global operational meteorological centers. A regional version (over South America) was produced in 2006 at CPTEC-INPE with 40km spatial resolution for the period from 2000 to 2004. This work presents a new generation of regional reanalysis under development in the GDAD (Group for Data Assimilation Developments) at CPTEC-INPE. The main goal of this reanalysis is to make available atmospheric fields over South America at 20Km spatial resolution from 2000 to 2009. A regional data assimilation system PSAS (Physical-space Statistical Analysis System) coupled to the Eta regional model is being employed. The main differences with respect to the previous reanalysis in addition to the methodology for data processing and the schedule for data release are discussed in this work.

Palavras-chave: Reanálise, Assimilação de dados, Modelo Eta, PSAS.

1 – INTRODUÇÃO

O uso de produtos de Previsão Numérica de Tempo (PNT) operacionais em estudos sobre variabilidades de baixa frequência ou mudanças climáticas sobre a América do Sul são de extrema utilidade para caracterizar os estados da atmosfera, porém, alguns cuidados devem ser tomados, pois eventuais problemas de dados ou deficiências dos sistemas de assimilação/previsão utilizados podem introduzir viés nos resultados. Além disso, as devidas correções ou aperfeiçoamentos em tais sistemas, realizados com o decorrer do tempo, geram indesejadas discontinuidades, podendo comprometer seriamente a qualidade de tais produtos. A utilização de processos denominados Reanálise visa contribuir com esse tema em particular, pois a partir do reprocessamento dos dados observados por um relativamente longo período de tempo, assegura-se que tais discontinuidades não sejam introduzidas nos resultados. Para isso é utilizado um sistema de assimilação de dados em conjunto com um de previsão de tempo, no qual suas versões são mantidas inalteradas durante todo o processamento no que diz respeito a modificações tanto na estrutura do código fonte quanto na melhoria dos processos físicos e dinâmicos. Os dados produzidos por este processo vêm sendo utilizados em um amplo espectro de aplicações meteorológicas, tanto para fins de tempo quanto de clima. Exemplos de projetos de reanálise global que devem ser destacados são: (a) A reanálise produzida pelo ECMWF (*European Centre for Medium-Range Weather Forecasts*) com um período de 15 anos iniciando em 1979, denominada ERA-15, na qual foi usado um modelo espectral T106 (Gibson et al. 1997), (b) a reanálise produzida pelo NCEP (*National Centers for Environmental Prediction*) e NCAR (USA) com início em 1948, resolução T62 (Kalnay et al. 1996); (c) A reanálise de 16 anos (com início em 1980) produzida pelo DAO (*Data Assimilation Office*) NASA (USA), com resolução T62; (d) A

reanálise de 40 anos produzida pelo ECMWF (ERA-40) com resolução T159 (Uppala et al. 2006). Além dessas há diversas outras reanálises já finalizadas e outras ainda em andamento. Embora livres das descontinuidades geradas pela modificação do sistema de assimilação/previsão, os produtos gerados na reanálise ainda continuam com as imperfeições inerentes à PNT, e por isso apresentam uma série de problemas e limitações. Com a melhoria dos sistemas de assimilação/previsão novas gerações de reanálises são oferecidas em um processo naturalmente evolutivo.

Com a finalidade de que reanálises de maior resolução espacial sejam viabilizadas, o mesmo processo tem sido desenvolvido com sistemas de PNT regionais. Sobre a América do Sul o Centro de Previsão de Tempos e Estudos Climáticos (CPTEC) disponibilizou em 2006 sua primeira reanálise regional (Aravequia et al. 2007). Nessa versão o sistema de assimilação PSAS (*Phicical-space Statistical Analysis System*) (Da Silva et al. 1995) foi utilizado juntamente com o modelo regional Eta na resolução horizontal de 40 km, para o período de 2000 a 2004. Seguindo a mesma tendência ocorrida nos demais projetos citados acima o sucesso de uso junto á comunidade foi significativo, assim como problemas e limitações foram identificados. Como consequência, uma nova geração de reanálise está em desenvolvimento no GDAD (Grupo de Desenvolvimento em Assimilação de Dados) do CPTEC-INPE, na qual está sendo utilizado o sistema de assimilação regional PSAS acoplado ao modelo Eta, versão *WorkStation*. Nessa versão a resolução horizontal utilizada é de 20 km e o período que a reanálise abrangerá é 2000 a 2009. O objetivo do presente trabalho é apresentar as características dessa nova versão da reanálise do CPTEC-INPE e destacar os pontos mais relevantes associados a sua concepção e execução. Para isso, na próxima seção são apresentados os detalhes do sistema de assimilação/previsão utilizado nessa versão, bem como as melhorias implementadas que a diferenciam da versão anterior. Na seção 3 são apresentadas informações sobre os dados observacionais, as condições de contorno e forçantes do modelo. Na seção 4 são discutidos os recursos computacionais exigidos nesse processo e a estratégia de processamento empregada. Na seção 5 são apresentados os comentários finais e discutido o cronograma para a futura disponibilidade das versões planejadas.

2 – DETALHES DO SISTEMA DE ASSIMILAÇÃO/PREVISÃO UTILIZADO

Para a versão a ser utilizada na reanálise, o acoplamento do sistema de assimilação PSAS no modelo regional Eta, versão *WorkStation* foi realizado em um ambiente de processamento escalar, utilizando um supercomputador com 1100 processadores (UNAI) do CPTEC-INPE. Nessa versão foi realizada uma modificação na metodologia empregada no fornecimento dos campos das variáveis de superfície, como temperatura e umidade do solo, para o modelo de previsão. Com essa modificação, ao invés de utilizar os campos previstos pelo modelo global, como na versão anterior, são utilizadas as previsões de curto prazo do próprio modelo regional. O objetivo foi evitar o impacto negativo ocasionado pela descontinuidade dos campos de temperatura e umidade do solo gerada pelas diferenças nos esquemas de parametrizações de tais variáveis entre os modelos global e regional. O benefício em potencial dessa modificação é um ganho em qualidade nas previsões dos campos de temperatura e umidade do solo e nas variáveis próximas a superfície, importantes para uma correta estimativa da camada limite planetária. Um estudo do impacto na atmosfera, das perturbações nos campos de umidade e temperatura do solo está em desenvolvimento. Como parte das novas implementações, a escolha pela versão Eta *WorkStation* permite a utilização de diferentes parametrizações físicas. Como benefício dessa mudança espera-se obter produtos relacionados com os fenômenos de menor escala espacial mais robustos. Uma outra implementação que merece destaque é a inclusão de atualizações da máscara de neve dinâmica ao invés de se utilizar uma máscara fixa, como em versões anteriores. Embora na

América do Sul, a cobertura de neve se restringe aos pontos mais altos dos Andes, trata-se de um processo cíclico e longo (10 anos), de tal forma que o impacto pode ser significativo. Portanto, é algo que merecendo atenção. Por fim, destaca-se a inclusão de um filtro digital mais eficiente que o filtro digital presente no código original do Eta. Diversos testes foram feitos com diferentes filtros e os resultados confirmaram essa eficiência.

3- ORGANIZAÇÃO, TRATAMENTO E AVALIAÇÃO DAS OBSERVAÇÕES, CONDIÇÕES DE CONTORNO E RESULTADOS.

Na concepção, e posteriormente, no desenvolvimento de uma reanálise uma etapa inquestionavelmente complexa e trabalhosa é a fase inicial, onde a suíte de processamento dos dados é implementada e os arquivos (dados e condições de contorno) a serem utilizados são organizados. Esse processo de implementação da suíte é bastante lento, pois envolve diversos testes para cada uma das novas implementações realizadas, para que a rodada principal e definitiva seja consistente em todo o período, sem que haja novas reinicializações. Por outro lado, como é um processo computacionalmente longo, há a necessidade de que o mesmo não sofra interrupções, que atrasem e/ou comprometam o produto final. A causa mais comum dessas interrupções está relacionada com dados de entrada ou com os dados de contorno. Por essa razão, uma organização eficiente dos dados de forma a antecipar e evitar esses problemas é muito importante. Como essa atividade de reanálise no CPTEC é relativamente nova, diversas ferramentas de organização, tratamento e conferência dos dados estão ainda em desenvolvimento.

Como a reanálise é um processo composto da repetição de muitos ciclos há a necessidade de que certos cuidados sejam tomados para evitar pequenos intervalos de tempo desnecessários que como resultado acarretará em um aumento do tempo final de processamento. Um exemplo disso são algumas modificações na suíte operacional da reanálise, as quais permitem que o PSAS possa ser disparado de forma independente da integração do modelo Eta. Essa modificação reduziu em 20% o tempo estimado de cada rodada. Nessa suíte operacional foram acrescentadas algumas ferramentas para conferência dos arquivos de dados a serem assimilados, bem como dos arquivos de contornos e para a avaliação dos arquivos de saída. Está sendo acrescentado também um processo de alimentação de um banco de dados, no qual os dados assimilados e os rejeitados serão armazenados. O objetivo é monitorar a quantidade dos dados assimilados, bem como os dados rejeitados. A partir desse banco de dados o software GMT (*Generic Mapping Tools*) mostra a distribuição espacial dos dados assimilados, bem como a quantidade de dados aceitos e rejeitados em cada uma das rodadas de forma gráfica e compacta.

Com relação aos dados a serem assimilados, todos os 10 anos de dados foram previamente pré-processados e armazenados. Nessa etapa está incluído o controle de qualidade inicial dos dados e a conversão das variáveis meteorológicas observadas, para as variáveis efetivamente utilizadas pelo PSAS. Para a conclusão desse processo foram utilizados tanto os arquivos gerados no NCEP, como os arquivos gerados a partir do banco de dados meteorológico do CPTEC. Esses últimos foram empregados nos últimos anos do período (de 2007 em diante) uma vez que sobre a América do Sul a quantidade de dados recebidos pelo CPTEC cresceu significativamente nesse período. Após organizados os dados a serem assimilados foram convertidos para o formato ODS (*Observation Data Stream*). Como última fase do pré-processamento, foi realizada a conferência tanto dos dados ODS como de todos os demais arquivos utilizados para rodar a reanálise. Para a realização desse processo de conferência foi criado um procedimento que além de conferir a ocorrência de arquivos faltantes, e se estavam incompletos, também conferia sua sanidade, identificando os arquivos danificados decorrentes de bugs ocorridos durante o processo de escrita no disco.

Para a avaliação sistemática dos resultados gerados está em desenvolvimento uma ferramenta de avaliação de modelos cujo principal algoritmo está em linguagem Fortran, o qual permite o cálculo de forma eficiente de diversas métricas baseadas na comparação com a análise e com campos climatológicos (de Mattos et al. 2010). Algumas implementações recentes tem sido feitas para permitir o processamento de dados no formato grib (formato comprimido de saída da reanálise) e para calcular de forma mais eficiente a correlação de anomalia, utilizando interpolações otimizadas dos campos climatológicos.

4 – ESTRATÉGIA DE PROCESSAMENTO

Um estudo inicial foi feito para determinar o número adequado de processadores utilizados na integração do modelo Eta, bem como para rodar o PSAS, considerando o recurso computacional disponível atualmente no CPTEC. Com a configuração definida, o desempenho total por rodada é de 1 hora, sendo 30 minutos para o PSAS e 30 minutos para os seis processos do Eta. Esse desempenho totaliza um número de 6 dias de reanálise por dia de processamento. Consequentemente, o tempo total de máquina exigido para o processamento dos 10 anos de reanálise é de 1 ano e 8 meses. Até o momento, diversos pequenos períodos de processamento foram realizados como testes, tanto da suíte operacional, como do filtro digital implementado, como outras modificações feitas. O processo principal foi disparado em agosto de 2009, mas teve que ser interrompido para que correções fossem feitas. No atual cronograma de processamento três processos em paralelos estão sendo disparados. A rodada principal inicia-se em janeiro de 2000 e processará até dezembro de 2009. Uma segunda inicia-se em janeiro de 2004 e rodará até dezembro de 2007. A terceira rodada inicia-se em janeiro de 2007 e processará até dezembro de 2009. Essa estratégia de processamento foi adotada com o objetivo de antecipar a disponibilidade da reanálise, ao criar uma versão preliminar formada por trechos processados simultaneamente. A Tabela 1 mostra destacadas em azul a rodada principal e em vermelho o mosaico formado pelos diferentes trechos em processamento nas 3 rodadas. O destaque em vermelho ilustra as partes me montarão a primeira versão a ser disponibilizada. Observa-se nessa tabela, entre o fim de um trecho para o início do próximo, há um ano de intersecção para que o salto na série temporal gerado nesse processo seja minimizado. Procedimento similar foi feito na primeira reanálise do CPTEC (Aravequia et al. 2007). Estudos adicionais sobre o impacto desse procedimento serão realizados e os resultados serão apresentados junto com a divulgação dessa versão preliminar, prevista para meados do próximo ano.

Tabela 1. Processamento simultâneo de três rodadas independentes.
Em vermelho o mosaico da versão preliminar a ser disponibilizada.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Versão preliminar (Julho de 2011)										
Versão final (Março de 2012)										
Rodada 1	X	X	X	X	X	X	X	X	X	X
Rodada 2					X	X	X	X		
Rodada 3								X	X	X

5 – COMENTÁRIOS FINAIS

Como toda reanálise, dois são os desafios a serem vencidos durante a execução desse trabalho: O primeiro, é o espaço em disco, pois se considerado todos os arquivos envolvidos no processamento dos 10 anos de reanálise, o espaço requerido sem redundância é de 5Tb. Cabe salientar que em paralelo ao processamento da reanálise encontra-se em desenvolvimento no grupo, o novo sistema de assimilação, denominado *Local Ensemble*

Transform Kalman Filter, o qual requer também muito espaço em disco. Esse desafio tem sido amenizado com as implementações de soluções envolvendo projetos no GDAD, mas ainda é algo que merece atenção; O segundo e mais preocupante desafio, está relacionado à necessidade de se cumprir um cronograma pré-estabelecido, respeitando as horas-máquina de processamento disponíveis atualmente no CPTEC. Com o crescimento de novos produtos e a melhoria da resolução dos modelos os sistemas de super computadores desse centro encontram-se no limite de sua capacidade. Com a aquisição do novo sistema de super computação em fase de aquisição pelo CPTEC isso será resolvido, e possivelmente a disponibilidade das versões seguindo o cronograma atual poderão ser até mesmo antecipada. Considerando a demanda e esse novo cenário dos recursos computacionais do CPTEC, está em projeto no GDAD o desenvolvimento de uma nova reanálise utilizando outros sistemas de assimilação, períodos mais longos e melhor resolução espacial. Esses futuros projetos serão beneficiados, não apenas pelos processos e ferramentas desenvolvidas ao gerar essa reanálise, mas também com a experiência que os membros do GDAD têm adquirido nessa nova e promissora frente de pesquisa.

6 – REFERÊNCIAS BIBLIOGRÁFICAS

ARAVEQUIA, J. A.; HERDIES, D. L.; e outros. Reanálise regional 2000-2004 sobre a América do Sul com o modelo RPSAS/ETA: Descrição do experimento e dos produtos derivados. Bol. da Soc. Brasileira de Met., v. 32, n. 1, p. 71-77, 2008.

DA SILVA, A. M.; PFAENDTNER, J.; GUO, J.; SIENKIEWICZ, M.; COHN, S. Assessing the effects of data selection with DAO's physical-space statistical analysis system. In: International WMO Symposium, 13-17 Mar. 1995, Tokyo: WMO, 1995. WMO/TD.651.

DE MATTOS, J. G. Z; SAPUCCI, L. F.; HERDIES, D. L.; DE GONÇALVES, L. G. G.; SCAMTeC: Sistema Comunitário de Avaliação de Modelos numéricos de previsão de Tempo e Clima. Anais do XVI Congresso Brasileiro de Meteorologia. Belém, 2010.

GIBSON, J.K., P. KALLBERG, S. UPPALA, A. HERNANDEZ, A. NOMURA AND E. SERRANO, 1997: ERA Description. ECMWF Re-Analysis Project. Report Series, 1.

KALNAY, E. AND KANAMITSU, M. and many others; The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470, 1996

SCHUBERT, S., PARK, C. K., WU, C. Y., HIGGINS, W., and others. 'A multi-year assimilation with the GEOS-1 system: Overview and results'. NASA Tech. Rep. series on global modeling and data assimilation, No. 6. NASA, Greenbelt, Maryland, USA, 1995.

UPPALA, S. M., P. W. KÅLLBERG, A. J. SIMMONS, U. Andrae and many others. The ERA-40 re-analysis. Quarterly Journal of the Royal Meteorological Society Volume 131, Issue 612, Pages 2961-3012, 2006.

AGRADECIMENTOS

Os autores desse trabalho agradecem ao Dr. Nelson Jesus Ferreira, coordenador dos projetos ADAPT e PROTIM 3, pelo apoio técnico e à Financiadora de Estudos e Projetos (FINEP) pelo apoio financeiro a esses projetos e consequentemente a essa pesquisa (Convênios de N° 01.06.1120.00 e .05.0912.00, respectivamente).