



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/06.20.12.39-TDI

ASSIMILAÇÃO DE DADOS GLOBAL HÍBRIDA POR CONJUNTO-VARIACIONAL NO CPTEC

Carlos Frederico Bastarz

Tese de Doutorado do Curso de Pós-Graduação em Meteorologia, orientada pelo Dr. Dirceu Luis Herdies, aprovada em 18 de julho de 2017.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3P5MHFL>>

INPE
São José dos Campos
2017

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

E-mail: pubtc@inpe.br

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)

Membros:

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas
(CEA)

Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)

Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial
(ETE)

Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos
Climáticos (CPT)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Duca Barbedo - Serviço de Informação e Documentação
(SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2017/06.20.12.39-TDI

ASSIMILAÇÃO DE DADOS GLOBAL HÍBRIDA POR CONJUNTO-VARIACIONAL NO CPTEC

Carlos Frederico Bastarz

Tese de Doutorado do Curso de Pós-Graduação em Meteorologia, orientada pelo Dr. Dirceu Luis Herdies, aprovada em 18 de julho de 2017.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3P5MHFL>>

INPE
São José dos Campos
2017

Dados Internacionais de Catalogação na Publicação (CIP)

Bastarz, Carlos Frederico.

B295a Assimilação de dados global híbrida por conjunto-variacional no CPTEC / Carlos Frederico Bastarz. – São José dos Campos : INPE, 2017.

xxvi + 49 p. ; (sid.inpe.br/mtc-m21b/2017/06.20.12.39-TDI)

Tese (Doutorado em Meteorologia) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2017.

Orientadores : Dr. Dirceu Luis Herdies.

1. Assimilação de dados. 2. Assimilação variacional tridimensional. 3. Filtro de Kalman por conjuntos. 4. Asssimilação de dados híbrida. 5. Matriz de covariâncias. I.Título.

CDU 551.509.313.3



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

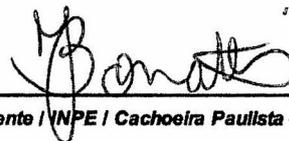
This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Carlos Frederico Bastarz**

"ASSIMILAÇÃO DE DADOS GLOBAL HÍBRIDA POR CONJUNTO-VARIACIONAL NO CPTEC"

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Doutor(a)** em
Meteorologia

Dr. José Paulo Bonatti



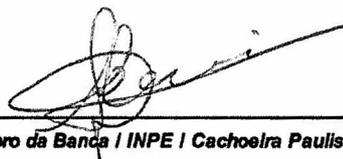
Presidente | INPE | Cachoeira Paulista - SP

Dr. Dirceu Luis Herdies



Orientador(a) | INPE | Cachoeira Paulista - SP

Dr. Luiz Fernando Sapucci



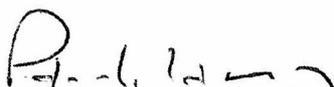
Membro da Banca | INPE | Cachoeira Paulista - SP

Dr. Clemente Augusto Souza Tanajura



Convidado(a) | UFBA | Salvador - BA

Dr. Pedro Leite da Silva Dias



Convidado(a) | USP | São Paulo - SP

Este trabalho foi aprovado por:

maioria simples

unanimidade

São José dos Campos, 18 de julho de 2017

“... When you are studying any matter, or considering any philosophy, ask yourself only what are the facts and what is the truth that the facts bear out. Never let yourself be diverted either by what you wish to believe, or by what you think would have beneficent social effects if it were believed. But look only, and solely, at what are the facts.”

BERTRAND RUSSELL
em “Face to Face with John Freeman”. BBC, 1959.

À memória de meus pais.

AGRADECIMENTOS

Em primeiro lugar, a Deus pela minha vida e pela família que me proveu.

A meus pais, cuja memória me é cara e é sempre lembrada em toda a minha trajetória.

O desenvolvimento deste trabalho não seria possível sem a colaboração de muitas pessoas, as quais gostaria de expressar meus sinceros agradecimentos.

Aos meus companheiros de trabalho dentro do Grupo de Desenvoltimentos em Assimilação de Dados, os quais faço questão de nomear: Eduardo Khamis, Fábio Diniz, Gustavo Gonçalves, Helena Barbieri, Lucas Amarante e demais companheiros que passaram pelo grupo e que seguiram outros caminhos. Aos meus amigos João Gerd e Bruna Silveira pela amizade, cumplicidade e constante incentivo. Por diversas vezes me auxiliaram nos vários desenvolvimentos, entendimentos e discussões ao longo de meu trabalho.

Aos funcionários e colaboradores do CPTEC/INPE da área administrativa e do programa de pós-graduação, pelo auxílio na organização e atenção aos prazos e etapas envolvidas durante o desenvolvimento do programa de doutoramento.

A minha família, especialmente minha irmã Beatriz Bastarz, pelo incentivo e atenção dados em diversos momentos.

A minha Helena Cachanhuk, pela paciência, amor, compreensão, amizade, companheirismo, incentivo, conselhos e encorajamentos. Sem a sua presença em minha vida, eu não saberia para onde caminhar. Obrigado por fazer parte da minha vida e me ajudar a melhorar como pessoa a cada dia.

Ao Silvio Nilo pela confiança em mim depositada e paciência na conclusão deste trabalho. Ao Luiz Sapucci pela confiança, incentivo, interesse e também pelos conselhos em diversos momentos durante o desenvolvimento do trabalho.

Ao meu orientador Dirceu Herdies, pela oportunidade e incentivo. Agradeço por fazer parte de minha formação profissional desde o meu começo no CPTEC como aluno de mestrado e agora, como funcionário do INPE, instituição da qual tenho muito orgulho de fazer parte.

Ao Ricardo Todling do GMAO/NASA pela boa vontade em me receber no GMAO durante os 9 meses em que estive lá. Agradeço pela paciência, orientação, entusiasmo

e boa vontade para comigo. Ao Stephen Cohn, também do GMAO/NASA, pela acolhida, paciência e orientação. Foi um grande privilégio poder aprender com um grande matemático, do qual pude observar e absorver características que agora fazem parte do meu pensamento matemático e crítico.

Aos demais colegas com os quais tive contato durante a realização do programa de doutoramento e que direta ou indiretamente me auxiliaram no trabalho. Meus sinceros agradecimentos.

Ao INPE pela concessão da bolsa institucional (via CAPES), no período de 2012 a 2013.

Ao CNPq pela bolsa de doutorado no período de 2013 a 2014 (processo número 140938/2014-1).

À CAPES pela concessão da bolsa para a realização do doutorado sanduíche no exterior, no período de 2014 a 2015 (processo número BEX 99999.008036/2014-04).

RESUMO

Esta tese de doutorado é dedicada a estudar a determinação e a aplicação das covariâncias dos erros de previsão, por meio da aplicação de um sistema híbrido 3DVar de assimilação de dados. A característica híbrida deste sistema de assimilação de dados refere-se a forma da matriz de covariâncias. Uma matriz de covariâncias híbrida é, no contexto deste trabalho, uma combinação linear entre uma matriz de covariâncias estática (i.e., fixa no tempo) e covariâncias obtidas a partir de um sistema de filtro de Kalman por conjunto (e.g., *Ensemble Kalman Filter*). A combinação entre estas duas espécies de covariâncias tem o benefício de fazer com a matriz de covariâncias estática seja capaz de enxergar, com certa limitação, as variações diárias do fluxo atmosférico, uma característica antes limitada a sistemas computacionalmente mais custosos e complexos. Atualmente, sistemas híbridos tem sido aplicados em centros operacionais de Previsão Numérica de Tempo sob diferentes metodologias. A metodologia aplicada neste trabalho para adicionar as covariâncias dos erros de um conjunto de previsões a matriz de covariâncias estática, utiliza uma extensão da variável de controle do sistema de assimilação de dados variacional. A partir desta metodologia, foi estabelecido um ciclo de assimilação de dados em que análises e previsões são geradas em uma única resolução (TQ0062L028), gerando previsões para até 5 dias. O sistema foi experimentado por durante dois meses do verão austral de 2013, testando-se diferentes porcentagens de contribuições das covariâncias do conjunto. Os resultados obtidos mostram que com a matriz de covariâncias híbrida, as análises produzidas pelo sistema permitem que o modelo de previsão utilizado produza previsões consistentes, comparáveis em termos de habilidade de previsão às previsões obtidas com o sistema 3DVar puro e de referência. Em diversas situações, na avaliação da habilidade da previsão de algumas variáveis e em regiões como o Hemisfério Sul e a região tropical, as análises e previsões provenientes do sistema híbrido foram melhores do que as análises e previsões produzidas pelo sistema variacional puro.

Palavras-chave: Assimilação de Dados. Assimilação Variacional Tridimensional. Filtro de Kalman por Conjuntos. Assimilação de Dados Híbrida. Matriz de Covariâncias. Previsão Numérica de Tempo.

GLOBAL HYBRID ENSEMBLE VARIATIONAL DATA ASSIMILATION AT CPTEC

ABSTRACT

This thesis is dedicated to the study of background error covariances by means of the application of a hybrid 3DVar data assimilation system. The hybrid characteristic of this system refers to the combination of a static (i.e., fixed in time) background error covariance matrix and covariances drawn from an ensemble of backgrounds through an ensemble data assimilation system (e.g., Ensemble Kalman Filter). The combination between the two kinds of covariances has the benefit of enabling the static covariances to account for the day-to-day variations of the background flow. Recently, hybrid systems have been in use in several Numerical Weather Prediction Centers under different methodologies. The methodology used in this work in order to add the ensemble covariances to the static part, is through an extension of the control variable. From the application of this methodology, a data assimilation cycle was established at a single resolution (TQ0062L028) with forecasts up to 5 days. The system was exercised during two months of the 2013 austral summer, where different amounts of the ensemble contribution to the static covariance have been experienced. The results shows that with the hybrid background error covariance matrix, the system was able to produce consistent forecasts. In terms of the analysis and forecasts skill, for some regions, like the South Hemisphere and the Tropics, the analysis and forecasts for some variables were better than the pure variational analysis.

Keywords: Data Assimilation. 3D Variational Assimilation. Ensemble Kalman Filter. Hybrid Data Assimilation. Covariance Matrix. Numerical Weather Prediction.

LISTA DE FIGURAS

	<u>Pág.</u>
3.1 Morfologia idealizada da matriz de covariâncias dos erros de previsão (B).	40
3.2 Distribuição das amplitudes calculadas para a matriz B na resolução TQ0299L064.	43
3.3 Idem Figura 3.2, para a velocidade potencial ($\chi \times 10^{-6}$).	45
3.4 Idem Figura 3.2, para a temperatura (T).	46
3.5 Idem Figura 3.2, para a umidade (q).	47
3.6 Idem Figura 3.2, para a pressão em superfície ($ps, \times 10^2$).	48
3.7 Idem Figura 3.2, para o ozônio ($oz \times 10^7$) e o conteúdo de água ($cw \times 10^4$).	49
3.8 Diagramas de dispersão entre as amplitudes representadas nas matrizes de covariâncias B CPTEC e NCEP.	52
3.9 Aspectos dos incrementos de análise utilizando a nova matriz de covariâncias do CPTEC.	54
3.10 Idem Figura 3.9, para a matriz de covariâncias do NCEP.	56
4.1 Exemplos das curvas de minimização da função custo variacional para os experimentos realizados (20130101).	65
4.2 Diagrama esquemático do sistema híbrido 3DVar.	71
5.1 Viés ponderado da inovação dos conjuntos de análises e previsões dos experimentos, válido para as 00Z.	81
5.2 Estatísticas de inovação dos conjuntos de análises dos experimentos, válido para as 00Z.	85
5.3 Correlação de Anomalia com teste de significância t-Student (95%) das previsões para até 120 horas para os horários das 00 e 12Z, sobre as regiões Hemisfério Norte (HN), Trópicos (TR) e Hemisfério Sul (HS).	88
5.4 Idem Figura 5.3, mas para as regiões Global (GL) e América do Sul (AS).	90
5.5 Médias espaciais das previsões de precipitação de 24 horas (em mm/dia) válido para Janeiro de 2013 às 12Z.	92
5.6 Médias espaciais das previsões de precipitação de 24 horas (em mm/dia) válido para Janeiro de 2013 às 12Z, para a região GL (Global).	94
5.7 Idem Figura 5.6, para as regiões (a) precipitação para todo o Hemisfério Norte (HN); (b) Trópicos (TR); (c) Hemisfério Sul (HS); (d) América do Sul (AS).	95
5.8 Séries temporais do perfil vertical da componente zonal (u) do vento horizontal durante Janeiro de 2013, para o experimentos realizados.	100

5.9	Imagens infravermelho dos satélites GOES 12 ($10.2 - 11.2 \mu m$) em composição com o METEOSAT 9 ($10.8 \mu m$), entre os dias 09 e 14 de Janeiro de 2013 às 10BRT (12UTC) mostrando o continente Sul-Americano, parte do continente Africano, partes do Oceano Pacífico e o oceano Atlântico Sul.	102
5.10	Idem Figura 5.9, mas entre os dias 26 e 31 de Janeiro de 2013.	103
5.11	Média da precipitação MERGE para Janeiro de 2013 às 12UTC (em mm/dia).	105
5.12	Médias da divergência de umidade (D_q - sombreado, $\times 10^{-7} Kgs^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 850 hPa para o período de 09 a 14 de Janeiro de 2013, dos experimentos REF, 3DVar e EnKF75. . .	108
5.13	Idem Figura 5.12, mas para a divergência do vento horizontal (D_w - sombreado, $\times 10^{-5} s^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 200 hPa.	110
5.14	Idem Figura 5.12, mas para a velocidade vertical (Ω - sombreado, $\times 10^{-1} Pas^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 500 hPa.	111
5.15	Análises dos experimentos REF e 3DVar para a umidade em 850 hPa no dia 2013011200 e diferenças entre as análises e as suas respectivas previsões para até 120 horas.	113
5.16	Idem Figura 5.15, mas para a análise do experimento EnKF75.	114

LISTA DE TABELAS

	<u>Pág.</u>
2.1 Principais opções das versões MCGAv4 e BAMv0 do modelo de circulação geral do CPTEC.	22
2.2 Observações tipicamente assimiladas pelo sistema GSI no CPTEC. . . .	33
3.1 Valores máximos e mínimos das amplitudes das versões das matrizes de covariâncias do CPTEC na resolução TQ0299L064 a partir dos pares de previsões de 48 e 24 horas do modelo MCGAv4.	49
3.2 Idem Tabela 3.1, para os comprimentos de escala horizontais.	49
3.3 Idem Tabela 3.1, para os comprimentos de escala verticais.	50
3.4 Coeficiente de Correlação de Pearson (r) e Erro Quadrático (r^2) para as amplitudes de ψ e χ , entre as matrizes B CPTEC e NCEP (TQ0299L064) sobre as regiões HN, TR e HS e para os horários AllZ (todos) e 00Z, 06, 12Z e 18Z. Em vermelho, estão destacados os valores de correlação menores do que 0,9.	51
3.5 Idem Tabela 3.4, para as variáveis temperatura (T) e umidade (q).	51
3.6 Parâmetros de configuração utilizados no sistema GSI para a aplicação das matrizes de covariâncias do CPTEC e do NCEP.	53
4.1 Experimentos com o sistema híbrido 3DVar.	74
5.1 Médias espaciais sobre a região Global (GL) para a referência (GPCPv2.2) e os experimentos realizados das previsões de 24 horas da precipitação total às 12Z (em mm/mês). Em azul, estão indicados os experimentos cujas médias são mais próximas a média de precipitação da referência. . . .	93
5.2 Médias espaciais, desvios-padrão e erro médio absoluto (MAE, em relação ao GPCP) das previsões de 24 horas da precipitação total às 12Z (em mm/mês) para as regiões Global (GL), Hemisfério Norte (HN), Trópicos (TR), Hemisfério Sul (HS) e América do Sul (AS). Em azul, estão indicados os experimentos cujas médias são mais próximas a média de precipitação do GPCPv1.2 (considerando o desvio-padrão do GPCP). Em verde, estão indicados os valores de MAE mais baixos em relação ao GPCP.	97
5.3 Períodos de atividade e duração das ZCAS em Janeiro de 2013.	98

LISTA DE ABREVIATURAS E SIGLAS

3DVar	–	3-Dimensional Variational
4DVar	–	4-Dimensional Variational
4DEnsVar	–	4-Dimensional Ensemble-Variational
AIRS	–	Atmospheric InfraRed Sounder
AMSU/A	–	Advanced Microwave Sounding Unit A
BAM	–	Brazilian Atmospheric Model
BoM	–	Bureau of Meteorology
BUFR	–	Binary Universal Form for the Representation of meteorological data
CODAS	–	CPTEC Ocean Data Assimilation System
CRPS	–	Continuous Rank Probability Skill Score
CPTEC	–	Centro de Previsão de Tempo e Estudos Climáticos
CRTM	–	Community Radiative Transfer Model
CS	–	Correções Sucessivas
DTC	–	Developmental Testbed Center
DWD	–	Deutsche Wetterdienst
E4DVAR	–	Ensemble 4-Dimensional Variational
ECMWF	–	European Centre for Medium Range Weather Forecasts
EKF	–	Extended Kalman Filter
EnKF	–	Ensemble Kalman Filter
EnSRF	–	Ensemble Square Root Filter
ETKF	–	Ensemble Transform Kalman Filter
HNEnKF	–	Hybrid Nudging Ensemble Kalman Filter
FCOR	–	Função de Corrente
FK	–	Filtro de Kalman
G3DVAR	–	Global 3-Dimensional Variational
GFS	–	Global Forecast System
GMAO	–	Global Modelling and Assimilation Office
GPCP	–	Global Precipitation Climatology Project
GPM	–	Global Precipitation Measurement
GPS	–	Global Positioning System
GPS-RO	–	Global Positioning System Radio Occultation
GSFC	–	Goddard Space Flight Center
GSI	–	Gridpoint Statistical Interpolation
GTS	–	Global Telecommunication System
HN	–	Hemisfério Norte
HRS4	–	Hyperspectral Research Satellite 4
HS	–	Hemisfério Sul
IASI	–	Infrared Atmospheric Sounding Interferometer
IBM	–	International Business Machines
INPE	–	Instituto Nacional de Pesquisas Espaciais

IO	–	Interpolação Ótima
JCSDA	–	Joint Center for Satellite Data Assimilation
JMA	–	Japan Meteorological Agency
JNWP	–	Joint Numerical Weather Prediction Unit
KMA	–	Korea Meteorological Administration
LAF	–	Lagged Average Forecasting
LEKF	–	Local Ensemble Kalman Filter
LETKF	–	Local Ensemble Transform Kalman Filter
MAE	–	Mean Absolute Error
MCGA	–	Modelo de Circulação Geral da Atmosfera
MERGE	–	Precipitação observada do satélite TRMM combinada com dados de estações de superfície
MetOffice	–	Meteorological Office
MHS	–	Microwave Humidity Sounder
MLEF	–	Maximum Likelihood Ensemble Filter
MOM4	–	Modular Ocean Model 4
MSC	–	Meteorological Service of Canada
NASA	–	National Aeronautics and Space Administration
NCAR	–	National Center for Atmospheric Research
NCEP	–	National Centers for Environmental Predictions
NESDIS	–	National Environmental Satellite and Data Information Service
NMC	–	National Meteorological Center
NMCRWF	–	National Centre for Medium Range Weather Forecasting
NMMB	–	Nonhydrostatic Multiscale Model on the B-grid
NOAA	–	National Oceanic and Atmospheric Administration
OmA	–	Observation minus Analysis
OmF	–	Observation minus Forecast
OMM	–	Organização Meteorológica Mundial
OSE	–	Observing System Experiments
PNT	–	Previsão Numérica de Tempo
PrepBUFR	–	Pre-processed Binary Universal Form for the Representation of meteorological data
RHMC	–	Hydrometeorological Center of Russia
RO-GPS	–	Radio Occultation-Global Positioning System
ROPP	–	Radio Occultation Processing Package
SLAF	–	Scaled Lagged Average Forecast
SPEEDY	–	Simplified Parameterizations, primitive-Equation DYNAMICS
SSI	–	Spectral Statistical Interpolation
SSiB	–	Simplified SiB (Simple Biosphere) Model
TR	–	Trópicos
TRMM	–	Tropical Rainfall Measuring Mission
UMD	–	University of Maryland

- WMO – World Meteorological Organization
- WRF – Weather Research and Forecasting model
- WRFDA – Weather Research and Forecasting model Data Assimilation
- ZCAS – Zona de Convergência do Atlântico Sul

LISTA DE SÍMBOLOS

J	– Função Custo variacional
\mathbf{x}	– Vetor de estado a ser analisado (variável de controle)
\mathbf{x}^b	– Vetor de estado do modelo
\mathbf{B}, \mathbf{B}_c	– Matriz de covariâncias (estática) dos erros de previsão variacional
\mathbf{y}^o	– Vetor observação
$\delta\mathbf{x}, \delta\mathbf{x}'$	– Vetores incremento de análise
\mathbf{R}	– Matriz de covariância dos erros de observação
H	– Operador observação não linear
S	– Espalhamento do conjunto de previsões
\mathbf{H}	– Operador observação linear
\mathbf{x}^a	– Vetor análise
\mathbf{x}_k^a	– k -ésimo membro do conjunto de análise
$\bar{\mathbf{x}}^a$	– Vetor com a média das variáveis do conjunto de análise
\mathbf{X}^a	– Matriz de perturbação do conjunto de análise
K	– Tamanho do conjunto
\mathbf{X}^b	– Matriz de perturbação do conjunto de previsões
\bar{x}^b	– Média do conjunto de previsões
\mathbf{x}_k^e	– k -ésimo membro do conjunto de previsões
$\tilde{\mathbf{K}}$	– Matriz ganho no espaço das perturbações do conjunto
\mathbf{P}^a	– Matriz de covariância dos erros do conjunto de análise
\mathbf{P}^b	– Matriz de covariância dos erros do conjunto de previsão
\mathbf{B}_{IO}	– Matriz de covariâncias dos erros de previsão do método de Interpolação Ótima
\circ	– Produto Schur
\mathbf{B}_{4dvar}	– Matriz de covariâncias dos erros de previsão do método 4DVar
\mathbf{B}_e	– Matriz de covariâncias dos erros do conjunto de previsões
p	– Pressão atmosférica
\propto	– Proporcional
D	– Divergência
D_q	– Divergência de Umidade
D_w	– Divergência do Vento Horizontal
ζ	– Vorticidade
ψ	– Função de Corrente
χ	– Velocidade Potencial
T	– Temperatura
q	– Umidade
oz	– Ozônio
cw	– Conteúdo de Água de Nuvens
ps	– Pressão em Superfície
tsm	– Temperatura da Superfície do Mar
u	– Componente Zonal do Vento Horizontal

v	–	Componente Meridional do Vento Horizontal
w	–	Vento Horizontal
r	–	Coefficiente de Correlação de Pearson
r^2	–	Erro Quadrático
μ	–	Média (espacial ou temporal)
σ	–	Desvio-Padrão/Coordenada vertical
δ	–	Viés
\mathbf{x}'	–	Vetor incremento de análise híbrido
\mathbf{a}	–	Vetor extensão da variável de controle do conjunto
\mathbf{a}_k	–	k -ésima extensão da variável de controle do conjunto
\mathbf{A}	–	Matriz bloco diagonal de correlação espacial associada a \mathbf{a}
\mathbf{C}	–	Matriz de correlação espacial horizontal associada \mathbf{a}_k
J_{3dvar}	–	Parcela da função custo referente ao 3DVar
J_e	–	Parcela da função custo referente ao conjunto
J_o	–	Parcela da função custo referente as observações
L	–	Comprimento de escala
V_ψ	–	Variância da função de corrente
V_ζ	–	Variância da vorticidade
V^1, V^2	–	Desvios-padrão
V^2	–	
$\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z$	–	Aplicação do filtro recursivo em x, y e z
$\mathbf{B}^1, \mathbf{B}^2$	–	Aplicação do filtro recursivo nas escalas horizontal e vertical
$\alpha, \lambda, \rho, \alpha_c^2, \alpha_e^2, \alpha_1, \alpha_2$	–	Coefficientes (escalares)
β_t	–	Viés do modelo
ω^f	–	Estado previsto do modelo
ω^t	–	Estado real do modelo
ϵ^t	–	Erro aleatório
$\langle \epsilon^t \rangle$	–	Esperança do erro aleatório
Ω	–	Velocidade vertical

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1 Motivação	3
1.2 Revisão Bibliográfica	7
1.2.1 Análise Objetiva	7
1.2.2 Assimilação de Dados por Conjuntos	10
1.2.3 Assimilação de Dados Híbrida	11
1.2.4 Modelagem de Covariâncias	16
1.3 Objetivos	18
1.3.1 Objetivo Geral	19
1.3.2 Objetivos Específicos	19
1.4 Estrutura do Documento	19
2 DADOS E METODOLOGIA	21
2.1 Modelo de Circulação Geral do CPTEC	21
2.1.1 Descrição e opções do Modelo de Circulação Geral	21
2.2 <i>Gridpoint Statistical Interpolation</i> (GSI)	23
2.2.1 Sistema Observacional do GSI	24
2.2.2 <i>3-Dimensional Variational Assimilation</i> (3DVar)	25
2.2.3 <i>Ensemble Kalman Filter</i> (EnKF)	26
2.2.4 <i>Ensemble Square Root Filter</i> (EnSRF)	30
2.3 Dados	32
3 MATRIZ DE COVARIÂNCIAS DOS ERROS DE PREVISÃO .	35
3.1 Importância da Matriz B	36
3.2 Método <i>National Modeling Center</i> (NMC)	37
3.3 Características Principais e Diagnósticos	42
3.3.1 Correlações entre B CPTEC X NCEP	50
3.4 Aspecto do Incremento de Análise	53
4 ASSIMILAÇÃO DE DADOS HÍBRIDA ENSEMBLE-VARIACIONAL GLOBAL	57
4.1 Incorporação das Covariâncias do filtro de Kalman por conjunto na Estrutura Variacional do GSI	57

4.2	Ajustes dos Parâmetros Principais	61
4.2.1	Matriz de Covariâncias Estática (B)	61
4.2.2	Assimilação de Dados Variacional	64
4.2.3	Assimilação de Dados por Conjuntos	66
4.3	Ciclo de Assimilação de Dados do Sistema Híbrido Ensemble-Variacional	69
4.4	Exemplo da Análise do Campo de Vento Horizontal	72
4.5	Experimentos com o Sistema Híbrido 3DVar	74
5	RESULTADOS	77
5.1	Assimilação de Dados Híbrida 3DVar	77
5.1.1	Viés Ponderado dos Conjuntos de Análises	77
5.1.2	Inovação dos Conjuntos de Análises	82
5.1.3	Habilidade das Previsões até 120 horas	86
5.1.4	Avaliação das Previsões de 24 horas de Precipitação	91
5.2	Estudo de Caso	98
6	CONCLUSÕES	115
6.1	Sugestões para Trabalhos Futuros	121
	REFERÊNCIAS BIBLIOGRÁFICAS	123
	APÊNDICE A - Derivação da Equação de Análise a partir da Função Custo Variacional Tridimensional	137
	APÊNDICE B - Pré-condicionamento da Função Custo Variacional com base em B Completa	143
	APÊNDICE C - Filtro Recursivo 1D - Grau 1	147
	APÊNDICE D - Filtro Recursivo 1D - Grau 2	153
	ANEXO A - Artigo Científico I	159
	ANEXO B - Artigo Científico II.	195

1 INTRODUÇÃO

Na década de 1950 Jules Charney indicou que mesmo com o contínuo melhoramento dos modelos de Previsão Numérica de Tempo (PNT), ainda sim haveria um limite para a habilidade dos modelos em prever os estados futuros da atmosfera. Este limite estaria então, relacionado a inevitáveis erros e imprecisões de modelagem e à erros nas condições iniciais dos modelos. Durante a década seguinte, Edward Lorenz mostrou que por mais perfeitos que fossem os modelos numéricos de PNT, sempre haveria um limite para a previsão de fenômenos atmosféricos. Ele se referia ao resultados de experimentos com um modelo atmosférico simples de baixa ordem (12 variáveis) em que dadas duas condições iniciais ligeiramente diferentes, seus resultados divergiam drasticamente após duas semanas, sendo tão diferentes quanto duas previsões aleatórias fornecidas pelo modelo. Estes experimentos ficaram conhecidos como “experimentos gêmeos”, em que dado o mesmo modelo e duas condições iniciais ligeiramente diferentes, após a evolução temporal das equações do modelo, os resultados encontrados seriam completamente distintos. Este tipo de comportamento encontrado nos experimentos gêmeos, foi atribuído por Lorenz as instabilidades inerentes a qualquer sistema de natureza dinâmica não linear, tal como a atmosfera. Originalmente, a ideia de Lorenz era mostrar que previsões estatísticas não tinham capacidade de conferir precisão as previsões de modelos com dinâmica não linear, mas que a previsão numérica de tempo tinha potencial suficiente para ser preditiva através de métodos estatísticos. Através desta experiência, Lorenz descobriu o teorema fundamental da previsibilidade, estabelecendo que “sistemas instáveis tem um limite de previsibilidade finito (enquanto a estabilidade do fluxo se mantém), enquanto que sistemas estáveis (periódicos ou estacionários, por manterem a estabilidade de seus fluxos), tem um limite de previsibilidade infinito”. Desde as descobertas de Lorenz, a ideia de que a atmosfera se comporta como um sistema dinâmico não linear, sendo ela previsível até alguns poucos dias (e depois disso seu comportamento é caótico, imprevisível), fez com que a ideia de determinação do estado do tempo mudasse. Esta mudança se refletiu na forma como as previsões numéricas de tempo são realizadas, deixando de serem simplesmente determinísticas, passando a serem probabilísticas (ou estocásticas), permitindo-se explorar a natureza caótica da atmosfera e conseqüentemente a extensão da validade da PNT operacional.

O período de duas semanas observado por Lorenz é o limite prático da previsão numérica de tempo e permanece válido ainda hoje. Naquela época, pouco mais de uma década após a introdução do computador eletrônico para a integração numérica dos modelos de previsão de tempo, a precisão e a habilidade das previsões não

era tão alta quando dois dias, ou seja, a validade das previsões daquela época era bastante curta e não alcançava 72 horas. Apesar deste limite encontrado por Lorenz, iniciativas (e.g., *The Subseasonal to Seasonal (S2S) Prediction Project Database*) tem sido colocadas em prática para que seja possível realizar previsões entre as escalas subsazonal e climática (VITART et al., 2017).

A previsibilidade da atmosfera é extremamente dependente da forma como ela mesma evolui, pois em alguns dias a previsão pode se comportar muito bem permanecendo válida por vários dias em sequência, mas em questão de poucas horas, a previsão para o dia seguinte pode ser comprometida. Este comportamento da atmosfera, já apontado por Lorenz como sendo caótico, fez com que fosse necessário também, se considerar a natureza estocástico-dinâmica da atmosfera.

A assimilação de dados, como um conjunto de técnicas para a determinação das análises dos modelos de PNT, tem mostrado entre seus principais avanços, alternativas para a modelagem e a representação das covariâncias dos erros de previsão. Entre estas alternativas, destacam-se a integração entre sistemas de assimilação de dados com diferentes abordagens, levando ao que se denomina de sistemas híbridos, com o objetivo de se combinar as características de diferentes sistemas de assimilação para representar de forma mais adequada a evolução espaço-temporal dos erros das previsões dos modelos, necessárias para a determinação das análises. Esta evolução espaço-temporal, é o que se denomina por “erros do dia” e representam uma das principais fontes de incertezas do processo de modelagem como um todo.

Fontes de incerteza são uma característica intrínseca a qualquer sistema dinâmico que apresente comportamento aleatório e relações não lineares entre as suas variáveis de estado. A atmosfera, por exemplo, é um sistema caótico em que diversos parâmetros (como a umidade e a temperatura) podem ser determinados por leis de conservação (ie., conservação de massa e energia) que regem a sua evolução temporal. Outros parâmetros, como os ventos (que são determinados pela lei de conservação do *momentum*) podem ser modelados e a sua dinâmica em larga escala é bem determinada; entretanto, quando se variam as escalas espaciais e temporais destas quantidades, as leis que antes eram aplicadas para a sua determinação, passam a ser incompletas. Isto significa que relações não lineares e o comportamento aleatório destas quantidades passam a ter alguma importância e a sua representação torna-se necessária. Um exemplo deste tipo de fenômeno é o movimento Browniano observado pelo botânico Robert Brown na segunda década do século XIX e descrito pelo físico Albert Einstein no início do século XX.

Dada a complexidade da atmosfera, a modelagem dos processos físicos e a sua representação pelos modelos de PNT é um grande desafio. Muitas fontes de incertezas e aproximações fazem com que a natureza caótica da atmosfera tenha que ser modelada, muitas vezes, utilizando-se métodos probabilísticos ao invés de se tentar alcançar uma solução determinística. Logo, uma aproximação para esta necessidade, envolve um conjunto de soluções, as quais podem ser obtidas mediante a imposição de condições iniciais e de contorno diferentes ou mesmo perturbações aleatórias. Como resultado, obtém-se uma pequena amostra de um enorme conjunto (talvez infinito) de possíveis soluções. Uma segunda aproximação para se abordar esta característica caótica dos sistemas dinâmicos (tal como a atmosfera), é a combinação de diferentes modelos (que não necessariamente possuem diferentes condições iniciais) para tentar gerar uma solução que seja uma amostra mais representativa do universo das possíveis soluções. Neste contexto, na assimilação de dados, os sistemas híbridos tem sido explorados durante os últimos anos com o objetivo de tentar reproduzir em um problema determinístico, a característica caótica e evolutiva que os erros associados a modelagem possuem e que são fundamentais para a determinação dos campos de análise.

1.1 Motivação

O Centro de Previsão de Tempo e Estudos Climáticos (CPTEC) do Instituto Nacional de Pesquisas Espaciais (INPE), desde meados do ano de 2010, desenvolveu algumas aplicações com alguma variação do filtro de Kalman por conjunto utilizando as previsões do Modelo de Circulação Geral da Atmosfera (MCGA-CPTEC/INPE). Algumas destas aplicações, foram baseadas no *Local Ensemble Transform Kalman Filter* (LETKF) (HUNT et al., 2007), valendo-se de vantagens como baixo custo computacional e fácil implementação. Estas características tornam o LETKF uma opção atrativa para estudos acadêmicos e aplicações com diversos modelos. Entretanto, para estas aplicações com o LETKF no CPTEC, limitações estiveram presentes no tratamento das observações não convencionais e outras fontes de dados observacionais, as quais demandariam o desenvolvimento de operadores específicos e outras rotinas (e.g., ciclo de assimilação de dados) para apenas então viabilizar uma possível aplicação operacional robusta.

Hoffman et al. (2008) desenvolveram com o LETKF no CPTEC o *CPTEC Ocean Data Assimilation System* (CODAS). Em estudos preliminares, os autores mostraram que a análise oceânica com até 12 membros do CODAS utilizando como conjunto de previsões as previsões do modelo *Modular Ocean Model 4* (MOM4), se mostrou

bastante próxima a realidade (neste estudo o modelo foi considerado como perfeito). Aravéquia (2008) realizou estudos com a assimilação de radiâncias do sensor *Atmospheric InfraRed Sounder* (AIRS) utilizando o modelo de transferência radiativa *Community Radiative Transfer Model* (CRTM) do *Joint Center for Satellite Data Assimilation* da *National Oceanic and Atmospheric Administration* (JCSDA/NOAA) e encontrou que a assimilação das medidas diretas de radiâncias exercem um impacto positivo sobre a América do Sul. Souza et al. (2010) realizaram comparações entre dois esquemas de convecção *cumulus* (Grell e Kuo), do MCGA-CPTEC/INPE e qual o seu impacto para as análises do LETKF, utilizando as previsões de curto prazo do modelo na resolução TQ0126L028. Os resultados mostraram que com o esquema de Grell houve redução do erro quadrático médio em relação as análises do *National Centers for Environmental Predictions* (NCEP). Cintra (2010) simularam o LETKF utilizando redes neurais para o modelo *Simplified Parameterizations, primitive-Equation Dynamics* (SPEEDY) (MOLTENI, 2003). Neste trabalho, foi encontrado que as redes neurais conseguiram simular as principais características do filtro, com a principal vantagem de ser computacionalmente ainda mais barato. Medeiros (2011) realizou um estudo para avaliar o impacto da assimilação de dados de radiâncias na temperatura do ar produzida pela análise do LETKF. Diniz (2012), realizou um estudo sobre o impacto que os diversos tipos de observações disponíveis para o LETKF exercem nas previsões de curto prazo produzidas pelo MCGA-CPTEC/INPE a partir das análises do LETKF. Avanço et al. (2013) e Sapucci et al. (2016) desenvolveram um operador observação baseado no software *Radio Occultation Processing Package* (ROPP) para a assimilação de dados de *Radio Occultation-Global Positioning System* (ROGPS) para o LETKF e mostraram que a inclusão destes dados exerce impacto positivo nas análises do sistema sobre a América do Sul.

Mais recentemente, no CPTEC tem-se aplicado o sistema *Gridpoint Statistical Interpolation* (GSI) (WU et al., 2002; KLEIST et al., 2009), para gerar análises em escala global utilizando o MCGA-CPTEC/INPE e também (a partir de 2016), o modelo *Brazilian Atmospheric Model* (BAM). Diferente dos sistemas de assimilação de dados que utilizam o filtro de Kalman por conjunto, o GSI é um sistema variacional que minimiza uma função custo para se determinar o estado da análise. Diversos trabalhos com estudos de impacto e diagnósticos foram realizados com o sistema GSI acoplado com o MCGA-CPTEC/INPE (sendo denominado de G3DVAR) e que foram importantes também na manutenção do sistema operacional. Azevedo (2014), aplicou a técnica de *Observing System Experiments* (OSE) para avaliar o impacto dos diferentes tipos de sistemas de observações nas análises e previsões do sistema. Araujo (2015) estudou a parametrização do comprimento de rugosidade térmica,

um importante parâmetro na determinação da temperatura da superfície terrestre, utilizando o modelo *Simplified Simple Biosphere* (SSiB) no sistema G3DVAR. Penna et al. (2015) avaliou a sensibilidade dos canais sensíveis a superfície terrestre, utilizando as radiâncias do AMSU/A a partir dos satélites 15 e 19 da série NOAA. Rodrigues (2017) realizou um estudo com o sistema G3DVAR para avaliar o impacto da injeção de umidade, um importante característica do sistema de assimilação GSI que permite controlar soluções não físicas da umidade (negativa e supersaturada).

Um sistema sequencial baseado no filtro de Kalman por conjunto e um sistema variacional em 3 dimensões, são os componentes necessários para um sistema híbrido de assimilação de dados que possa ser utilizado operacionalmente em um futuro próximo pelo CPTEC. Além disso, nos principais centros operacionais de PNT e assimilação de dados (e.g., NCEP, *European Centre for Medium Range Weather Forecasts* (ECMWF) e *Meteorological Office* (MetOffice), *Deutsche Wetterdienst* (DWD), *Hydrometeorological Center of Russia* (RHMC) e outros), já desenvolveram ou estão desenvolvendo sistemas híbridos (BROWN; JAKOB, 2011). Durante o *Sixth Symposium on Data Assimilation* realizado pela WMO nos Estados Unidos (em outubro de 2013), ficou bastante evidente a tendência pela escolha dos centros operacionais por sistemas híbridos, cujas vantagens já haviam sido evidenciadas principalmente nos trabalhos de Wang et al. (2007), Wang et al. (2008a), Wang et al. (2008b), sendo que boa parte dos resultados mostrados se referem a um conjunto de previsões provenientes do sistema baseado em 4DVar. Os motivos pelos quais os centros operacionais seguem esta tendência, em resumo, são os seguintes:

- a) Atualmente, o limite técnico dos sistemas operacionais de assimilação de dados está na capacidade computacional, o qual limita também o aproveitamento do crescente número de observações não convencionais (i.e., radiâncias);
- b) Tendo-se em vista este cenário, sabe-se que para calcular a análise pela abordagem variacional, é necessária a inversão da matriz de covariâncias dos erros de previsão. Devido a quantidade de variáveis e pontos de grade representando os estados do modelo, é praticamente impossível armazenar a matriz de covariâncias por completo. Por isso, ela é simplificada em sua representação atual e deve ser pré-condicionada para a minimização da função custo;
- c) Uma abordagem alternativa para se representar as covariâncias dos erros de previsão é através de técnicas estatísticas e uma técnica que permite

esta representação é o filtro de Kalman por conjunto (e suas variações);

- d) Sistemas variacionais tratam as covariâncias dos erros de previsão como sendo informações estáticas (considerando que os erros crescem linearmente nas primeiras 48 horas de previsão, e.g., Zhang e Krishnamurti (1999), e que sua distribuição é Gaussiana), o que implica na falta de habilidade do sistema em incorporar as variações diárias dos erros associados as previsões e esta importante informação que deixa de ser utilizada pela análise;
- e) Sistemas sequenciais baseados no filtro de Kalman por conjunto tem a característica de fornecer uma estimativa dos “erros do dia”, mas podem ser computacionalmente caros, além do que há que se lidar com conjuntos de análises/previsões relativamente grandes (e.g., 40 membros ou mais) quando se trabalha com sistemas caóticos (como a atmosfera) e, dificuldades com estes conjuntos podem incluir problemas de amostragem (*undersampling*), crescimento inapropriado das perturbações (*imbreding*), divergência do filtro e correlações espúrias;
- f) Através da aplicação de sistemas híbridos, mostra-se que é possível incluir a estimativa dos “erros do dia” dentro do ciclo de assimilação de dados variacional (i.e., completando a informação representada na matriz de covariâncias estática) ponderando-se a contribuição das covariâncias dos erros de previsão da parte estática/estacionária (provenientes do sistema variacional) e dinâmico/não estacionária (provenientes do sistema baseado no filtro de Kalman por conjunto). Além disso, um sistema híbrido pode não requerer um conjunto muito grande a ser mantido (devido a utilização da própria matriz estática), possibilitando a diminuição do custo computacional envolvido;
- g) Sistemas híbridos podem representar o acoplamento entre sistemas síncronos (i.e., que consideram o tempo como dimensão, como o 4DVar) com o filtro de Kalman por conjunto, ou ainda um conjunto de realizações dos sistemas 3DVars ou 4DVars. Ou seja, os sistemas híbridos introduzem uma alternativa para a modelagem de erros não Gaussianos.

Um dos maiores desafios para a assimilação de dados é a especificação da matriz de covariâncias dos erros de previsão. As técnicas sequenciais e variacionais, como a IO, o 3DVar e o 4DVar, preconizam que a matriz de covariâncias dos erros de previsão seja pré-determinada, sendo esta uma característica estabelecida pela metodologia

utilizada nestes sistemas. Por outro lado, os sistemas baseados no filtro de Kalman por conjunto, preconizam que a matriz de covariâncias dos erros de previsão é determinada com base nas diferenças entre os membros do conjunto e a sua média. Esta é uma grande vantagem destes sistemas em comparação com os demais. Porém, um conjunto infinito de previsões seria necessário para que se possa acessar toda a variabilidade atmosférica e então se amostrar todos os graus de liberdade necessários para a construção de uma matriz de covariâncias completa. Além disso, toda a estrutura relacionada ao tratamento das observações e (e.g., operadores de observações de radiâncias, a determinação e a aplicação de injunções (i.e., *constraints*), fazem com que esses métodos sejam mais vantajosos pela sua infraestrutura.

Diante deste fato, ao longo da evolução das técnicas de assimilação de dados, inúmeros desenvolvimentos foram realizados com o objetivo de se amostrar de forma satisfatória as quantidades necessárias para se ponderar as contribuições das previsões na determinação das análises. As seções a seguir apresentam as técnicas mais comuns em assimilação de dados meteorológicos e como é feito o tratamento das covariâncias dos erros de previsão em cada uma delas.

1.2 Revisão Bibliográfica

1.2.1 Análise Objetiva

O emprego de análise objetiva em assimilação deu-se no início da década de 1950, em que eram utilizados polinômios para realizar a interpolação local ou global de dados observacionais com o objetivo de se inicializar um modelo numérico de previsão de tempo.

Até 1954 os primeiros experimentos em PNT foram realizados por meio da interpolação manual e empírica das observações em ponto de grade, com a finalidade de se prognosticar o tempo a partir de informações previamente obtidas e as poucas observações de tempo disponíveis. Esta atividade compreendia um conjunto bastante reduzido de técnicas e era bastante dependente da habilidade e conhecimento do meteorologista. A técnica ficou conhecida como análise subjetiva. As primeiras tentativas de se produzir uma análise objetiva deu-se, no entanto, no final de década de 1940. [Panofsky \(1949\)](#) deduziu um polinômio interpolador global (i.e., o mesmo polinômio era utilizado em toda a grade a ser analisada). Já na década de 1950, [Gilchrist e Cressman \(1954\)](#) utilizando as ideias iniciais de Panofsky, desenvolveram uma versão local de um polinômio interpolador, o qual podia ser definido para cada ponto de grade a ser analisado. Mas mesmo com estes avanços que permitiram a evolução da

análise subjetiva para a análise objetiva, ainda restava o seguinte problema: como automatizar o procedimento de análise objetiva a fim de torná-lo operacional? Neste caso, o sentido da palavra operacional, tem o seguinte significado: baixo tempo de computação e erro comparável com a da análise subjetiva.

A partir de 1950, vários fatos importantes ocorreram. O ano de 1954 marca, por exemplo, o lançamento do primeiro computador de uso comercial da *International Business Machines* (IBM, modelo 704). Em seguida, o ano de 1955 é marcado com início da PNT operacional. Além disso, há também a publicação do trabalho de Bergthórsson e Döös (1955) em que os autores agregam à recém formulada análise objetiva os elementos necessários para torná-la operacional. O método de análise objetiva espacial de Bergthórsson e Döös (1955) permite automatizar o processo de análise objetiva de forma rápida e eficiente (se comparada com a análise subjetiva convencional) e este método, foi o precursor do método de Correções Sucessivas (CS).

O método proposto por Bergthórsson e Döös (1955) representou um grande avanço para as técnicas da época e ajudou a cunhar as principais características que hoje estão presentes na assimilação de dados moderna. Estas características incluem a utilização de uma previsão anterior a análise, fosse uma previsão, uma climatologia ou a combinação de ambos. No método proposto, os pesos atribuídos as previsões e as observações é empírico e dado em função da distância entre o ponto de grade analisado e ponto da estação (i.e., o ponto em que a observação está situada). Segundo Daley (1996), na formulação do método proposto Bergthórsson e Döös (1955) são feitas as seguintes considerações: as previsões e as observações contém erros e estes são não correlacionados; o erro da previsão é homogêneo (i.e., igual em todos os pontos de grade) e conseqüentemente a sua variância é independente do local; o erro da observação é espacialmente não correlacionado (i.e., cada estação tem o seu próprio erro) e este é dado apenas em função do erro do instrumento de medição. De forma geral, a formulação proposta possui o seguinte aspecto:

$$f_A(\mathbf{r}_i) = \frac{E_B^{-2} f_B(\mathbf{r}_i) + E_O^{-2}(k)w(\mathbf{r}_k - \mathbf{r}_i)[f_B(\mathbf{r}_i) + f_O(\mathbf{r}_k) - f_B(\mathbf{r}_k)]}{E_B^{-2} + E_O^{-2}(k)w(\mathbf{r}_k - \mathbf{r}_i)} \quad (1.1)$$

onde:

- f_A é o valor obtido pelo procedimento de análise objetiva no ponto de grade;
- f_B é o valor do campo previsto (anterior a observação) no ponto de grade;

- f_O é o valor da observação no ponto da estação;
- \mathbf{r}_i é a posição do ponto de grade;
- \mathbf{r}_k é a posição do ponto da estação;
- w é o peso;
- E_B é a variância do erro da previsão (homogêneo);
- E_O é a variância do erro da observação (espacialmente não correlacionado).

Na Eq. 1.1 estão relacionadas previsões (ou uma combinação destas com alguma climatologia) anteriores a análise e as observações, com seus pesos atribuídos em função da distância entre os pontos de grade das previsões e os pontos de estação das observações. Neste equação, é importante notar as seguintes propriedades: quando $\mathbf{r}_k = \mathbf{r}_i$ (i.e., quando os pontos de grade e da estação coincidem), então $w(\mathbf{r}_k - \mathbf{r}_i) = 1$. De outra forma, quando $|\mathbf{r}_k - \mathbf{r}_i| \rightarrow \text{inf}$ (i.e., quando a distância entre os pontos aumenta), então $w(\mathbf{r}_k - \mathbf{r}_i) \rightarrow 0$.

Com os desenvolvimentos da análise objetiva apontando para um futuro operacional, Cressman (1959) estabelece o primeiro sistema de análise objetiva operacional nos Estados Unidos no *Joint Numerical Weather Prediction Unit* (JNWPU, o centro de pesquisas em PNT precedente ao *National Meteorological Center* (NMC, atual NCEP). O sistema operacional planejado por Cressman foi baseado no trabalho de Bergthórsson e Döös (1955) e neste trabalho o termo *first guess* é utilizado para designar um “chute inicial”, ou uma previsão de curto prazo. Neste sistema, as correções (das previsões de curto prazo a partir das observações) são feitas com as previsões interpoladas nos pontos das estações de forma que estas correções são feitas por algumas varreduras, as quais são mais suaves a cada nova varredura. O domínio é esférico e centrado na América do Norte e considera um raio de influência variável em que a distância entre as estações e o ponto de grade é considerada como peso.

Os métodos de análise objetiva evoluíram ao longo do Século XX, e acompanharam a evolução do sistema de observação terrestre. Com a disponibilidade das observações remotas da Terra a partir dos anos 1950, os métodos de assimilação de dados passaram também a incluir outros tipos de observações, que não apenas as observações geofísicas (e.g., temperatura, umidade, pressão etc). Eventualmente, observações de *retrievals* (i.e., perfis recuperados) passaram a incrementar o conjunto de dados de observações a serem assimiladas e em seguida, dados de radiâncias (observações de satélite)

também passaram a fazer parte desse conjunto de dados. O método variacional representa esta evolução (SASAKI, 1958; SASAKI, 1970) nos métodos e técnicas e o seu estabelecimento promoveu o desenvolvimento de algoritmos de ingestão desses dados. Portanto, dentre todos os métodos de assimilação de dados existentes, a estrutura agregada ao método variacional permite não apenas a assimilação de uma grande variedade de dados de observação meteorológica, mas permite também o desenvolvimento de operadores de observação não lineares, os quais são fundamentais para a assimilação dos dados não convencionais.

1.2.2 Assimilação de Dados por Conjuntos

Com o avanço da tecnologia, o computador eletrônico tornou-se cada vez mais acessível e novas possibilidades começaram a ser exploradas no campo da PNT. No final da década de 1960 e início da década de 1970, Epstein (1969) e Leith (1974) perceberam que, dada a natureza caótica da atmosfera apontada por Lorenz e o seu limite de previsibilidade, ao invés de se calcular deterministicamente apenas uma previsão de tempo, seria possível calcular um conjunto delas e, a partir deste conjunto, extrair um conjunto de momentos estatísticos, através dos quais seria possível construir a distribuição de probabilidades das incertezas das condições iniciais. Neste sentido, com um conjunto infinito de previsões perturbadas, seria possível descrever a incerteza da análise (condição inicial). Com isso, iniciaram-se os estudos da modelagem numérica por conjunto.

Epstein realizou os primeiros estudos com previsão estocástico-dinâmica, os quais consideravam explicitamente a incerteza das previsões dos modelos e o desenvolvimento de técnicas que consideravam a construção de distribuições de probabilidades com base na estimativa dos momentos estatísticos fornecidos pelos próprios modelos dinâmicos. No entanto, mesmo para modelos com poucos graus de liberdade, esta aproximação para a determinação de todos os momentos estatísticos de uma distribuição de probabilidades mostrou-se computacionalmente cara, e então uma nova aproximação foi tentada, desta vez, considerando apenas dois momentos estatísticos: média e covariância. Então, percebeu-se que com apenas estes dois momentos estatísticos era possível construir-se uma distribuição de probabilidades que representasse as incertezas da condição inicial. Em seguida, na década de 1970, Leith mostrou que as previsões obtidas com a técnica de Epstein eram válidas apenas se a distribuição de probabilidades gerada pelo conjunto fosse uma amostra representativa da distribuição de probabilidades relacionadas a atmosfera. Com isso, Leith propôs a utilização de um conjunto de membros - ao invés de se utilizar apenas um único membro, como

em uma única previsão determinística. A proposta de Leith empregava a técnica de Monte Carlo, em que a solução é aproximada de forma exaustiva por meio de repetidas simulações aproximadas do fenômeno. Esta similaridade com o aspecto da ideia da previsão de tempo por conjunto possibilitou a adequação da técnica para gerar um conjunto de previsões independentes.

Ao mesmo tempo que Epstein e Leith desenvolviam as bases da previsão dinâmico-estocástica, e cujas ideias foram posteriormente utilizadas por, e.g., Hoffman e Kalnay (1983) no desenvolvimento da técnica de *Lagged Average Forecasting* (LAF), surgia o filtro de Kalman-Bucy (KALMAN; BUCY, 1961). O filtro de Kalman-Bucy, ou simplesmente filtro de Kalman, é um algoritmo desenvolvido com o objetivo de se utilizar medições ruidosas observadas ao longo do tempo e realizar estimativas de estados futuros do modelo do sistema observado. Estas estimativas são melhores do que aquelas que seriam produzidas utilizando-se apenas uma única medição observada no tempo (no sentido Bayesiano). O filtro de Kalman é um algoritmo aplicado em problemas tipicamente lineares e foi empregado com sucesso em diversos desenvolvimentos associados a exploração espacial, indústria aeronáutica e militar. Uma principais características do filtro de Kalman linear está na estimativa dos momentos estatísticos, os quais são utilizados na predição dos estados do sistema observado, e que são fundamentais na determinação das relações entre os estados do sistema observados, i.e., a estimativa das variâncias e das covariâncias dos estados do sistema observados.

Tais características do filtro de Kalman o tornaram atrativo para aplicações em problemas com dinâmica não linear, em que a estimativa das covariâncias é necessária para a determinação do estado. O filtro de Kalman por conjunto (*Ensemble Kalman Filter* - EnKF, Evensen (1994)), foi então desenvolvido como uma simplificação do filtro de Kalman original, em que as covariâncias são então atualizadas utilizando os próprios estados do sistema, ao invés de um modelo para a sua representação. Nas Seções 2.2.3 e 2.2.4 do Capítulo 2, será apresentada uma introdução mais detalhada de dois tipos de filtros de Kalman por conjunto.

1.2.3 Assimilação de Dados Híbrida

Combinando-se as características de dois ou mais sistemas de assimilação de dados, pode-se obter o que se chama de sistema híbrido. Um sistema híbrido, é então um sistema cuja solução ou é variacional, ou é sequencial, mas que apresenta entre suas componentes, uma combinação - em geral linear, de dois ou mais parâmetros. O exemplo mais comum de sistema híbrido em aplicação na assimilação de dados

atmosféricos, são os sistemas por conjunto-variacionais, em que uma combinação linear entre as matrizes de covariâncias do filtro de Kalman por conjunto e do sistema 3DVar (ou 4DVar) é feita.

Os primeiros trabalhos a trazerem a tona esta nova abordagem para o tratamento do problema de análise meteorológica atmosférica, é o trabalho de Hamill e Snyder (2000) no qual os autores propuseram um sistema híbrido entre o *Ensemble Kalman Filter* (EnKF) e o 3DVar utilizando um modelo quasi-geostrófico perfeito (ou seja, considerando a climatologia do modelo como a verdade). Etherton e Bishop (2004) testaram a resistência do sistema híbrido proposto por Hamill e Snyder (2000) utilizando o mesmo modelo quasi-geostrófico perfeito e fizeram testes com um sistema híbrido semelhante utilizando um *Ensemble Transform Kalman Filter* (ETKF) aplicado ao mesmo modelo. Posteriormente, outros autores também realizaram testes com este tipo de sistema híbrido (WANG et al., 2008a; WANG et al., 2008b; TODLING; AKKRAOUI, 2013) e desenvolveram também outros tipos de híbridos com o objetivo de se acessar características diferentes dos sistemas envolvidos (BUEHNER, 2005; ZUPANSKI, 2005; WANG et al., 2007; ZHANG et al., 2009; LEI et al., 2012; CLAYTON et al., 2012; AKKRAOUI; TODLING, 2013).

Como pode-se observar, diversos testes foram realizados com diferentes modelos e diferentes esquemas de assimilação de dados. Modelos completos e simplificados e técnicas variacionais de 3 e 4 dimensões, em conjunto com sistemas sequenciais como o filtro de Kalman por conjunto e a Interpolação Ótima, geraram um conjunto de aproximações para a solução do problema de análise atmosférica em que o senso determinístico é alcançado através de técnicas - em sua maioria estatísticas, como o filtro de Kalman. Esta combinação tem se mostrado bastante promissora porque permite que os sistemas em conjunto, amenizem suas deficiências mútuas. Com isso, é notável que alguns sistemas híbridos apresentem como resultado a melhoria da análise variacional (e.g., sistemas do tipo EnKF-Var, ETKF-Var, Hamill e Snyder (2000), Etherton e Bishop (2004) respectivamente) e outros a melhoria da análise do sistema por conjunto (e.g., *Hybrid Nudging Ensemble Kalman Filter* - HNEKF, Lei et al. (2012)).

Uma das principais características dos sistemas híbridos é que os seus sistemas componentes podem ser melhorados mutuamente. Desta forma, pode-se utilizar as covariâncias dos erros de previsão (ou modelagem) provenientes de algum tipo de EnKF para melhorar a especificação das covariâncias na minimização da Função Custo (J) de um sistema variacional, em que estão relacionados os vetores de estados

do modelo e observações e as suas respectivas matrizes de erros (e.g., Eq. 2.1). Por outro lado, sabe-se que o sistema por conjunto possui posto (matricial) incompleto (*rank deficient*) e que a representação das covariâncias dos erros de previsão dos sistemas variacionais possui posto completo (*full rank*, ou seja, contém todos os graus de liberdade definidos). Logo, o sistema variacional pode melhorar o ciclo de assimilação de dados de um sistema baseado no filtro de Kalman por conjunto, pois, do contrário, para que o conjunto possua posto completo, esse deve ser infinito. Como consequência disto, a resposta entre as análises e as previsões provenientes dos dois sistemas, pode melhorar o sistema variacional. Além disso, o filtro de Kalman por conjunto, resolve um problema linear, enquanto que a minimização de uma função custo variacional é um problema não linear (a função custo variacional, 3D ou 4D é uma função quadrática). Este cenário favorece o conjunto, pois a previsão fornecida para o sistema por conjunto possui como características estas relações não lineares, que são impressas na análise e propagadas pelo modelo de PNT.

Hamill e Snyder (2000) foram os primeiros a propor um sistema híbrido. Neste trabalho os autores propuseram um sistema híbrido entre o EnKF e o 3DVar utilizando um modelo quasi-geostrófico perfeito (ou seja, quando a climatologia do modelo é considerada como verdade). Neste híbrido, os autores calcularam uma matriz de covariâncias dos erros de previsão (matriz \mathbf{B}) utilizando a Eq. 1.2 (as equações abaixo foram modificadas a partir dos seus respectivos originais para facilitar a comparação com as equações que serão utilizadas no trabalho de tese).

$$\mathbf{B} = (1 - \alpha)\mathbf{P}^b + \alpha\mathbf{S}\mathbf{C}\mathbf{S}^T \quad (1.2)$$

Esta equação representa uma combinação linear entre a matriz de covariâncias do conjunto \mathbf{P}^b e um modelo de covariâncias “estático” dado por $\mathbf{S}\mathbf{C}\mathbf{S}^T$ (em que \mathbf{S} é um operador que transforma do espaço espectral para ponto de grade e \mathbf{C} uma matriz com as variâncias dos coeficientes espectrais), em que o peso α é atribuído a cada uma das partes. Etherton e Bishop (2004) testaram a resistência do sistema híbrido proposto por Hamill e Snyder (2000) utilizando o mesmo modelo quasi-geostrófico perfeito e fizeram testes com um sistema híbrido semelhante utilizando um ETKF aplicado ao mesmo modelo. A matriz de covariâncias utilizada pelos autores neste trabalho é a Eq. 1.3 a seguir:

$$\mathbf{B} = (1 - \alpha)\lambda\mathbf{P}^b + \alpha\rho\mathbf{B}_{3dvar} \quad (1.3)$$

A Eq. 1.3 é semelhante aquela utilizada por Hamill e Snyder (2000), mas utilizando uma matriz de covariâncias estática \mathbf{B}_{3dvar} de um sistema variacional de 3 dimensões e com a inclusão dos coeficientes λ e ρ , que também são pesos (assim como α) e possuem a função de capturar as contribuições de cada uma das parcelas da combinação linear. Estes pesos são calculados a cada ciclo de assimilação de dados e refletem as influências que as matrizes \mathbf{P}^b e \mathbf{B}_{3dvar} exercem sobre a matriz \mathbf{B} calculada. Buehner (2005) em um estudo sobre a modelagem das covariâncias dos erros de previsão em um sistema de PNT quasi-operacional, avaliou as covariâncias dos erros de previsão geradas pelo sistema híbrido de Hamill e Snyder (2000) em comparação com resultados obtidos com outras técnicas utilizando o filtro e Kalman por conjunto e utilizando técnicas variacionais. Zupanski (2005) desenvolveu o *Maximum Likelihood Ensemble Filter* (MLEF), permitindo que um operador não linear fosse utilizado na minimização da função custo variacional, sendo este realizado no espaço do conjunto. Esta abordagem permite, por exemplo, a inclusão de dados de precipitação na assimilação de dados por conjunto. Esta é uma abordagem interessante porque os erros associados a precipitação não possuem distribuição Gaussiana, enquanto que, em geral, os erros de previsão são considerados com distribuição Gaussiana. Wang et al. (2007) propuseram um sistema híbrido combinando o ETKF e a técnica de Interpolação Ótima (*Optimum Interpolation - OI*) e fizeram comparações com o esquema *Ensemble Square Root Filter* (EnSRF). A matriz de covariâncias utilizada pelos autores neste estudo (Eq. 1.4) também é semelhante aquela utilizada por Hamill e Snyder (2000) e por Etherton e Bishop (2004), mas com a diferença de que o método de assimilação de dados (para a parte estática do sistema híbrido) foi o método de Interpolação Ótima:

$$\mathbf{B} = (1 - \alpha)\mathbf{P}^b + \alpha\mathbf{B}_{IO} \quad (1.4)$$

Wang et al. (2008a) e Wang et al. (2008b), com base no sistema híbrido elaborado por Hamill e Snyder (2000), propuseram um sistema híbrido para o modelo WRF (um modelo de PNT de física completa) definindo um novo incremento de análise através de uma extensão da variável de controle, uma técnica originalmente proposta por Lorenc (2003). A matriz de covariâncias utilizada pelos autores (Eq. 1.5) utiliza uma abordagem um pouco diferente, por utilizar um produto Schur entre a matriz de covariâncias do conjunto e uma matriz de correlação, cuja função é localizar o novo incremento de análise.

$$\mathbf{B} = (1 - \alpha)\mathbf{B}_{3dvar} + \alpha\mathbf{P}^b \circ \mathbf{C} \quad (1.5)$$

Zhang et al. (2009) elaboraram um acoplamento entre o EnKF e o sistema 4DVar (denominado E4DVAR) aplicado ao modelo Lorenz96. Neste estudo, os autores também definiram uma matriz de covariâncias utilizando uma combinação linear entre as matrizes de covariâncias do EnKF e do 4DVar (Eq. 1.6).

$$\mathbf{B} = \alpha\mathbf{P}^b + (1 - \alpha)\mathbf{B}_{4dvar} \quad (1.6)$$

Lei et al. (2012) desenvolveram um sistema híbrido baseado na técnica de *Nudging* e EnKF (denominado HNEKF) aplicado também ao modelo Lorenz96. Clayton et al. (2012) avaliaram a implementação operacional de um sistema híbrido utilizando o sistema de previsão por conjunto do MetOffice (*MetOffice Global and Regional Ensemble Prediction System* - MOGREPS, baseado no ETKF) e o sistema 4DVar aplicado ao modelo global operacional do centro. A matriz de covariâncias utilizada nesta implementação (Eq. 1.7) também inclui o produto Schur entre a matriz de covariâncias do ETKF e uma matriz de correlação \mathbf{C} que tem a função de localizar correlações muito pequenas provenientes do conjunto e que podem ser dominadas por erros de amostragem. Na Eq. 1.7, os coeficientes α_c^2 e α_e^2 são os pesos dados as matrizes \mathbf{B}_c (do 4DVar) e \mathbf{B}_e (Eq.1.8, do ETKF) respectivamente:

$$\mathbf{B} = \alpha_c^2\mathbf{B}_c + \alpha_e^2\mathbf{B}_e \quad (1.7)$$

$$\mathbf{B}_e = \mathbf{P}^b \circ \mathbf{C} \quad (1.8)$$

Mais recentemente, Todling e Akkraoui (2013) estudam um sistema híbrido sem a necessidade de se empregar um sistema de análise por conjunto baseado em filtro de Kalman. Wang e Lei (2013) estudam uma extensão da técnica aplicada por Wang et al. (2008a), Wang et al. (2008b) (Variável de Controle Estendida) e realizam um sistema híbrido entre um filtro de Kalman por conjunto e o 4DVar (formando um 4DEnsVar) utilizando as previsões do modelo GFS. A grande vantagem deste acoplamento é utilizar as covariâncias do conjunto para evitar a necessidade de se utilizar o modelo tangente linear para propagar as covariâncias no tempo e assimilando as observações no tempo correto.

1.2.4 Modelagem de Covariâncias

Como o problema de se especificar as covariâncias para uso na minimização de uma função custo variacional ou para aplicação direta em um sistema mais simples com poucos graus de liberdade, foi necessário desenvolver-se modelos matemáticos capazes de descrever a natureza não linear das relações multivariadas das quantidades de interesse. Ao longo desse desenvolvimento, diferentes técnicas foram propostas, e estas técnicas podem ser divididas em basicamente, dois grupos: i) técnicas que calculam explicitamente as covariâncias, utilizando algum tipo de metodologia paramétrica; ii) técnicas que amostram a incerteza de um determinado conjunto de informações utilizando um procedimento do tipo Monte Carlo.

Sistemas híbridos de assimilação de dados utilizando um filtro de Kalman por conjunto e uma sistema variacional foram introduzidos no início dos anos 2000 (HAMILL; SNYDER, 2000; LORENC, 2003; ZUPANSKI, 2005) e tem sido aplicados com ênfase na amostragem e na representação das variações diárias das covariâncias dos erros de previsão, introduzindo os chamados “erros do dia” (CORAZZA et al., 2003) a parte estática das covariâncias. A representação das covariâncias dos erros de previsão é um dos principais problemas a serem endereçados na assimilação de dados operacional e a sua determinação em sistemas determinísticos (e.g., sistemas variacionais) é de grande importância.

Sistemas variacionais como o 3DVar utilizam uma matriz estática para representar as covariâncias dos erros de previsão. Isto significa que estas covariâncias - embora tenham sido especificadas com base em previsões distribuídas ao longo do tempo, elas não variam em conjunto com a evolução temporal das análises e previsões e são, portanto, fixas durante todo o processo de assimilação de dados. Uma situação similar ocorre com o 4DVar em que estas mesmas covariâncias são utilizadas entre as janelas de assimilação de dados, embora dentro desta janela de assimilação, as covariâncias são propagadas no tempo durante a assimilação das observações utilizando a versão tangente linear do modelo de previsão. As diferenças na forma como as estatísticas de erro variam com o tempo de acordo com as variações no fluxo atmosférico levam ao que se chama de “dependência de fluxo”. Anisotropia é uma característica relacionada com o aspecto geométrico das funções de correlação que fazem com que as covariâncias se ajustem aos gradientes impressos no fluxo atmosférico (e.g., como em uma zona frontal). Inomogeneidade é uma outra característica que está relacionada com a distribuição espacial das covariâncias em cada ponto de grade e como as funções de correlação são definidas. Estas características inerentes a representação

das covariâncias são atingidas quando as estruturas de covariâncias são modeladas levando-se em consideração a natureza física da dinâmica da atmosfera.

Covariâncias dependentes do fluxo atmosférico são a característica principal dos sistemas modernos de assimilação de dados e sua determinação, no início ou durante a janela de assimilação de dados, permite que a análise a ser produzida contabilize a dependência temporal das inovações trazidas pelas observações nos incrementos de análise, nas direções horizontal e vertical. Além disso, características desejáveis das covariâncias dos erros de previsão também incluem a anisotropia e a inomogeneidade (i.e., estatísticas não homogêneas, Rabier (2005)). A dependência do fluxo atmosférico refere-se às alterações da dinâmica da atmosfera e suas variações com o tempo, as quais devem ser representadas dentro da matriz de covariâncias. Anisotropia e inomogeneidade são características da distribuição espacial das covariâncias e são governadas por funções de correlação espaciais. Funções de correlação são modeladas de forma a permitir que as covariâncias sejam não homogêneas (i.e., diferentes entre os pontos de grade).

Vários esforços já foram feitos de forma a conferir algum grau de anisotropia e dependência de fluxo às covariâncias estáticas. O desenvolvimento das técnicas variacionais para uso operacional durante os anos 1990, incluiu modificações na estrutura da assimilação de dados variacional. Desroziers (1997) introduziu uma transformação de coordenadas para permitir que sistemas de assimilação de dados pudessem contabilizar as estruturas de sistemas frontais. Este desenvolvimento proporcionou a produção de uma análise com covariância fluxo-dependente e correlações anisotrópicas.

Os gradientes representados no fluxo atmosférico também incluem importantes informações que podem ser utilizadas para se determinar a anisotropia. Os gradientes espaciais das previsões podem, então, serem utilizados para modelar as estatísticas dos erros. Riishøjgaard (1998) mostrou como utilizar as alterações no campo de umidade para modelar funções de correlação que podem ser “esticadas” para acomodar as covariâncias de acordo com os gradientes nas previsões. Outras aproximações para a modelagem de funções de correlação que implicam em covariâncias fluxo-dependentes, envolveram a aplicação do método de ondeletas em conjunto com o método espectral (FISHER, 2003). Neste caso, dificuldades surgem com a especificação das ondeletas na esfera. Filtros recursivos são também um outro método conhecido para se derivar formas (de funções) quase-Gaussiana, que podem ser utilizadas na modelagem de covariâncias. Este método tem sido utilizado com sucesso em várias aplicações; Hayden e Purser (1995) aplicaram filtros recursivos isotrópicos para o

pré-processamento do *National Environmental Satellite and Data Information Service* (NESDIS); recentemente, filtros recursivos foram também aplicados ao sistema GSI para modelar a aplicação das covariâncias dos erros de previsão (WU et al., 2002).

Dentre estes desenvolvimentos, outros métodos tem sido desenvolvidos e estão mais relacionados com a representação de não-linearidades do fluxo atmosférico nas estatísticas dos erros de previsão. O 4DVar (THÉPAUT; COURTIER, 1991) é uma extensão do 3DVar (LORENC, 1986) e pode implicitamente evoluir as covariâncias dentro de uma janela de assimilação. Corazza et al. (2003) adicionou a dependência do fluxo atmosférico as covariâncias contabilizando os “erros do dia” utilizando *bredd-vectors*. Métodos híbridos (HAMILL; SNYDER, 2000) foram introduzidos aproveitando as vantagens dos métodos de filtro de Kalman por conjunto. Esta nova metodologia apresenta várias vantagens, incluindo o fato de que muitos centros operacionais já experimentaram algum método variacional e tem alguma experiência também com filtros de Kalman por conjunto. Cada um destes métodos e metodologias podem ser tomados como um complemento para se melhorar as deficiências do outro (WANG et al., 2007; WANG et al., 2009). Recentemente, métodos híbridos tem sido aplicados em situações reais (WANG et al., 2013; CLAYTON et al., 2012), utilizando covariâncias estimadas a partir de um filtro de Kalman por conjunto em combinação com as covariâncias estáticas de um 3DVar.

1.3 Objetivos

Esta tese de doutorado insere-se no contexto dos avanços científicos realizados durante as duas últimas décadas (2000-presente) na área de assimilação de dados meteorológicos. Nesta área, os maiores desafios encontram-se na determinação das covariâncias dos erros de previsão e a sua aplicação para a determinação das análises (condições iniciais) dos modelos de PNT.

Com a premissa de que o CPTEC é um dos maiores centros de PNT do Hemisfério Sul e em consonância com a sua missão, esta tese de doutorado tenta responder ao seguinte questionamento: **qual é a contribuição das covariâncias do filtro de Kalman por conjunto na determinação das análises e previsões de um sistema de assimilação de dados global?**

Partindo-se do princípio de que existe a necessidade de se aprimorar a forma como as covariâncias globais são tratadas no sistema de assimilação de dados em uso no CPTEC, propõem-se um aprimoramento para o atual sistema de forma a incluir os “erros do dia” dentro da etapa de cálculo da análise atmosférica.

1.3.1 Objetivo Geral

Obter um conjunto de análises a partir da minimização de uma função custo variacional tridimensional, utilizando uma combinação de covariâncias estacionárias e dinâmicas.

1.3.2 Objetivos Específicos

Para se obter um conjunto de análises calculadas com base na combinação de covariâncias estacionárias e dinâmicas, as seguintes tarefas devem ser cumpridas:

- Calcular uma nova matriz de covariâncias, baseada na versão do modelo atmosférico utilizado;
- Identificar as principais estruturas associadas as covariâncias dos erros de previsão do modelo de circulação atmosférico do CPTEC;
- Habilitar o sistema de assimilação de dados GSI (componente do sistema G3DVAR) para ler as covariâncias de um filtro de Kalman por conjuntos;
- Estabelecer uma rotina de realizações cíclicas do sistema híbrido habilitado;
- Investigar os efeitos da atualização cíclica das covariâncias híbridas nas análises e previsões.

1.4 Estrutura do Documento

Para se endereçar os objetivos propostos da Seção 1.3, este documento foi organizado da seguinte maneira: no Capítulo 2, são apresentadas as metodologias e os dados utilizados no estudo. A descrição do filtro de Kalman por conjunto utilizado, está detalhado em duas seções do capítulo, bem com uma descrição do modelo de circulação geral do CPTEC e o sistema de assimilação de dados variacional. A metodologia utilizada para o cálculo da matriz de covariâncias estática está destacada e detalhada no Capítulo 3. A metodologia para a incorporação das covariâncias do conjunto na estrutura variacional do sistema de assimilação de dados bem como a descrição do ciclo de assimilação de dados, estão também destacados e são apresentados no Capítulo 4. No Capítulo 5 são apresentados os resultados, divididos em duas partes: na primeira parte, são apresentados os resultados obtidos com a aplicação da matriz de covariâncias estática e o seu impacto na assimilação das observações e nas previsões de curto prazo; na segunda parte, são apresentados os resultados obtidos com a

aplicação das covariâncias do conjunto em combinação com as covariâncias estáticas dentro da estrutura variacional, e o seu impacto nas análises e previsões até 5 dias.

Foram adicionados ao documento principal da tese, 4 apêndices que trazem resultados da teoria variacional e a aplicação da matriz de covariâncias na estrutura variacional. Estes documentos são registros de estudos prévios sobre os sistemas envolvidos no trabalho e foram adicionados para dar suporte ao entendimento de algumas seções. Com os resultados obtidos e, dentro do tempo disponível para realizar o trabalho, foram elaborados dois artigos científicos. O primeiro - submetido para a Revista Brasileira de Meteorologia, traz alguns resultados e contribuições ao entendimento da importância e aplicação das covariâncias na assimilação de dados. O segundo artigo - submetido para revista *Journal of Applied Meteorology and Climatology*, traz os resultados obtidos com o sistema híbrido e contribui para o entendimento na forma como as covariâncias híbridas são aplicadas. Estes dois artigos encontram-se nos dois últimos apêndices do documento principal.

2 DADOS E METODOLOGIA

2.1 Modelo de Circulação Geral do CPTEC

O Modelo de Circulação Geral da Atmosfera (MCGA) do CPTEC/INPE utilizado nos experimentos deste trabalho é a mesma versão utilizada pela versão operacional do sistema G3DVARv1.1.3 (versão 1.1.3, disponível para uso interno em <https://projetos.cptec.inpe.br/projects/g3dvar/repository/show/tags/v1.1.3/G3DVAR>). Esta versão do modelo é frequentemente referenciada como MCGAv4 (versão 4) e ela possui um pacote de opções de configuração da dinâmica bastante semelhante a mais nova versão do modelo (mais detalhes adiante). As diferenças principais entre estas duas versões do modelo atmosférico global do CPTEC, está no conjunto de parametrizações físicas utilizado (comunicação pessoal Silvio N. Figueroa).

O *Brazilian Atmospheric Model* (BAM), ou Modelo Atmosférico Brasileiro do CPTEC/INPE, foi lançado em Janeiro de 2016. Dada as semelhanças entre esta versão inicial do BAM e o MCGAv4 e a fim de evitar discrepâncias e confusões sobre qual versão foi de fato utilizada neste trabalho, doravante será utilizado o termo BAMv0 para designar a versão do modelo empregada nos experimentos utilizando a infraestrutura do sistema G3DVAR. Portanto, o termo BAMv0 representa o modelo MCGAv4 do sistema G3DVARv1.1.3 com algumas características (de opções da física) do modelo BAM (utilizado pela Divisão de Modelagem e Desenvolvidos, DMD). Conseqüentemente, a versão gerada a partir do estabelecimento do sistema de assimilação de dados resultante desta pesquisa, poderia ser denominada G3DVARv2 uma vez que traz significativas alterações no sistema em termos de características e funcionalidades, como a atualização do modelo atmosférico, uma nova matriz de covariâncias e outras funcionalidades no esquema de assimilação de dados (que são tratadas ao longo do escopo deste trabalho).

2.1.1 Descrição e opções do Modelo de Circulação Geral

As características principais do MCGAv4 estão descritas em [Cavalcanti et al. \(2002\)](#), enquanto que as características principais do BAM estão descritas em [Kubota \(2012\)](#) e, principalmente em [Figueroa et al. \(2016\)](#). Nesta seção, entretanto, são descritas as características principais do BAMv0, as suas principais opções e configurações utilizadas (Tabela 2.1) e como é feita a sua aplicação nos experimentos realizados.

Os testes com o cálculo da matriz de covariâncias utilizando os pares de previsões

na resolução TQ0299L064, foram feitos utilizando-se a versão MCGAv4. Um diagnóstico das amplitudes da matriz calculada nesta resolução são mostrados na Seção 3.3. Um diagnóstico semelhante para a matriz de covariâncias calculada na resolução TQ0062L028, é mostrada na Figura 2 do artigo científico do Apêndice . Os experimentos realizados com o sistema híbrido 3DVar, foram feitos, de fato, com a matriz de covariâncias na resolução TQ0062L028, utilizando-se a versão BAMv0. Ambas as configurações são bastante semelhantes entre si, sendo que as diferenças podem ser encontradas na estrutura do modelo (o BAMv0 possui uma organização diferente do MCGAv4), além do desempenho computacional. Apesar disso, com a versão BAMv0 na resolução TQ0062L028, que corresponde a uma aplicação de modestos requerimentos computacionais (e.g., o modelo foi realizado nesta resolução com apenas 48 processadores).

Tabela 2.1 - Principais opções das versões MCGAv4 e BAMv0 do modelo de circulação geral do CPTEC.

	MCGAv4	BAMv0
Resolução	TQ0299L064/TQ0062L028	TQ0062L028
Passo de Tempo	200s/1200s	1200s
Inicialização	Diabática/Modos Normais Não Lineares	Diabática/Modos Normais Não Lineares
Dinâmica	Euleriana	Semi-Lagrangiana
Transporte de Umidade	Lagrangiano	Lagrangiano
Conservação de Massa	$ln(ps)$	$ln(ps)$
Radiação de Onda Curta	CLiRAD (CHOU, 1999)	CLiRAD (CHOU, 1999)
Radiação de Onda Longa	Harshvardhan et al. (1987)	Harshvardhan et al. (1987)
Topo Camada Limite	Holtslag e Boville (1993)	Holtslag e Boville (1993)
Base Camada Limite	Mellor e Yamada (1974)	Mellor e Yamada (1974)
Esquema de Superfície	SSiB (SELLERS et al., 1996)	IBIS
Convecção Profunda	Grell (1993)	Arakawa
Convecção Rasa	Grell (1993)	Tiedtke (1983)

O modelo de circulação geral da atmosfera do CPTEC (seja o MCGAv4 ou o BAMv0) tem como base o modelo do *Center for Ocean–Land–Atmosphere Studies* (COLA) (BONATTI, 1996). O modelo é utilizado operacionalmente no CPTEC para previsão de tempo até 7 dias, previsão de tempo estendida até 15 dias (sendo uma componente do Sistema de Previsão por Conjunto (SPCON)) e previsões climáticas. As equações primitivas do modelo são resolvidas utilizando o método espectral, sendo o movimento do vento horizontal dado em termos da divergência (D) e da vorticidade (ζ). Na vertical, o modelo é representado pela coordenada vertical sigma (σ), e atualmente já estão sendo realizados testes com uma coordenada vertical

híbrida, a qual pode contribuir para a melhoria da representação dos processos físicos (que são parametrizados), principalmente sobre regiões de topografia íngreme (e.g., Cordilheira dos Andes, Himalaia, Montanhas Rochosas).

Em relação à condição inicial e condições de fronteira, o modelo de circulação do CP-TEC utiliza os coeficientes espectrais de logaritmo da pressão ($\ln(ps)$), temperatura virtual (T_v), divergência (D), vorticidade (ζ) e umidade específica (q). As condições de contorno são dadas pela temperatura da superfície do mar, neve e umidade do solo. O modelo possui um pacote de pré-processamento, que extrai as informações necessárias da condição inicial, e a prepara para a integração numérica do modelo. No pré-processo, uma das etapas principais é o *Chopping*, que realiza (entre outros procedimentos), o truncamento da condição inicial, caso esta seja fornecida em uma resolução diferente daquela em que o modelo será integrado. As previsões são escritas por meio dos coeficientes espectrais das quantidades prognósticas, a partir das quais, o pós-processamento recompõe para ponto de grade e interpola para níveis de pressão as quantidades requeridas pelo usuário.

2.2 *Gridpoint Statistical Interpolation (GSI)*

O *Gridpoint Statistical Interpolation (GSI)* (WU et al., 2002; KLEIST et al., 2009) é um sistema de assimilação de dados em espaço físico (i.e., a trajetória do modelo é corrigida a partir da minimização das incertezas das previsões no espaço das observações). Este sistema é capaz de utilizar observações convencionais e não convencionais, gerando análises variacionais em 3 ou 4 dimensões (3DVar ou 4DVar) e híbridas, em que uma matriz de covariâncias híbrida é utilizada na função custo variacional. No caso da análise híbrida do GSI, pode-se optar pelos algoritmos de filtro de Kalman por conjunto (*Ensemble Kalman Filter*, EnKF) ou o filtro raiz quadrada por conjunto (*Ensemble Square Root Filter*, EnSRF).

O GSI é um sistema em constante desenvolvimento e é uma evolução do sistema *Spectral Statistical Interpolation (SSI)*. Diversos centros operacionais ao redor do mundo tem aplicado este sistema para gerar análises independentes, as quais são utilizadas na inicialização de modelo globais e regionais. A análise independente de um centro de PNT é representada não apenas pela análise proveniente de um ciclo de assimilação de dados, mas também utilizando-se a própria matriz de covariâncias e o próprio controle de qualidade das observações.

Nas seções a seguir, são apresentadas as principais componentes que constituem o sistema GSI e que efetivamente foram abordadas no processo de elaboração do

sistema híbrido 3DVar.

2.2.1 Sistema Observacional do GSI

Um sistema observacional, dentro do contexto da assimilação de dados meteorológicos, é a componente do sistema que envolve tudo o que está relacionado ao tratamento das observações. Em inglês, o termo *observer* (ou sistema observacional), já foi empregado por Griffith e Nichols (1994) em um estudo sobre a utilização de sistemas observacionais em assimilação de dados. O termo original não tem sido extensivamente explorado na literatura, porém será empregado aqui para uma descrição do sistema observacional do GSI.

Seguindo Griffith e Nichols (1994), um sistema observacional é a componente do sistema que se utiliza das observações meteorológicas para conduzir o sistema de modelagem (neste caso, o modelo de PNT) para o estado representado pelas observações. Neste sentido, uma descrição mais ampla poderia envolver, além das observações em si, técnicas como o *Nudging* (ou relaxação Newtoniana), em que dados observacionais são utilizados para conduzir (ou forçar) o modelo a um estado mais próximo das observações, a partir da inclusão de termos extras nas equações prognósticas para esta finalidade. De forma mais restrita, um sistema observacional no contexto da assimilação de dados, é aquele que envolve as inovações (i.e., $\mathbf{y}^o - H(\mathbf{x}^b)$) e, portanto, as observações (\mathbf{y}^o), o estado do modelo (\mathbf{x}^b), o operador observação (H ou \mathbf{H}) e a matriz de covariância dos erros de observação (\mathbf{R}). Sistemas de observações meteorológicas também podem ser considerados na descrição de um sistema observacional para um sistema de assimilação de dados, porque representam metodologias específicas para a representação das observações em si (as quais podem incluir a coleta, o processamento, o armazenamento e a disseminação dos dados), as quais podem estar incluídas no operador observação.

No sistema GSI, o sistema observacional é realizado durante a minimização da função custo variacional e etapas adicionais são requeridas quando o sistema de assimilação é realizado de forma híbrida, por envolver alguma técnica de assimilação de dados por conjunto, a qual irá requerer a realização do sistema observacional para cada membro do conjunto. Tipicamente, o 3DVar calcula as inovações das observações durante a minimização da função custo e, a partir das matrizes de covariâncias dos erros de observação e previsão, parte das inovações são convertidas em incrementos de análise. Esta transformação é o que pode-se referenciar como *observer*, seguindo a ideia de Griffith e Nichols (1994) de que o sistema observacional contribui para a condução do estado do modelo para o estado representado pelas observações.

Dentro da infraestrutura do GSI, o sistema observacional é construído de forma a armazenar as informações referentes as estatísticas de *Observation minus Analysis* (OmA, ou Observação menos Análise), *Observation minus Forecast* (OmF, ou Observação menos Previsão) e convergência da minimização da função custo. A informação da convergência da função custo, também pode ser extraída do sistema observacional, porque esta é dependente da quantidade de observações que são inseridas no sistema como um todo; a cada novo ciclo de assimilação de dados, mais ou menos observações podem estar disponíveis, as quais impactam diretamente na convergência da função custo.

Na estrutura do sistema híbrido, as informações contidas no sistema observacional são importantes, pois é a partir delas que os membros do EnKF serão realizados, a partir dos quais as covariâncias do conjunto são extraídas.

2.2.2 3-Dimensional Variational Assimilation (3DVar)

Na assimilação tridimensional variacional (3DVar), assume-se que a distribuição dos erros de observação e de modelagem é Gaussiana, ou seja, assume-se que média da distribuição dos erros das variáveis de estado seja zero e que o desvio padrão é característico do erro da observação analisada. Esta é a suposição mais largamente utilizada e a sua especificação é fundamental para o sucesso do processo de assimilação de dados variacional (LORENC, 1986).

No sistema 3DVar do GSI, a função custo (em sua forma mais geral) é representado pela Eq. 2.1:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)}_{J_b} + \underbrace{\frac{1}{2}[\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x})]}_{J_o} \quad (2.1)$$

onde:

- \mathbf{x} é o vetor de estado a ser analisado (dimensão: $n \times 1$);
- \mathbf{x}^b é o vetor de estado previsto ($n \times 1$);
- $\mathbf{x} - \mathbf{x}^b$ é o vetor incremento de análise ($n \times 1$);
- \mathbf{B} é a matriz de covariâncias dos erros de previsão ($n \times n$);
- \mathbf{y}^o é o vetor de estado observado ($p \times 1$);

- $\mathbf{y}^o - H(\mathbf{x})$ é o vetor inovação ou incremento de observação ($p \times n$);
- H é o operador observação não linear (responsável por interpolar e relacionar as variáveis de estado e observação, nos respectivos pontos onde se encontram as observações);
- \mathbf{R} é a matriz de covariâncias dos erros de observação ($p \times p$);
- J_b representa o termo da função custo referente a previsão de curto prazo;
- J_o representa o termo da função custo referente as observações.

A função custo representada pela Eq. 2.1 é quadrática e o resultado da sua minimização iterativa pode ser interpretada como uma medida da redução da distância entre a posição da variável de controle (i.e., a variável em torno da qual é realizada a minimização da função custo) e o ponto de mínimo local da função.

A análise 3DVar calculada pelo GSI está representada na Eq. 2.2. A análise é calculada a partir do mínimo da função custo representada pela Eq. 2.1, quando $\nabla J_{\mathbf{x}} = 0$ implicando em $\mathbf{x} = \mathbf{x}^a$ (a derivação completa da Eq. 2.2 e uma forma alternativa equivalente, a partir da Eq. 2.1, é apresentada no Apêndice A).

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} [(\mathbf{H}^T \mathbf{R}^{-1})(\mathbf{y}^o - H(\mathbf{x}^b))] \quad (2.2)$$

Na Eq. 2.2 tem-se que o estado da análise \mathbf{x}^a é igual ao estado da previsão de curto prazo \mathbf{x}^b mais a contribuição da inovação trazida pelas observações ($\mathbf{y}^o - H(\mathbf{x}^b)$), ponderado pela razão entre as matrizes de covariâncias dos erros de observação (\mathbf{R}) e modelagem (\mathbf{B}). A presença da matriz \mathbf{H}^T aplicada em \mathbf{R} , indica que as operações com as matrizes são realizadas em ponto de grade, i.e., \mathbf{R}^{-1} é interpolada para o espaço do modelo.

Tipicamente, em aplicações com um modelo atmosférico de circulação geral, o GSI analisa as seguintes variáveis: a função de corrente, a parte desbalanceada da velocidade potencial, a parte desbalanceada da temperatura virtual, a parte desbalanceada da pressão em superfície e a umidade relativa normalizada.

2.2.3 *Ensemble Kalman Filter (EnKF)*

O filtro de Kalman por conjunto (*Ensemble Kalman Filter*, EnKF) (EVENSEN, 2003) é um algoritmo de assimilação de dados que gera um conjunto de análises em intervalos

de tempo regulares que refletem real estado da atmosfera e a sua incerteza, através da média do conjunto e do seu espalhamento (HARLIM; HUNT, 2005).

O desenvolvimento do EnKF foi baseado em diferentes evoluções do filtro de Kalman original. O filtro de Kalman (FK) original é aplicado a problemas com dinâmica linear e uma extensão do FK para aplicações em problemas com dinâmica não linear, é encontrada no filtro de Kalman estendido (*Extended Kalman Filter*, EKF). A solução dada pelo EKF está na linearização sucessiva da trajetória do modelo através da aplicação de um modelo tangente linear (O'KANE; FREDERIKSEN, 2008). O EnKF, por outro lado, utiliza o próprio conjunto de previsões para fazer uma estimativa das covariâncias das previsões e das análises. Há que se assinalar, entretanto, que o EnKF introduzido originalmente por Evensen (2003) é estocástico, no sentido de que as observações são perturbadas para gerar um conjunto de análises. Existe uma versão determinística do EnKF em que as observações não são perturbadas e o conjunto de análises é obtido utilizando-se um conjunto de previsões e observações (determinísticas) e que é utilizada pelo sistema GSI.

As equações de análise, ganho de Kalman e estimativas das covariâncias dos erros de previsão e análise do FK linear são dadas por:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)] \quad (2.3)$$

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T (\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (2.4)$$

$$\mathbf{P}^b = \mathbf{M} \mathbf{P}^a \mathbf{M}^T + \mathbf{Q} \quad (2.5)$$

$$\mathbf{P}^a = (\mathbf{I} + \mathbf{K} \mathbf{H}) \mathbf{P}^b \quad (2.6)$$

onde:

- \mathbf{x}^b : vetor de estado do modelo, representando pelos *priors* (i.e., previsões; dimensão: $m \times 1$);
- \mathbf{x}^a : vetor de estado atualizado, representado pelos *posteriors* (i.e., análises) em ponto de grade ($m \times 1$);
- \mathbf{y}^o : vetor observação ($p \times 1$);
- \mathbf{H} : operador observação linear;
- \mathbf{P}^b : matriz de covariâncias dos erros de previsão ($m \times m$);

- \mathbf{P}^a : matriz de covariâncias dos erros de análise ($m \times m$);
- \mathbf{R} : matriz de covariâncias dos erros de observação ($p \times p$);
- \mathbf{K} : matriz ganho de Kalman;
- \mathbf{M} : é o modelo com dinâmica linear;
- \mathbf{Q} : é a matriz de covariâncias dos erros aleatórios do modelo linear.

Na Eq. 2.5, a matriz \mathbf{P}^a é calculada em um instante de tempo anterior (i.e., $t = t^{n-1}$), de forma que as equações são calculadas no instante de tempo atual (i.e., $t = t^n$) e conseqüentemente a Eq. 2.6 fornece uma atualização de \mathbf{P}^a a partir de \mathbf{P}^b no instante de tempo atual (i.e., $t = t^n$). A Eq. 2.5 junto com o prognóstico do modelo (i.e., $\mathbf{M}_{t^n \rightarrow t^{n+1}}$) são chamadas de equações de previsão e as Eqs. 2.3, 2.4 e 2.6 são chamadas de equações de correção. A Eq. 2.3 representa a atualização do estado da análise utilizada pelo FK. A Eq. 2.4 representa o ganho de Kalman (\mathbf{K}) e nela está representada a covariância multivariada entre as estimativas anteriores das observações (i.e., as observações interpoladas nos pontos de grade do modelo) e as variáveis de análise do modelo, além de um mapeamento dos incrementos das observações nos pontos das observações a fim de analisar as variáveis de análise nos pontos de grade do modelo. As Eqs. 2.3 e 2.4 são chamadas também de equações de atualização e guardam semelhança com a equação de análise do esquema de IO, em que a matriz peso \mathbf{W} é fixa (i.e., não é atualizada no tempo pelas covariâncias dos erros de previsão e observação). Além disso, o FK linear é dito ótimo e conseqüentemente minimiza uma função custo semelhante a função custo do 3DVar.

No caso do FK linear, a estimativa da covariância dos erros de previsão é dada em função da dinâmica linear do sistema. Para o caso em que é considerado um conjunto de previsões, as Eqs. 2.3 e 2.4 podem ser reescritas da seguinte forma:

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_e [\mathbf{y}^o - \mathbf{H}(\mathbf{x}_k^b)] \quad (2.7)$$

$$\mathbf{K}_e = \mathbf{P}_e^b \mathbf{H}^T (\mathbf{H} \mathbf{P}_e^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (2.8)$$

em que o subscrito e representa o conjunto de previsões e o subscrito k cada membro do conjunto de K previsões.

Na Eq. 2.8, a matriz de covariâncias dos erros de previsão do ganho de Kalman, é estimada por:

$$\mathbf{P}_e^b = \frac{1}{K-1} \sum_{k=1}^K (\mathbf{x}_k^b - \bar{\mathbf{x}}^b)(\mathbf{x}_k^b - \bar{\mathbf{x}}^b)^T \quad (2.9)$$

onde:

- K : é o número de membros do conjunto de previsões;
- \mathbf{x}_k^b : representa o k -ésimo membro do conjunto de previsões;
- $\bar{\mathbf{x}}^b$: representa a média do conjunto de previsões (i.e., $\bar{\mathbf{x}}^b = \frac{1}{K} \sum_{i=1}^K \mathbf{x}_i^b$).

Na Eq. 2.9, a média do conjunto ($\bar{\mathbf{x}}^b$) sobre K membros tende a enviesar a variância do erro de previsão porque todos os membros são utilizados na estimativa de sua própria covariância (KALNAY, 2003). Logo, a estimativa das covariâncias é feita com base na normalização em $K - 1$ membros a fim de se permitir que a estimativa da variância não seja feita com base em toda a população.

A média do conjunto de análises, a partir da Eq. 2.7, pode ser escrita como:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}_e[\mathbf{y}^o - \mathbf{H}(\bar{\mathbf{x}}^b)] \quad (2.10)$$

A covariância do erro conjunto de previsões pode também ser convenientemente expressa da seguinte forma:

$$\mathbf{P}_e^b = (\mathbf{X}^{tb})(\mathbf{X}^{tb})^T \quad (2.11)$$

onde:

- \mathbf{X}^{tb} : é a matriz de perturbação do conjunto de K previsões;

$$\mathbf{X}^{tb} = \frac{1}{\sqrt{K-1}} (\mathbf{x}_1^b - \bar{\mathbf{x}}^b, \mathbf{x}_2^b - \bar{\mathbf{x}}^b, \mathbf{x}_3^b - \bar{\mathbf{x}}^b, \dots, \mathbf{x}_k^b - \bar{\mathbf{x}}^b)$$

- \mathbf{X}^b : é a matriz do conjunto de K previsões.

$$\mathbf{X}^b = \frac{1}{\sqrt{K-1}}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_k)$$

Além de uma estimativa da covariância dos erros de previsão conjunto dada pela Eq. 2.11, o EnKF fornece também uma estimativa da covariância dos erros do conjunto das análises:

$$\mathbf{P}_e^a = (\mathbf{I} - \mathbf{K}_e \mathbf{H}) \mathbf{P}_e^b (\mathbf{I} - \mathbf{K}_e \mathbf{H})^T \quad (2.12)$$

onde:

- \mathbf{P}_e^a : é a estimativa da covariância dos erros do conjunto de análises (dimensão: $m \times m$);
- \mathbf{K} : é a matriz ganho de Kalman do conjunto;
- \mathbf{P}_e^b : é a estimativa da covariância dos erros do conjunto de previsões ($m \times m$);

2.2.4 *Ensemble Square Root Filter (EnSRF)*

Um conjunto de previsões muito pequeno pode fazer com que o EnKF subestime a covariância do erro da análise (Eq. 2.6) devido a amostragem insuficiente da população (O'KANE; FREDERIKSEN, 2008). Uma forma de se melhorar o espalhamento (i.e., a medida da diferença entre os membros) do conjunto é através da perturbação das observações assimiladas em conjunto com a inflação do conjunto para aumentar o espalhamento. Outra possibilidade é escolher os modos de crescimento mais rápidos para perturbar a condição inicial e formar um conjunto de previsões (e.g., utilizando-se o método de Zhang e Krishnamurti (1999), em que os modos de perturbação são obtidos a partir da aplicação de Funções Ortogonais Empíricas). Alternativamente, outros tipos de filtros podem ser determinados com o objetivo de se amenizar problemas de amostragem. Exemplos de filtros alternativos, incluem os filtros do tipo raiz quadrada, que não requerem que as observações sejam perturbadas.

Um filtro do tipo raiz quadrada, pode então ser determinado com a finalidade de ser numericamente mais estável e preciso, além de não depender diretamente da estimativa das covariâncias dos erros de previsão (\mathbf{P}^b) na estimativa das covariâncias

dos erros de análise (\mathbf{P}^a). Para o caso do FK linear, o exemplo a ser dado é o algoritmo de Potter (POTTER; STERN, 1963) para um filtro do tipo raiz quadrada, em que \mathbf{P}^b é substituída pela sua raiz quadrada. Segundo Tippett et al. (2003), a não unicidade da raiz quadrada de uma matriz de covariâncias (pois diferentes conjuntos de previsões podem ter a mesma matriz de covariâncias) permite que diferentes algoritmos sejam desenvolvidos.

A implementação do filtro de Kalman por conjunto dentro do sistema GSI, segue a implementação do filtro tipo raiz quadrada por conjunto (*Ensemble Square Root Filter* - EnSRF, Whitaker e Hamill (2002)). Nesta implementação é possível alternar entre uma versão que se utiliza de perturbações adicionadas as observações (filtro estocástico EnKF) e uma versão que não perturba as observações (filtro determinístico EnSRF).

As equações de correção e ganho do EnKF são utilizadas para determinar o conjunto de análise e a estimativa da covariância do erro da análise para o EnSRF. Entretanto, no EnSRF utiliza-se a equação de perturbação do conjunto de análises (Eq. 2.17) dado em função da matriz ganho de Kalman e da matriz de perturbações do conjunto de previsões:

$$\alpha = \left[1 + \sqrt{\frac{R}{(\mathbf{H}\mathbf{P}_e^b\mathbf{H}^T + R)}} \right]^{-1} \quad (2.13)$$

$$\mathbf{K}_e = \mathbf{P}_e^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_e^b\mathbf{H}^T + \mathbf{R})^{-1} \quad (2.14)$$

$$\tilde{\mathbf{K}}_e = \alpha\mathbf{K}_e \quad (2.15)$$

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \tilde{\mathbf{K}}_e[\mathbf{y}^o - \mathbf{H}(\mathbf{x}_k^b)] \quad (2.16)$$

$$\mathbf{X}^{1a} = \mathbf{X}^{1b} - \tilde{\mathbf{K}}_e\mathbf{H}\mathbf{X}^{1b} \quad (2.17)$$

onde:

- \mathbf{X}^{1a} : é a perturbação (desvios de cada membro em relação a média) do conjunto de K análises;
- \mathbf{X}^{1b} : é a perturbação do conjunto de K previsões;
- \mathbf{K}_e : é a matriz ganho de Kalman (definida da mesma forma que na Eq. 2.8);

- $\tilde{\mathbf{K}}_e$: é a matriz ganho utilizada para atualizar o conjunto de perturbações do conjunto de análises;
- R : é um escalar representando o erro da observação assimilada;
- \mathbf{H} : é o operador observação linear;
- \mathbf{P}^b : é a matriz de covariância dos erros do conjunto de previsões;
- α : é um escalar.

Uma das características desta implementação do EnSRF dentro do sistema GSI, é a assimilação serial das observações, i.e., cada tipo de observação é assimilada por vez. Este procedimento pode ainda ser feito assimilando-se as observações em uma ordem diferente da outra (o que conseqüentemente pode levar a diferentes estados de balanço entre modelo e observações, com maior ou menor grau de erro da análise, ou ainda menor tempo computacional na assimilação [Nerger \(2015\)](#)). Nesse sentido, quando o primeiro tipo de observação é assimilado, o estado atualizado do conjunto de análises torna-se o estado do conjunto de previsões para o próximo tipo de observação a ser assimilado. Por este motivo, R está indicado na Eq. 2.13 ao invés de \mathbf{R} .

Na Eq. 2.15, o escalar α modula o ganho de Kalman na atualização das perturbações do conjunto de análise de acordo com o erro do tipo de observação assimilada.

Ao final do processo de atualização dos membros do conjunto de análises, a estimativa da covariância dos erros do conjunto de análises é dada pela Eq. 2.6.

2.3 Dados

Os dados a serem utilizados para a realização dos experimentos com a nova matriz de covariâncias e com o sistema híbrido 3DVar, são os mesmos utilizados na operação do G3DVAR. O conjunto de dados é constituído por dados de temperatura da superfície do mar e dados de cobertura de neve em ponto de grade (dados obtidos do NCEP) e análises do GSI no formato espectral (global, resolução TQ0062L028 e TQ0299L064), - estes dados são utilizados pelo pré-processamento do modelo de circulação geral do CPTEC; dados de observação no formato PrepBUFR (obtidos também através do NCEP), - estes dados são utilizados para assimilação pelo GSI e compreendem dados de observações convencionais (de superfície e ar superior) além de vento por satélite. Dados não convencionais no formato BUFR (e.g., dados de radiancias dos sensores AMSU/A, HRS4, IASI, AIRS e MHS, além de dados de refratividade de Rádio Ocultação GPS) com cobertura global. Além disso, são utilizados também dados

de correção do viés da massa e do ângulo dos satélites. Estes dados são atualizados ao longo do ciclo de assimilação de dados (com excessão dos dados para a correção da massa e ângulo dos satélites). O filtro de Kalman por conjunto (EnKF/EnSRF) também utiliza o mesmo conjunto de observações que a componente variacional do sistema GSI (Tabela 2.2), sendo este comparável a assimilação operacional em outros centros (e.g., NCEP).

Tabela 2.2 - Observações tipicamente assimiladas pelo sistema GSI no CPTEC.

Mnemônico	Tipo	Descrição
airsbufr	Não Convencional	Radiâncias do AMSU-A e AIRS do satélite AQUA
amsuabufr	Não Convencional	Radiâncias do AMSU-A 1b (temperatura de brilho) dos satélites NOAA-15, 16, 17, 18 e 19 e METOP-A
hirs4bufr	Não Convencional	Radiâncias do HIRS4 1b dos satélites NOAA 18, 19 e METOP-A
mhsbufr	Não Convencional	Sondagens de umidade do MHS dos satélites NOAA 18, 19 e METOP-A
gpsrobufr	Convencional	Refratividade de Rádio Ocultação GPS
iasibufr	Não Convencional	Sondagens do IASI do satélite METOP-A
prepbufr	Convencional	Observações convencionais incluindo <i>ps</i> , <i>t</i> , <i>q</i> , <i>pw</i> , <i>wv</i> , <i>spd</i> , <i>dw</i> e <i>sst</i> a partir de plataformas de observação como o METAR, SYNOP, radiossondas e outras

3 MATRIZ DE COVARIÂNCIAS DOS ERROS DE PREVISÃO

Na matriz de covariâncias dos erros de previsão estão representadas as relações espaço-temporais (dentro de um período fixo no tempo) da variação dos erros de uma determinada variável e as relações que esta tem com outras variáveis a serem analisadas pelo sistema de assimilação de dados. As covariâncias dos erros de previsão representam, portanto, uma espécie de climatologia dos erros da previsão do modelo. Diferentes técnicas foram propostas para a determinação destas covariâncias, sendo a mais largamente utilizada aquela denominada de método *National Modeling Center* (NMC), introduzida por Parrish e Derber (1992). Outras técnicas incluem o método observacional (HOLLINGSWORTH; LÖNNBERG, 1986) e mais recentemente, o método por conjunto (FISHER, 2003). O método de Hollingsworth e Lönnberg (1986) é baseado nas medidas do termo de inovação (i.e., $\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)$) e é portanto dependente das observações, além do modelo. Se as observações consideradas forem exatamente as mesmas representadas pelo modelo, então o operador observação é apenas um interpolador (na Seção 4.4 é dado um exemplo de como o operador H é aplicado no cálculo da análise do vento horizontal). Do contrário, o operador observação deve incluir também as relações físicas que permitem que estas quantidades possam ser relacionadas, portanto, tornando este método mais complicado e susceptível a erros. Por outro lado, ao invés de se considerar a diferença entre observações e previsões de curto prazo (e.g., 6 horas), o método NMC utiliza a diferença entre duas previsões (48 e 24 horas). Neste caso, não há necessidade de se utilizar um operador observação, visto que as quantidades avaliadas são oriundas do mesmo modelo. Uma das vantagens deste método em relação ao método observacional está, portanto, no fato de não serem necessárias observações para o cálculo das covariâncias. Consequentemente, a determinação das covariâncias para toda a grade global pode ser prejudicada sobre as regiões onde não se tem muitas observações (e.g., todo o HS, em relação ao HN), o que não deve acontecer com o método NMC, uma vez que este utiliza previsões para esta estimativa. O método por conjunto é semelhante ao método NMC no sentido do uso das informações do modelo para a estimativa das covariâncias. Porém, ao invés de utilizar a diferença entre pares de previsões, utiliza um conjunto delas para, de forma semelhante a Eq. 2.5.

Cada um destes métodos assume diferentes suposições, mas todos lidam com o fato de que não se conhece o verdadeiro estado da atmosfera e que uma amostragem deste verdadeiro estado é impossível.

3.1 Importância da Matriz \mathbf{B}

Para se verificar a importância da matriz de covariâncias dos erros de previsão (\mathbf{B}), pode-se tomar por base a Eq. 2.1, supondo-se apenas uma observação e um ponto de grade a ser analisado e que o operador observação seja linear (\mathbf{H}):

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[\mathbf{y}^o - \mathbf{H}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\mathbf{x})] \quad (3.1)$$

Neste caso mais simples, consideramos o operador observação como sendo,

$$\mathbf{H} = [0, \dots, 0, 1, 0, \dots, 0] \quad (3.2)$$

Partindo-se da Eq. A.20 com o operador observação linear, obtém-se:

$$0 = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)] \quad (3.3)$$

Resolvendo-se a Eq. 3.3 para o incremento de análise ($\mathbf{x}^a - \mathbf{x}^b$), obtemos:

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) = (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)] \quad (3.4)$$

$$\mathbf{x}^a - \mathbf{x}^b = \frac{\mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)]}{\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}} \quad (3.5)$$

Multiplicando-se e dividindo-se por \mathbf{B} o lado direito da Eq. 3.5, obtemos:

$$\mathbf{x}^a - \mathbf{x}^b = \frac{\mathbf{B} \mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)]}{1 + \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}} \quad (3.6)$$

$$\mathbf{x}^a - \mathbf{x}^b = \frac{\mathbf{B} \mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)]}{\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T} \quad (3.7)$$

$$\mathbf{x}^a - \mathbf{x}^b = \frac{\mathbf{B} \mathbf{H}^T [\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)]}{\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T} \quad (3.8)$$

Como a suposição inicial foi a de que há apenas uma observação e apenas um ponto de grade a ser analisado, os termos $\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)$ e $\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T$ são escalares. Com

isso, pode-se afirmar que:

$$\mathbf{x}^a - \mathbf{x}^b \propto \mathbf{B}\mathbf{H}^T \quad (3.9)$$

Ou seja, o incremento de análise $\mathbf{x}^a - \mathbf{x}^b$ é proporcional a uma coluna da matriz \mathbf{B} . O papel da matriz de covariâncias dos erros de previsão é o de localizar o incremento de análise espalhando a informação das observações e garantir que estes incrementos sejam dinamicamente consistentes nas dimensões horizontal e vertical.

3.2 Método *National Modeling Center* (NMC)

Para a aplicação do sistema híbrido 3DVar no contexto deste trabalho, foi determinada uma matriz de covariâncias para uso nos experimentos de acordo com a resolução escolhida. O método utilizado para a determinação da matriz de covariâncias, é o método conhecido como *National Modeling Center* (NMC) (PARRISH; DERBER, 1992). Este método foi escolhido pelas seguintes razões:

- a) É um método relativamente simples de ser empregado, pois requer apenas a utilização das previsões de 24 e 48 horas do modelo;
- b) Pode ser empregado para calcular a matriz de covariâncias apenas para o período de estudo proposto (não requer uma base de dados de previsão excessivamente extensa);
- c) É o método geralmente escolhido para calcular a matriz de covariâncias do GSI no NCEP e em outros centros operacionais que utilizam este sistema.

O método NMC supõe que a correlação espacial dos erros de previsão são semelhantes as correlações temporais entre as diferenças das previsões de 48 e 24 horas. Ou seja, a ideia principal do método NMC é escolher os erros que crescem mais rápido a partir das diferenças entre as previsões de 48 e 24 horas para uma grade global (ou entre as previsões de mais curto prazo para uma grade regional, considerando que a taxa de crescimento dos erros pode ser mais acentuada devido a erros na representação de processos de subgrade). A seguir será apresentado um algoritmo para o cálculo da matriz de covariâncias dos erros do previsão.

O cálculo da matriz de covariâncias dos erros de previsão para o modelo de circulação geral da atmosfera do CPTEC (MCGAv4 e BAMv0), segue um algoritmo semelhante ao proposto pela ferramenta “gen_be” (RIZVI et al., 2009) desenvolvida pelo

Developmental Testbed Center (DTC) do *National Center for Atmospheric Research* (NCAR), com a diferença de que este algoritmo não considera as variáveis de controle conteúdo de água líquida (cw), ozônio (oz) e temperatura da superfície do mar (sst), que são necessárias para o cálculo para um modelo espectral global. O algoritmo do programa “gen_be” foi criado para calcular as estatísticas dos erros de previsão do modelo Weather Research and Forecasting Model (WRF), sendo empregado na assimilação de dados com o sistema variacional do *Weather Research and Forecasting model Data Assimilation* (WRFDA). Apesar disso, o algoritmo principal para o cálculo da matriz de covariâncias para os casos regional e global, é fundamentalmente o mesmo. Portanto, o algoritmo aqui apresentado a partir do programa “gen_be” servirá como exemplo para o cálculo da matriz \mathbf{B} da componente variacional do GSI.

O cálculo da matriz de covariâncias dos erros de previsão utilizando-se o método NMC, envolve as seguintes variáveis: velocidade potencial (χ), função de corrente (ψ), temperatura do ar (T), umidade relativa (q) e pressão a superfície (ps). O algoritmo original utiliza as componentes u e v do vento para calcular as quantidades divergência (D) e vorticidade (ζ) a partir das Eqs. 3.10 e 3.11 a seguir:

$$D = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \quad (3.10)$$

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad (3.11)$$

Seguindo-se a equação de Poisson (Eqs. 3.12 e 3.13), são calculadas as quantidades função de corrente (ψ) e velocidade potencial (χ):

$$\nabla^2 \psi = \zeta \quad (3.12)$$

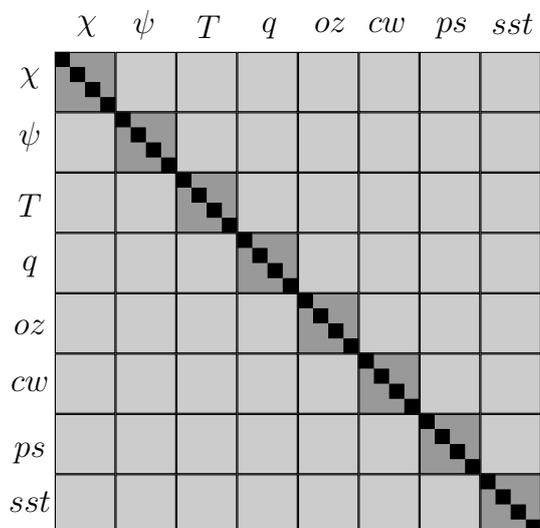
$$\nabla^2 \chi = D \quad (3.13)$$

Para o caso do modelo WRF, estes passos iniciais são necessários, uma vez que o modelo não escreve estas variáveis nos seus arquivos de previsão (e que são utilizados pelo software “gen_be”). Para o caso do modelo de circulação geral do CPTEC, este primeiro passo não é necessário uma vez que as quantidades velocidade potencial (χ) e função de corrente (ψ) são prognosticadas pelo modelo.

A seguir, o programa “gen_be” calcula os coeficientes de regressão para a velocidade potencial, temperatura e pressão em superfície com relação a função de corrente. Estes coeficientes são calculados para posteriormente montar o campo de velocidade potencial desbalanceado (parte rápida do fluxo - e.g., ondas de gravidade), temperatura desbalanceada e pressão em superfície desbalanceada. Estes campos são obtidos subtraindo-se a parte balanceada (parte lenta do fluxo - eg., onda de Rossby) do campo completo ou total de cada variável. Depois, as variâncias e os comprimentos de escala (horizontal e vertical) são calculados para cada uma das variáveis de análise. Os comprimentos de escala horizontal e vertical são bastante importantes por representarem o papel da localização das covariâncias dos erros de previsão a serem utilizadas no cálculo dos incrementos de análise.

Como resultado, depois que o arquivo com os desvios padrões, amplitudes e comprimentos de escalas é escrito, obtém-se uma matriz do tipo bloco-diagonal - representada de forma idealizada pela Figura 3.1) composta pelas variâncias de cada variável de estado na diagonal principal (no GSI estas variáveis são velocidade potencial (χ), função de corrente (ψ), temperatura (T), umidade (q), ozônio (oz), conteúdo de água de nuvens (cw), pressão em superfície (ps) e temperatura da superfície do mar (sst). As submatrizes que compõem a matriz \mathbf{B} representam autocovariâncias e covariâncias entre as variáveis.

Figura 3.1 - Morfologia idealizada da matriz de covariâncias dos erros de previsão (**B**).



No caso do GSI, as variáveis de estado analisadas são representadas: na diagonal principal os quadrados pretos representam as variâncias de cada variável; nos quadrados em cinza escuro, estão representados os elementos de autocovariância e em cinza claro, os elementos de covariância multivariados.

Fonte: Adaptado de [Petrie \(2012\)](#)

Para o cálculo da matriz de covariâncias do sistema GSI a partir do modelo de circulação geral do CPTEC, foram feitos dois testes. O primeiro teste considerou os pares de previsões de 48 e 24 horas que já estavam preparados na resolução TQ0299L064 (experimento denominado “TAG” - este experimento não faz parte do escopo deste trabalho) e extraídos do ciclo de assimilação de dados do sistema G3DVAR (versão 1.1.3) utilizando o modelo MCGAv4. O segundo teste considerou os pares de 48 e 24 horas do modelo BAMv0 na resolução TQ0062L028. Neste segundo teste, os pares de previsões foram obtidos a partir da realização do modelo BAMv0 usando as análises do NCEP. As justificativas para que o segundo teste tenha sido feito a com as análises do NCEP, são as seguintes: 1) foram encontradas inconsistências na representação das quantidades *oz* e *cw* na matriz de covariâncias quando esta foi calculada com base no MCGAv4; 2) embora seja possível calcular uma matriz de covariâncias para uma resolução menor do que a dos pares de entrada, estes pares não apresentaram valores consistentes de *oz* e *cw*, fato este que não foi amenizado com a troca de resolução; 3) o problema com as quantidades *oz* e *cw* foi temporariamente resolvido utilizando-se as quantidades de *oz* e *cw* da matriz de covariâncias do NCEP,

porém os primeiros testes do sistema híbrido 3DVar na resolução TQ0299L064 foram desencorajadores devido ao elevado custo computacional (e.g., dificuldades em alocar os recursos necessários para a realização dos membros do conjunto em tempo próximo do operacional); 4) não havia pares de previsões de 48 e 24 horas disponíveis para o MCGAv4 na resolução TQ0062L028 e portanto optou-se por testar uma versão mais recente do modelo de circulação geral do CPTEC (neste caso o modelo BAMv0) na tentativa de se sanar o problema com a representação de *oz* e *cw* (sem a necessidade de se copiar estas quantidades da matriz do NCEP) e facilitar a realização do sistema híbrido 3DVar.

A matriz de covariâncias resultante do segundo teste foi então escolhida para a realização dos experimentos com o sistema híbrido 3DVar (cujos resultados são apresentados no Capítulo 5). Nesta seção, portanto, são apresentadas as características principais das matrizes de covariâncias calculadas (nos dois testes mencionados), pois a matriz de covariâncias estática faz parte da metodologia aplicada ao longo deste trabalho.

O cálculo da matriz de covariâncias na resolução TQ0299L064 com os pares de previsões do modelo MCGAv4, foi feito com 1460 pares de previsões de 48 e 24 horas distribuídos ao longo de 1 ano (2013) no formato espectral. Os pares de previsões foram organizados na forma a seguir (válido também para a matriz de covariâncias do teste com o BAMv0): considerando-se o dia 2014010100, o primeiro par de previsões válido foi gerado com as análises dos dias 2013123100 (válido para uma previsão de 24 horas) e 2013123000 (válido para uma previsão de 48 horas).

O algoritmo para o cálculo da matriz de covariâncias, cumpre as seguintes etapas:

- a) Leitura e organização dos pares de previsões de 48 e 24 horas;
- b) Remoção de viés (em toda a coluna vertical);
- c) Cálculo das matrizes de balanço que permitirão as transformações entre função de corrente (ψ) e as componentes desbalanceadas de velocidade potencial (χ), pressão em superfície (p) e temperatura (T);
- d) Cálculo das variâncias dos erros de cada uma das variáveis de controle (ψ , χ , q , oz , cw , p);
- e) Cálculo dos comprimentos de correlação verticais (em unidades inversas em ponto de grade);

f) Cálculo dos comprimentos de correlação horizontais (em km).

No caso da matriz de covariâncias do MCGAv0 (na resolução TQ0299L064), nas etapas d), e) e f), as quantidades *oz* e *cw* foram posteriormente substituídas pelas quantidades correspondentes contidas na matriz de covariâncias do NCEP.

3.3 Características Principais e Diagnósticos

Nesta seção são apresentadas as características principais da matriz de covariâncias calculada no teste com o MCGAv4 (mencionado na seção anterior). As estruturas da matriz de covariâncias são apresentadas em termos de suas amplitudes. Como foram utilizados todos os pares de previsões de 48 e 24 horas disponíveis para o ano de 2013 na resolução TQ0299L028, foram calculadas também versões da matriz de covariâncias exclusivas para cada um dos horários sinóticos padrão (i.e., 00, 06, 12 e 18Z). Logo, uma comparação é feita entre estas versões da matriz de covariâncias do CPTEC com as versões completas das matrizes de covariâncias do CPTEC e NCEP. O objetivo destas comparações é caracterizar de forma quantitativa e qualitativa as matrizes calculadas.

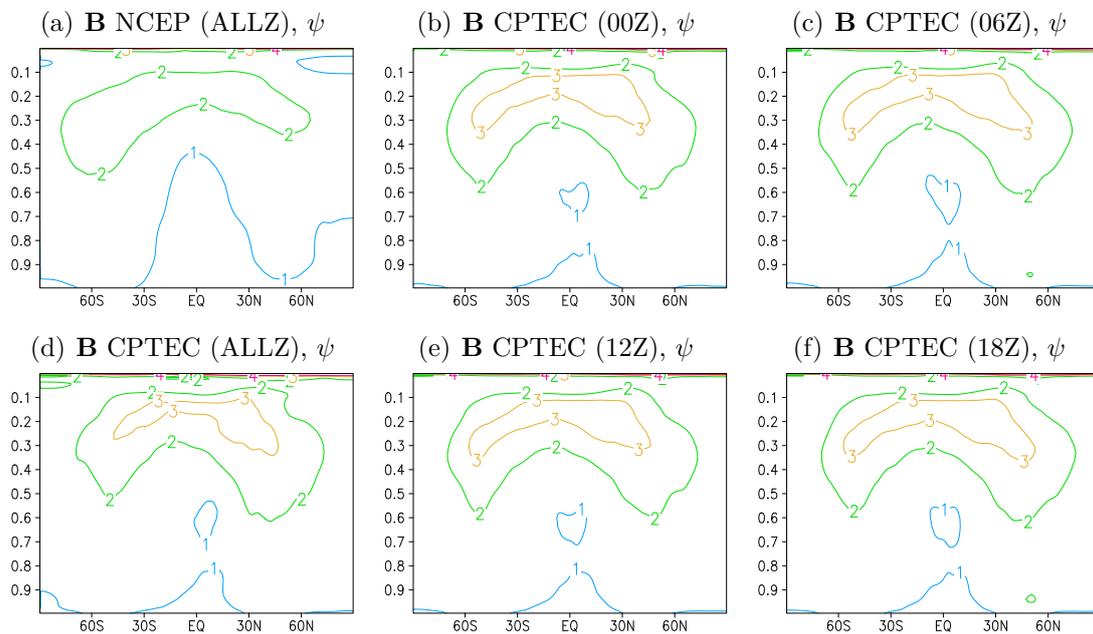
Características Principais da Matriz de Covariâncias do MCGA v4.x

A matriz de covariâncias do GSI possui as seguintes estruturas: amplitudes (representadas por variâncias e desvios-padrão), distâncias (representadas pelos comprimentos de escalas horizontal e vertical) e matrizes de projeção de balanço. A matriz de covariâncias utilizada pelo CPTEC (3DVar) varia apenas nas latitudes e na vertical. As variáveis de estado do GSI são a função de corrente, a velocidade potencial desbalanceada, a temperatura virtual desbalanceada, a pressão em superfície, a pseudo umidade relativa (ou umidade relativa normalizada), a razão de mistura de ozônio e a razão de mistura de condensação em nuvens. Para cada uma destas variáveis são definidas amplitudes e as distâncias. As variáveis velocidade potencial, temperatura e pressão em superfície, possuem matrizes de projeção que são responsáveis por projetar o incremento da função de corrente no perfil da parte balanceada do incremento de cada uma destas quantidades. Para a velocidade potencial, uma matriz de correlação é utilizada para contabilizar a correlação positiva entre divergência e vorticidade.

Para se verificar a qualidade das amplitudes calculadas a partir dos pares de previsões do modelo de circulação geral MCGAv4 na resolução TQ0299L064, foram calculadas as amplitudes para os horários sinóticos padrão (00, 06, 12 e 18Z) e para todos os horários juntos (“AllZ”), para o período de 1 ano. Para este período de tempo,

foi utilizado um total de 1460 pares (contabilizando todos os horários sinóticos) e 365 pares de previsões para as matrizes exclusivas de cada horário sinótico. Para comparação, foi utilizada a matriz de covariâncias do NCEP, calculada utilizando-se os pares de previsão do modelo GFS. Segundo Wu et al. (2002), esta matriz foi calculada utilizando-se 49 pares de previsões, distribuídos ao longo de 1 ano. Em todos os casos, as comparações são feitas na grade de dimensões 768x386x64 (lon x lat x lev).

Figura 3.2 - Distribuição das amplitudes calculadas para a matriz **B** na resolução TQ0299L064.



Amplitudes da função de corrente (ψ , $\times 10^{-6}$) ao longo das latitudes e níveis sigma, calculadas utilizando-se o método NMC para todos os horários sinóticos e todos os horários separados. A comparação é feita com o NCEP.

Fonte: Produção do autor.

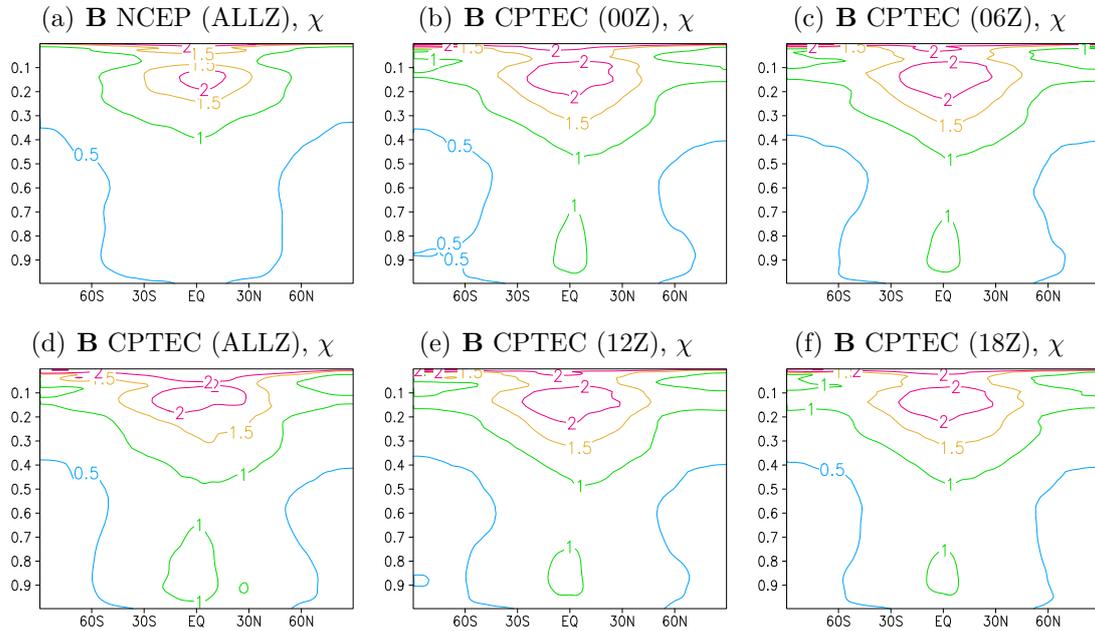
A Figura 3.2 mostra a distribuição vertical e latitudinal das amplitudes (variâncias/desvios-padrão) da função de corrente provenientes da matriz de covariâncias do NCEP e do CPTEC. Comparando-se as Figuras 3.3(a) e 3.3(d) (as duas matrizes contendo os pares de previsões de todos os horários sinóticos padrão), observa-se que a amplitude da função de corrente da matriz do CPTEC é mais concentrada, sendo confinada entre os níveis 0.3 a 0.1 sigma (aproximadamente entre

307.2 e 102.4 hPa). As isolinhas de amplitude da função de corrente da matriz do CPTEC, mostram valores maiores nesta região (3×10^{-6}) enquanto que a matriz do NCEP apresenta valores menores na mesma região (2×10^{-6}), mas espalhado entre os níveis sigma 0.5 e 0.1 (512 e 102.4). Além disso, pode-se observar também a distribuição da amplitude ao longo das latitudes (90N e 90S); nos trópicos ela é menor, apresentando um padrão em forma de “sino” (o qual pode ser melhor observado nas demais matrizes, calculadas com os pares de previsões dos horários sinóticos em separado) e maior nos polos.

Esta distribuição da amplitude (variância) da função de corrente para a matriz do CPTEC calculada utilizando-se todos os pares de previsão, não é exatamente semelhante a matriz correspondente do NCEP e esta comparação mostra que, de forma geral, a amplitude da função de corrente no Hemisfério Norte é maior na matriz do CPTEC (o valor em si é maior e a faixa de latitudes envolvida também), enquanto que esta mesma amplitude é menor no Hemisfério Sul, como representado na matriz de covariâncias do CPTEC. Isto pode estar relacionado a forma como os pares de previsão foram gerados: para o cálculo da matriz de covariâncias do CPTEC, foram utilizados pares de previsões de 48 e 24 horas provenientes do ciclo de assimilação de dados do CPTEC, utilizando-se observações regionalizadas (além daquelas provenientes do *Global Telecommunication System - GTS*). Consequentemente, a variância do erro sobre o Hemisfério Sul, sobretudo sobre a América do Sul (entre 0 e 60S) pode ser menor. Nas outras matrizes do CPTEC (calculadas com os pares exclusivos das 00, 06, 12 e 18Z), observa-se uma distribuição quase simétrica da amplitude da função de corrente, com valores mais próximos da matriz do NCEP, mas com a distribuição, mais localizada ao sul do continente sul-americano.

A Figura 3.3 mostra a distribuição das amplitudes da velocidade potencial. Assim como para a Figura 3.2, as amplitudes calculadas para a velocidade potencial mostram que sobre a região tropical (entre 30S e 30N), a variância do erro na matriz do CPTEC é maior do que a representada pela matriz do NCEP, sobretudo entre os níveis mais baixos do modelo até o nível sigma 0.6 (614 hPa). Comparativamente, cada uma das demais matrizes do CPTEC (00, 06, 12 e 18Z) apresentam similaridades entre si (principalmente entre as matrizes das 00 e 12Z e 06 e 18Z - Figuras 3.4(b) e 3.4(e); 3.4(c) e 3.4(f), respectivamente).

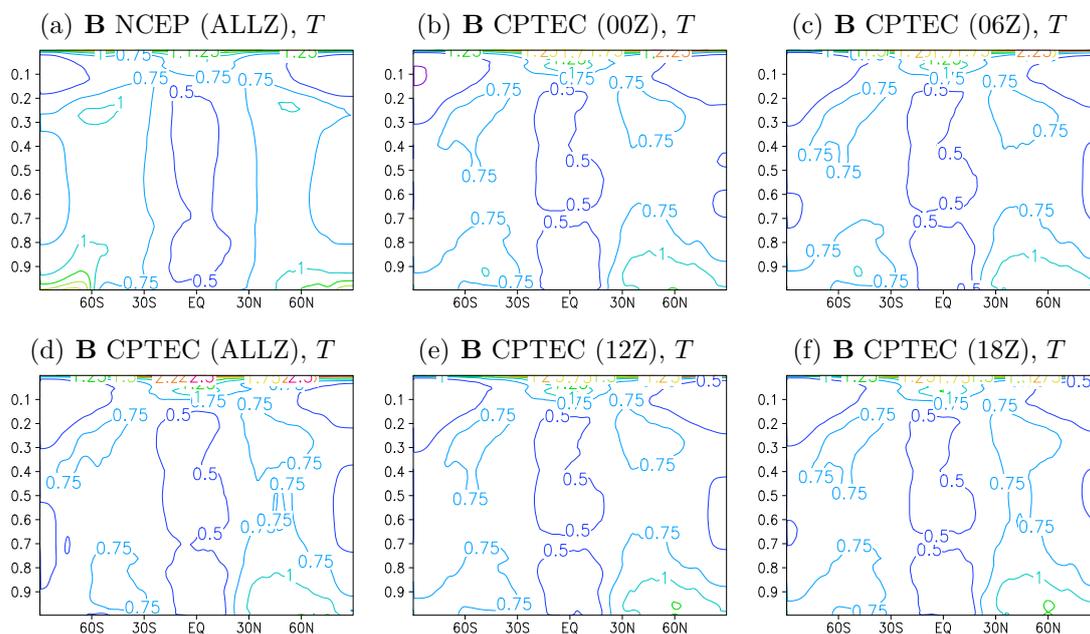
Figura 3.3 - Idem Figura 3.2, para a velocidade potencial ($\chi \times 10^{-6}$).



Fonte: Produção do autor.

A Figura 3.4 a seguir, apresenta a distribuição das amplitudes da temperatura, sendo que estes valores na matriz do CPTEC completa (i.e., BALLZ - Figura 3.5(b)), onde a maior parte da variância encontra-se concentrada sobre o Hemisfério Norte, mais precisamente a partir de 30N (isolinha de 1 na Figura 3.5(b)); esta mesma isolinha na matriz do NCEP (Figura 3.5(a)) está desenhada a partir de aproximadamente 45N. Em ambas as matrizes, esta amplitude é limitada as primeiras camadas do modelo (ie., entre a superfície e o nível sigma de 0.8 ou 819.2 hPa). Nas demais matrizes, observa-se um padrão semelhante em relação a distribuição das amplitudes, sendo que no Hemisfério Norte, estas alcançam os 819.2 hPa, enquanto que no Hemisfério Sul, não ultrapassam este nível.

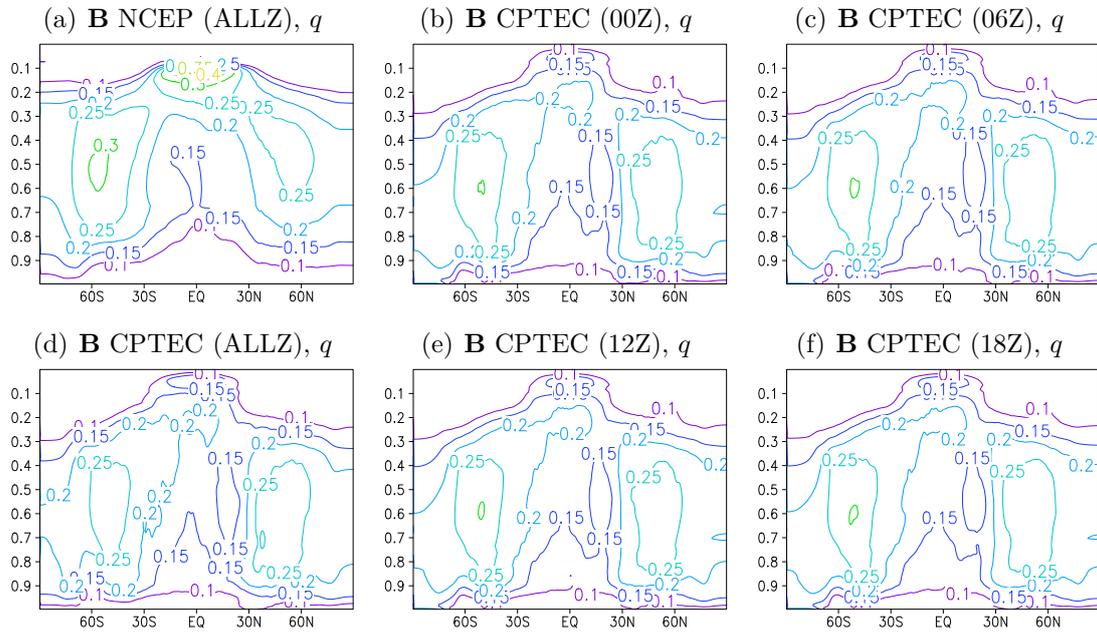
Figura 3.4 - Idem Figura 3.2, para a temperatura (T).



Fonte: Produção do autor.

As amplitudes da umidade apresentada na Figura 3.6(d) para a matriz CPTEC BALLZ, não difere muito em distribuição ao longo das latitudes como nas demais matrizes (i.e., 00, 06, 12 e 18Z - Figuras 3.6(b), 3.6(c), 3.6(e) e 3.6(f), respectivamente), mas a amplitude da variância nos subtropicais em ambos os hemisférios (entre 60S e 30S e entre 30N e 60N), é melhor definida e menos intensa também do que aquela representada em B NCEP (Figura 3.6(a)). Neste caso, a variância da umidade em ambos os hemisférios é bastante simétrica, principalmente nos casos calculados com as previsões do CPTEC. Isto pode reforçar a necessidade de se remover o viés do modelo do CPTEC. Neste caso, as isolinhas mais próximas a superfície (até o nível sigma 0.8 ou 819.2 hPa), apresentam valores com quase o dobro do valor, se comparados com a mesma região na matriz do NCEP. De outra forma, pode mostrar também alguma deficiência na física do modelo em representar com melhor precisão os processos úmidos (que dependem da física do modelo).

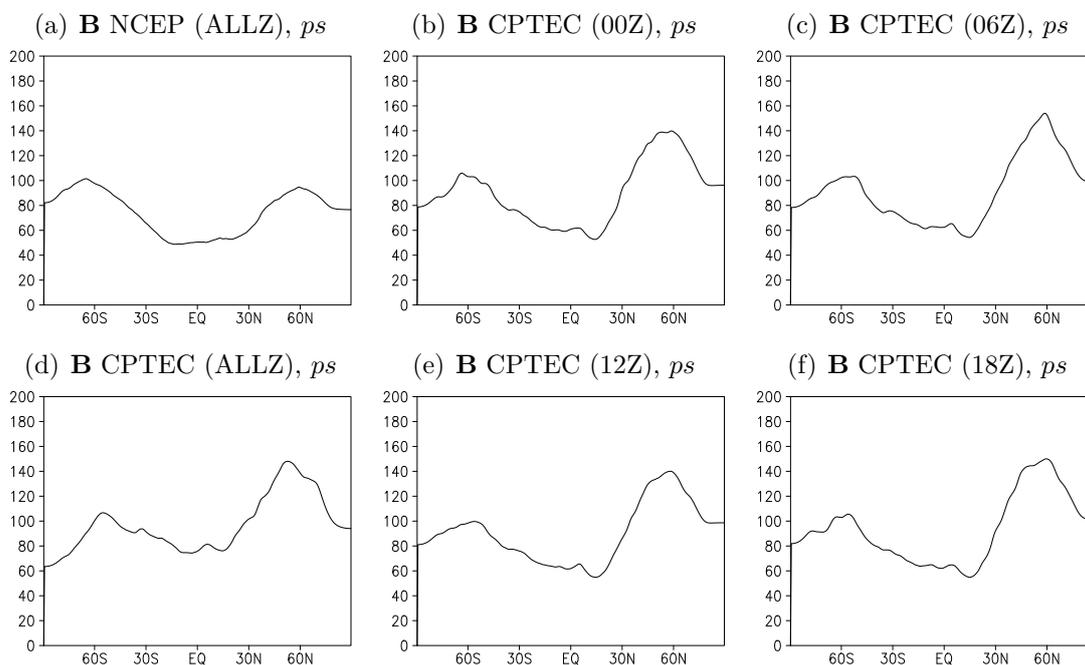
Figura 3.5 - Idem Figura 3.2, para a umidade (q).



Fonte: Produção do autor.

A Figura 3.6 apresenta a distribuição latitudinal das médias zonais das amplitudes da pressão em superfície. As amplitudes da matriz B NCEP (Figura 3.7(a)) apresentam uma distribuição simétrica, com dois picos sobre as latitudes de 60S e 60N. Nas matrizes do CPTEC (Figuras 3.7(b), 3.7(c), 3.7(d), 3.7(e) e 3.7(f)), a distribuição das amplitudes é assimétrica, sendo que os picos das amplitudes são maiores sobre a latitude de 60N do que sobre a latitude de 60S. A mesma amplitude apresentada em Wu et al. (2002), mostra que os picos da variância dos erros situam-se em torno destas latitudes, especialmente sobre o Hemisfério Norte, onde o pico local situa-se entre as latitudes de 30N e 60N.

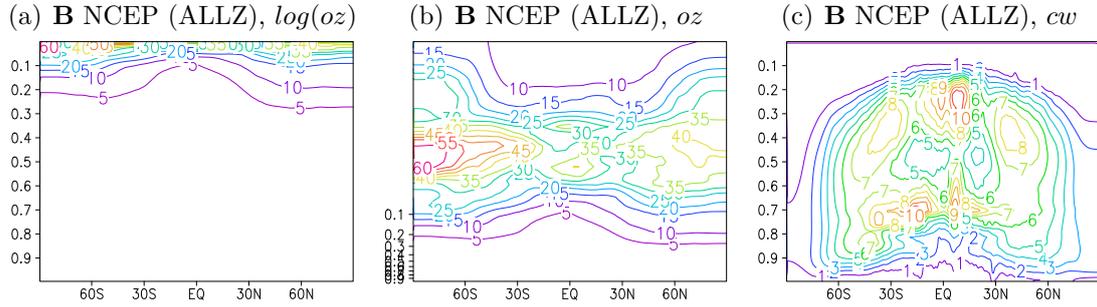
Figura 3.6 - Idem Figura 3.2, para a pressão em superfície (ps , $\times 10^2$).



Fonte: Produção do autor.

As Figuras 3.8(a), 3.8(b) e a Figura 3.8(c) mostram as amplitudes calculadas para as variáveis ozônio e razão de mistura de condensação em nuvem, respectivamente. O sistema G3DVAR (versão v1.1.3) não assimila observações e ozônio, e além disso, o ozônio utilizado pelo modelo de MCGAv4 utiliza uma climatologia do ozônio. As amplitudes calculadas para esta quantidade são bastante diferentes daquelas apresentadas pela matriz NCEP BA1Z (Figuras 3.8(a), 3.8(b)), e mostram valores muito mais altos do que qualquer matriz calculada utilizando-se os pares de previsões do CPTEC. Além disso, o MCGAv4 não representa a razão de mistura de condensação, como pode ser visto na Figura 3.8(c). Para melhorar a visualização da distribuição das amplitudes do ozônio, na Figura 3.8(a) estas amplitudes estão destacadas em escala logarítmica.

Figura 3.7 - Idem Figura 3.2, para o ozônio ($oz \times 10^7$) e o conteúdo de água ($cw \times 10^4$).



Fonte: Produção do autor.

Nas Tabelas 3.1, 3.2 e 3.3 são apresentados os valores máximos e mínimos das quantidades apresentadas nas Figuras 3.2, 3.3, 3.4, 3.5, 3.6 e 3.7 em adição as amplitudes e comprimentos de escalas da temperatura da superfície do mar.

Tabela 3.1 - Valores máximos e mínimos das amplitudes das versões das matrizes de covariâncias do CPTEC na resolução TQ0299L064 a partir dos pares de previsões de 48 e 24 horas do modelo MCGAv4.

	Bncep		Bcptec		B00Zcptec		B06Zcptec		B12Zcptec		B18Zcptec	
	Máx	Mín										
ψ	$7,99 \times 10^6$	$4,3 \times 10^5$	$2,8 \times 10^7$	$4,6 \times 10^5$	$1,7 \times 10^7$	$4,9 \times 10^5$	$1,9 \times 10^7$	$4,9 \times 10^5$	$1,7 \times 10^7$	$4,9 \times 10^5$	$1,7 \times 10^7$	$4,8 \times 10^5$
χ	$3,77 \times 10^6$	0	$1,07 \times 10^7$	0	$4,52 \times 10^6$	0	$3,74 \times 10^6$	0	$4,19 \times 10^6$	0	$4,94 \times 10^6$	0
T	2,23	0	4,04	0	3,01	0	3,11	0	3,13	0	3,03	0
q	0,41	$-4,06 \times 10^{-7}$	0,29	$-2,15 \times 10^{-7}$	0,3	$-2,12 \times 10^{-7}$	0,3	$-2,27 \times 10^{-7}$	0,3	$-2,59 \times 10^{-7}$	0,3	$-2,53 \times 10^{-7}$
oz	$6,4 \times 10^{-7}$	$-4,38 \times 10^{-7}$	$4,4 \times 10^{-7}$	$-2,2 \times 10^{-7}$	$6,4 \times 10^{-7}$	$-2,30 \times 10^{-7}$	$6,4 \times 10^{-7}$	$-2,33 \times 10^{-7}$	$6,4 \times 10^{-7}$	$-2,59 \times 10^{-7}$	$6,4 \times 10^{-7}$	$-2,63 \times 10^{-7}$
cw	$1,24 \times 10^{-4}$	$-4,4 \times 10^{-7}$	$1,24 \times 10^{-4}$	$-2,4 \times 10^{-7}$	$1,24 \times 10^{-4}$	$-2,3 \times 10^{-7}$	$1,24 \times 10^{-4}$	$-2,33 \times 10^{-7}$	$1,24 \times 10^{-4}$	$-2,59 \times 10^{-7}$	$1,24 \times 10^{-4}$	$-2,64 \times 10^{-7}$
ps	$1,01 \times 10^{-1}$	$4,88 \times 10^{-2}$	$1,48 \times 10^{-1}$	$6,35 \times 10^{-2}$	$1,4 \times 10^{-1}$	$5,27 \times 10^{-2}$	$1,54 \times 10^{-1}$	$5,43 \times 10^{-2}$	$1,4 \times 10^{-1}$	$5,48 \times 10^{-2}$	$1,5 \times 10^{-1}$	$5,48 \times 10^{-2}$
sst	$4,32 \times 10^{-1}$	2×10^{-1}										

Tabela 3.2 - Idem Tabela 3.1, para os comprimentos de escala horizontais.

	Bncep		Bcptec		B00Zcptec		B06Zcptec		B12Zcptec		B18Zcptec	
	Máx	Mín										
ψ	$1,3 \times 10^6$	$2,2 \times 10^5$	$1,75 \times 10^6$	$4,55 \times 10^5$	$1,32 \times 10^6$	$4,73 \times 10^5$	$1,4 \times 10^6$	$4,7 \times 10^5$	$1,31 \times 10^6$	$4,7 \times 10^5$	$1,27 \times 10^6$	$4,73 \times 10^5$
χ	$1,1 \times 10^6$	0	$1,82 \times 10^6$	0	$1,37 \times 10^6$	0	$1,45 \times 10^6$	0	$1,5 \times 10^6$	0	$1,42 \times 10^6$	0
T	$3,1 \times 10^5$	0	$5,9 \times 10^5$	0	$5,9 \times 10^5$	0	$5,9 \times 10^5$	0	$6,42 \times 10^5$	0	$6,37 \times 10^5$	0
q	$7,9 \times 10^6$	0	$2,22 \times 10^7$	0	$1,31 \times 10^7$	0	$1,48 \times 10^7$	0	$1,33 \times 10^7$	0	$1,36 \times 10^7$	0
oz	$7,99 \times 10^6$	0	$2,84 \times 10^7$	0	$1,73 \times 10^7$	0	$1,96 \times 10^7$	0	$1,75 \times 10^7$	0	$1,78 \times 10^7$	0
cw	$7,99 \times 10^6$	0	$2,84 \times 10^7$	0	$1,73 \times 10^7$	0	$1,96 \times 10^7$	0	$1,75 \times 10^7$	0	$1,78 \times 10^7$	0
ps	$1,89 \times 10^5$	$1,19 \times 10^5$	$4,18 \times 10^5$	$1,83 \times 10^5$	$3,9 \times 10^5$	$1,83 \times 10^5$	$3,9 \times 10^5$	$1,8 \times 10^5$	$3,8 \times 10^5$	$1,8 \times 10^5$	$3,9 \times 10^5$	$1,84 \times 10^5$
sst	8×10^2	$5,07 \times 10^2$										

Tabela 3.3 - Idem Tabela 3.1, para os comprimentos de escala verticais.

	Bncep		Bcptec		B00Zcptec		B06Zcptec		B12Zcptec		B18Zcptec	
	Máx	Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx	Mín
ψ	$9,1 \times 10^{-1}$	$8,3 \times 10^{-2}$	1,27	$8,41 \times 10^{-2}$	1,16	$8,52 \times 10^{-2}$	1,27	$8,59 \times 10^{-2}$	1,08	$8,54 \times 10^{-2}$	1,17	$8,54 \times 10^{-2}$
χ	1,22	0	1,5	0	1,85	0	1,74	0	1,78	0	1,65	0
T	1,75	0	1,67	0	1,68	0	1,63	0	1,62	0	1,68	0
q	$1,3 \times 10^6$	0	$1,75 \times 10^6$	0	$1,28 \times 10^6$	0	$1,33 \times 10^6$	0	$1,27 \times 10^6$	0	$1,23 \times 10^6$	0
oz	$1,3 \times 10^6$	0	$1,82 \times 10^6$	0	$1,37 \times 10^6$	0	$1,45 \times 10^6$	0	$1,47 \times 10^6$	0	$1,42 \times 10^6$	0
cw	$1,3 \times 10^6$	0	$1,82 \times 10^6$	0	$1,37 \times 10^6$	0	$1,45 \times 10^6$	0	$1,47 \times 10^6$	0	$1,42 \times 10^6$	0

3.3.1 Correlações entre B CPTEC X NCEP

Aproveitando-se do fato de que a matriz de covariâncias do CPTEC na resolução TQ0299L064 utilizando os pares de previsões de 48 e 24 horas do MCGAv4 foi calculada com dimensões de grade iguais as da matriz de covariâncias do NCEP, então é possível fazer uma comparação direta. A matriz de covariâncias global do NCEP - e que esteve em uso por um breve período na operação do sistema G3DVAR (v1.1.3) no CPTEC, foi calculada com base nos pares de previsões do modelo GFS. Ambos os modelos, MCGAv4 e GFS, são modelos espectrais e as mesmas variáveis foram utilizadas para o cálculo das matrizes de covariâncias. A comparação entre as duas matrizes foi feita a partir do diagrama de espalhamento considerando a distribuição das médias latitudinais nos 64 níveis de cada variável. Nas comparações feitas a seguir, foram consideradas as variáveis ψ , χ , T e q nas regiões Hemisfério Norte (HN - lons: 0-360; lats:20N-80N), Trópicos (TR - lons:0-360; lats:20S-20N) e Hemisfério Sul (HS - lons:0-360; lats: 80S-20S). O particionamento em regiões para esta comparação, tem como objetivo melhorar a identificação das correlações entre as variáveis indicadas. Para cada figura, foram calculadas a curva de ajuste, o erro quadrático (r^2) e o coeficiente de correlação de Pearson (r).

Na Figura 3.8, são apresentados diagramas de dispersão entre as matrizes de covariâncias do NCEP e CPTEC (todos os horários, AllZ), para as variáveis ψ , χ , T e q , sobre as regiões HN, TR e HS. Em geral, a correlação entre as quantidades consideradas das duas matrizes de covariâncias apresentam valores em torno de 0,9 (90%), o que reflete as semelhanças encontradas nas amplitudes apresentadas na seção anterior. Além disso, maiores correlações foram encontradas sobre o Hemisfério Sul, enquanto que sobre a região Tropical, foram encontradas as menores correlações. As variáveis que possuem a maior correlação são a função de corrente (ψ) e a velocidade potencial (χ) sobre o Hemisfério Sul (Figuras 3.9(c) e 3.9(f)). Entretanto, verificou-se também que as amplitudes de q apresentam os menores valores de correlação e maior erro nos

níveis mais altos do modelo. Isto indica que há uma discrepância na representação da umidade a partir do nível 22 (nível sigma 0,604374, aproximadamente 618,88 hPa) até o topo do modelo sobre o Hemisfério Norte. Sobre os Trópicos, as discrepâncias na representação das amplitudes da umidade, estão em níveis mais baixo. Para a umidade, a correlação sobre esta região é de 0,511 (aproximadamente 51% - Figura 3.9(k)). Entretanto, sobre o Hemisfério Sul, a correlação entre as amplitudes da umidade das duas matrizes é alta (aproximadamente 90%). Sobre esta região, as maiores diferenças estão nos níveis médios. Nas Tabelas 3.4 e 3.5, estão sumarizados os valores de correlação e erro quadrático calculados para todos os horários sinóticos individuais. Neste caso, ressalta-se que as matrizes calculadas para os horários das 00, 06, 12 e 18Z, foram correlacionadas para as regiões indicadas com a matriz de covariâncias do NCEP.

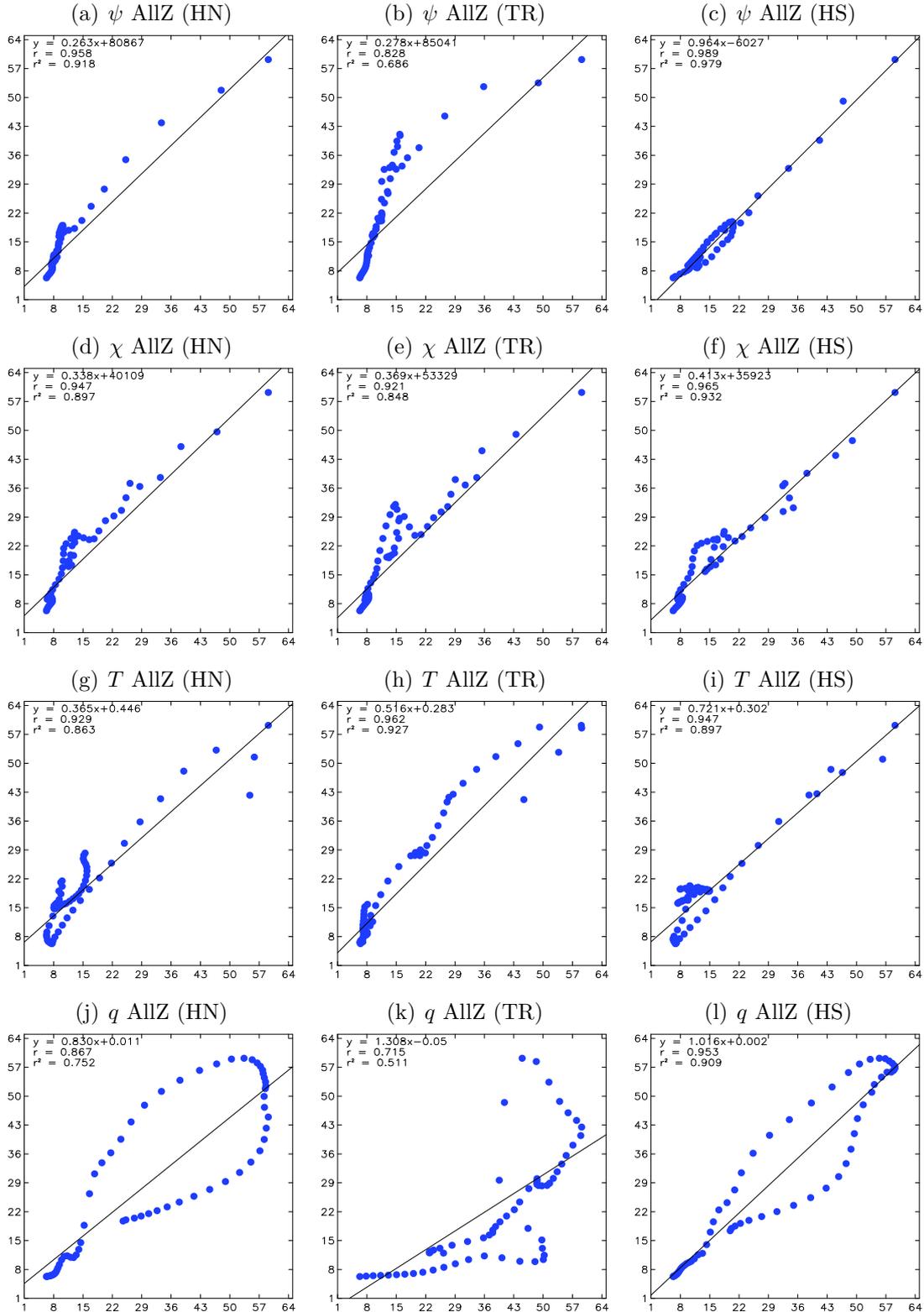
Tabela 3.4 - Coeficiente de Correlação de Pearson (r) e Erro Quadrático (r^2) para as amplitudes de ψ e χ , entre as matrizes **B** CPTEC e NCEP (TQ0299L064) sobre as regiões HN, TR e HS e para os horários AllZ (todos) e 00Z, 06, 12Z e 18Z. Em vermelho, estão destacados os valores de correlação menores do que 0,9.

	ψ						χ					
	HN		TR		HS		HN		TR		HS	
	r	r^2	r	r^2	r	r^2	r	r^2	r	r^2	r	r^2
00Z	0,972	0,945	0,925	0,855	0,953	0,908	0,979	0,959	0,977	0,955	0,972	0,946
06Z	0,966	0,933	0,934	0,873	0,962	0,925	0,976	0,952	0,972	0,946	0,974	0,950
12Z	0,975	0,951	0,940	0,885	0,963	0,927	0,978	0,956	0,977	0,955	0,970	0,941
18Z	0,977	0,955	0,932	0,870	0,965	0,931	0,971	0,944	0,978	0,957	0,973	0,947
AllZ	0,958	0,918	0,828	0,686	0,989	0,979	0,947	0,897	0,921	0,848	0,965	0,932

Tabela 3.5 - Idem Tabela 3.4, para as variáveis temperatura (T) e umidade (q).

	T						q					
	HN		TR		HS		HN		TR		HS	
	r	r^2	r	r^2	r	r^2	r	r^2	r	r^2	r	r^2
00Z	0,966	0,934	0,985	0,970	0,966	0,934	0,879	0,773	0,761	0,579	0,942	0,888
06Z	0,965	0,932	0,983	0,967	0,963	0,928	0,878	0,771	0,761	0,579	0,945	0,893
12Z	0,964	0,931	0,984	0,969	0,962	0,927	0,883	0,781	0,758	0,575	0,944	0,892
18Z	0,966	0,935	0,985	0,971	0,965	0,932	0,881	0,777	0,756	0,572	0,945	0,893
AllZ	0,929	0,863	0,962	0,927	0,947	0,897	0,867	0,752	0,715	0,511	0,953	0,909

Figura 3.8 - Diagramas de dispersão entre as amplitudes representadas nas matrizes de covariâncias **B** CPTEC e NCEP.



Diagramas de dispersão (níveis sigma), correlações e erros quadráticos entre as amplitudes representadas nas matrizes de covariâncias **B** CPTEC e NCEP em todos os horários (AllZ), para as regiões HN (coluna a esquerda), TR (coluna do meio) e HS (coluna da direita).

Fonte: Produção do autor.

3.4 Aspecto do Incremento de Análise

Como forma de verificação da influência e do impacto que a nova matriz de covariâncias do CPTEC exerce no incremento de análise do sistema G3DVAR, foi feito um experimento com a assimilação de uma única observação (sintética) no sistema. As Figuras 3.9 e 3.10 apresentam uma comparação entre os incrementos de análise produzidos pelas matrizes de covariâncias do CPTEC e do NCEP (respectivamente).

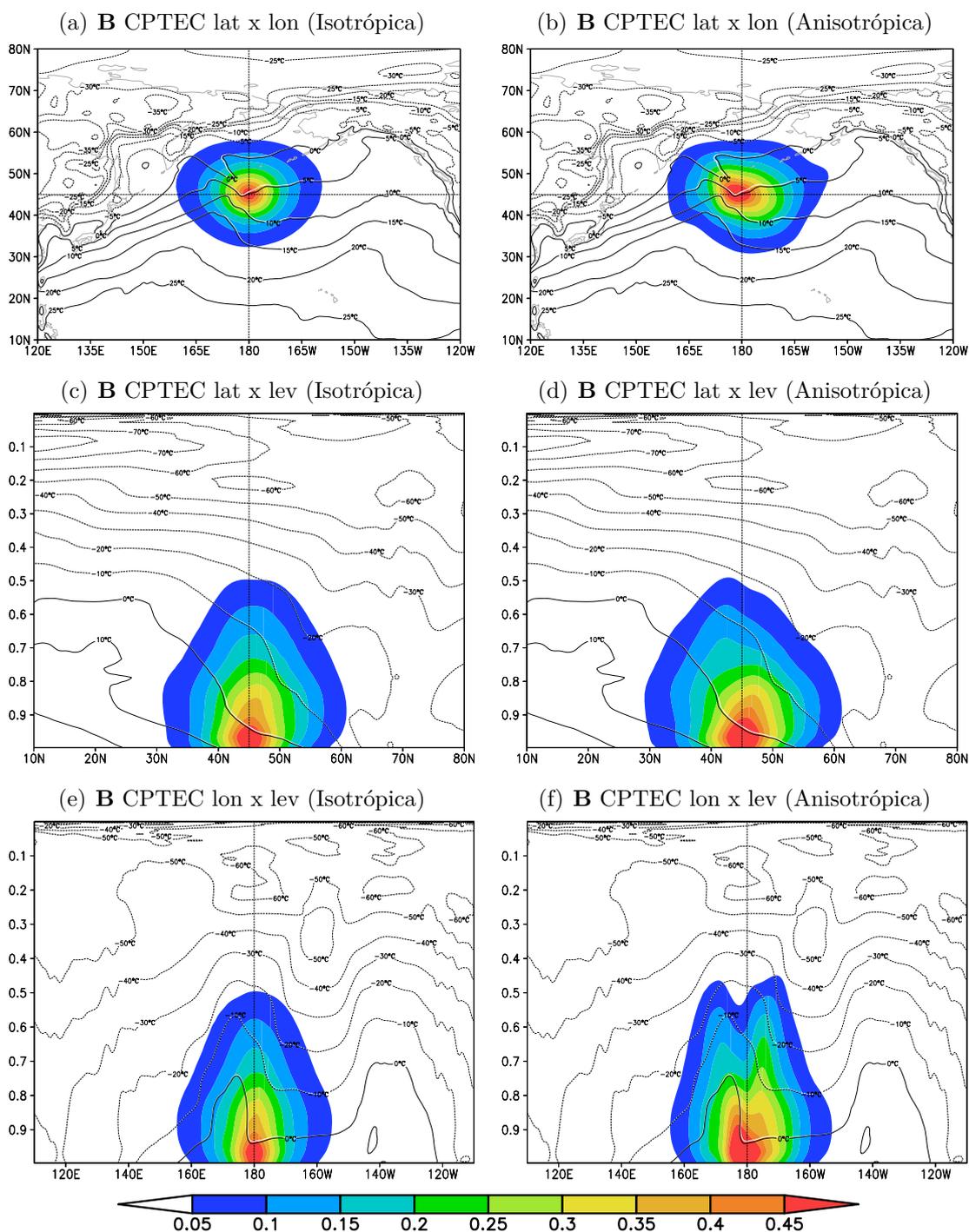
Na Figura 3.9, é mostrado o incremento de análise (Previsão subtraída da Análise, ou OmA) produzido pela assimilação de uma observação sintética de vento zonal em 1000hPa. Para isto, ajustou-se o erro da observação para 1 ms^{-1} , bem como a magnitude da inovação trazida pela observação (Previsão subtraída da Observação, ou OmF). Neste caso, o painel da esquerda mostra o incremento de análise isotrópico (em que os pesos dados para o Filtro Recursivo nas direções x e y são iguais), de forma que o incremento produzido possui aspecto quase igual em todas as direções. O formato achatado no topo evidencia apenas o fato de que a observação foi posicionada mais próxima ao polo norte, mais precisamente no ponto com coordenada 0 (longitude) e 45N (latitude). No painel da direita, o incremento de análise mostrado é anisotrópico, e os parâmetros principais ajustados no GSI para a sua produção, são apresentados na Tabela 3.6:

Tabela 3.6 - Parâmetros de configuração utilizados no sistema GSI para a aplicação das matrizes de covariâncias do CPTEC e do NCEP.

Parâmetro	Valor	Descrição
<i>vs</i>	0,7	Fator de escala do comprimento de correlação vertical
<i>hzscl</i>	1,7; 0,8; 0,5	Fator de suavização para as 3 escalas horizontais
<i>hswgt</i>	0,45; 0,3; 0,25	Pesos aplicados a cada escala horizontal
<i>bw</i>	0	Fator no cálculo da covariância do erro
<i>norsp</i>	4	Ordem de suavização horizontal das covariâncias do erro de previsão
<i>bkg_flowdep</i>	True/False	Flag para usar (True = anisotrópico) ou não (False = isotrópico) a dependência das variâncias do erro de previsão

O parâmetro “*vs*”, ajusta a escala vertical e sua unidade é dada em pontos de grade; o parâmetro “*hzscl*”, ajusta a escala horizontal e o parâmetro “*hswgt*” ajusta os pesos dados a aplicação do filtro recursivo nas direções x , y e z , respectivamente.

Figura 3.9 - Aspectos dos incrementos de análise utilizando a nova matriz de covariâncias do CPTEC.

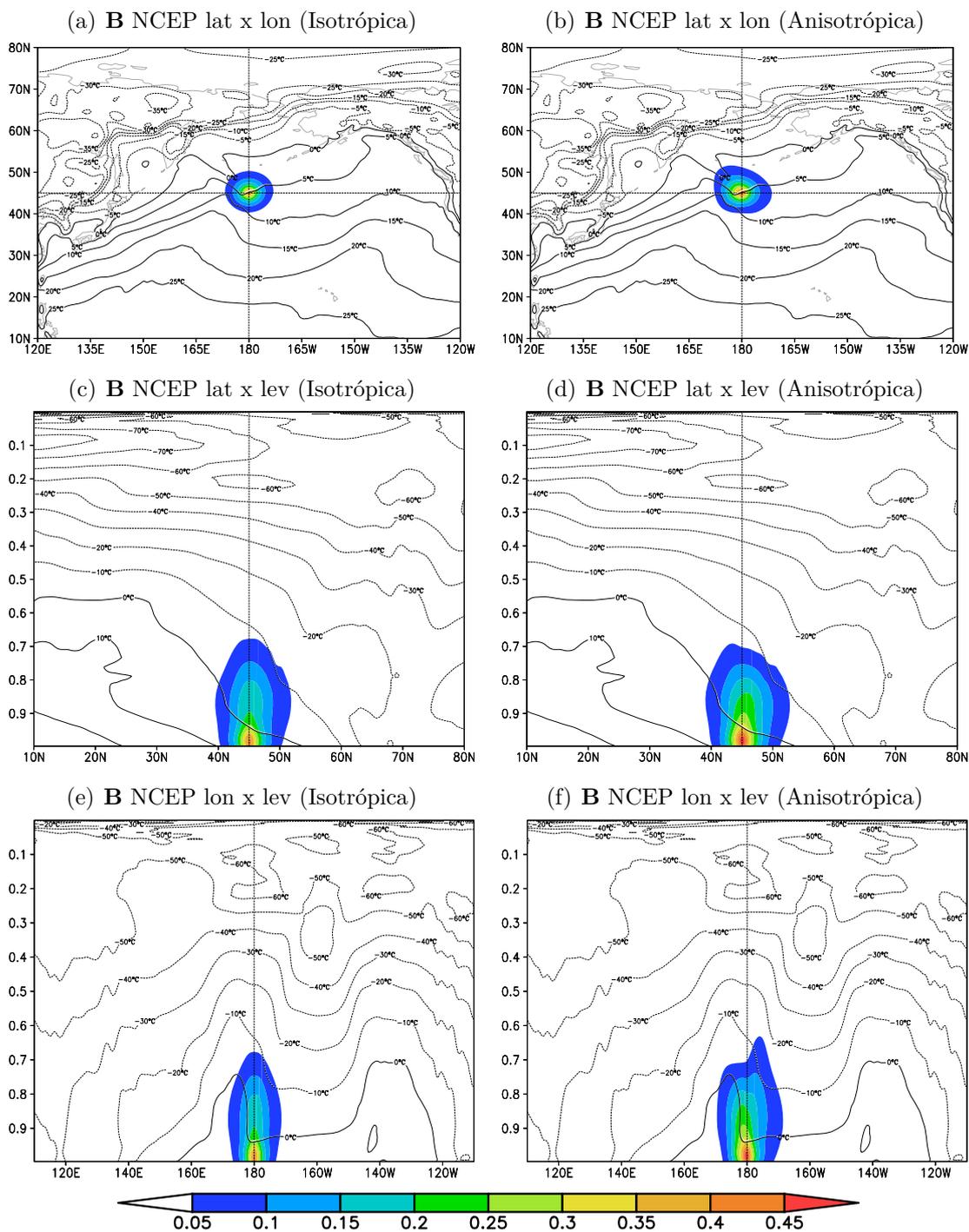


Incremento de análise da temperatura (T) ano nível de 1000 hPa, utilizando a nova matriz de covariâncias do CPTEC: a esquerda, incremento isotrópico; a direita incremento anisotrópico.

Fonte: Produção do autor.

A Figura 3.10 apresenta o incremento de análise produzido pela matriz de covariâncias do NCEP, aplicada ao sistema G3DVAR, assimilando a mesma observação sintética única, no mesmo nível e com as mesmas magnitudes de erro e inovação. De forma geral, para ambos os casos (isotrópico e anisotrópico), o incremento de análise calculado pela matriz de covariâncias do CPTEC (matriz CPTEC BALLZ), produz um incremento mais largo, abrangendo mais pontos de grade em todas as direções. Isso significa que o incremento de análise calculado, é espalhado sobre uma região maior ao redor da vizinhança da observação e que a sua influência pode ser transportada para mais longe. Em uma situação física em que, por exemplo há a incursão de um sistema frontal e, considerando-se o caso anisotrópico, a assimilação de dados pode trazer um benefício importante para a previsão gerada, porque a análise utilizada possuirá um balanço melhor entre observações e previsão. Isso mostra a importância da seleção das observações, do controle de qualidade e do *thinning* do sistema (i.e., o controle espacial das observações). Além disso, a magnitude do incremento de análise produzido pela matriz de covariâncias do CPTEC é mais suave, embora seja menos concentrado. Em contraste, a matriz de covariâncias do NCEP, apresenta incrementos muito mais concentrados e um pouco mais intensos também.

Figura 3.10 - Idem Figura 3.9, para a matriz de covariâncias do NCEP.



Incremento de análise da temperatura (T) ano nível de 1000 hPa, utilizando a matriz de covariâncias do NCEP: a esquerda, incremento isotrópico; a direita incremento anisotrópico.

Fonte: Produção do autor.

4 ASSIMILAÇÃO DE DADOS HÍBRIDA ENSEMBLE-VARIACIONAL GLOBAL

Neste capítulo é apresentada a metodologia para a incorporação das covariâncias do filtro de Kalman por conjunto (Seções 2.2.3 e 2.2.4) na estrutura variacional do sistema GSI (Seção 2.2.2). O código computacional dos sistemas EnKF e EnSRF já havia sido previamente programado e estava sendo testado pelo GMAO em suas aplicações de pesquisa no sistema híbrido proposto por eles. Para o caso do sistema híbrido 3DVar do CPTEC em uso com o modelo de circulação geral do centro, não foram necessárias modificações no código computacional, com exceção do tratamento do ozônio sendo este ajustado para ser univariado (i.e., não é considerada a sua covariância com as demais variáveis da matriz de covariâncias). Isso foi necessário devido aos problemas iniciais que foram encontrados com a representação do ozônio na matriz de covariâncias calculada para o MCGAv4.

Na seção a seguir é apresentada a metodologia de incorporação das covariâncias do filtro de Kalman por conjunto na estrutura variacional. Esta metodologia segue as ideias introduzidas por Lorenc (2003) e aplicadas por Wang et al. (2008a), Wang et al. (2008b), Wang (2010) no sistema GSI. A notação das equações foi alterada para ficar em conformidade com os demais desenvolvimentos apresentados no trabalho. Além da apresentação da metodologia para a incorporação das covariâncias, são apresentadas também as configurações utilizadas no sistema para que fosse possível realizar o ciclo de análises e previsões de forma satisfatória, o que inclui a determinação de parâmetros fundamentais para a assimilação de dados como a configuração das iterações da minimização da função custo, o controle da umidade não física, porcentagem de contribuição das covariâncias do conjunto entre outros. É apresentado também um diagrama esquemático ilustrando o ciclo de assimilação de dados do sistema híbrido. Neste diagrama são apresentadas também as principais equações envolvidas nas diferentes etapas do processo. Ao final do capítulo, é apresentado como é feita a análise do campo de vento horizontal no GSI, uma vez que diferentes quantidades para representar o vento estão envolvidas no processo de assimilação de dados e previsão.

4.1 Incorporação das Covariâncias do filtro de Kalman por conjunto na Estrutura Variacional do GSI

A técnica utilizada para incorporar as covariâncias dos erros de previsão provenientes do filtro de Kalman por conjunto (EnKF/EnSRF) na estrutura variacional tridimensional (3DVar) do GSI, é aquela apresentada por Lorenc (2003) denominada “Variável

de Controle Alpha” e também em Wang et al. (2008a), Wang et al. (2008b), Wang (2010) denominada “Variável de Controle Estendida”. A técnica de incorporação ou inclusão das covariâncias dos erros de previsão na estrutura variacional consiste na definição de um novo incremento de análise ($\delta\mathbf{x}'$) e de uma matriz de correlações (\mathbf{C}), cujos elementos são multiplicados pelos elementos da matriz de covariâncias dos erros de previsão proveniente do filtro de Kalman por conjunto (\mathbf{P}^b). Esta abordagem é semelhante aquela empregada nos trabalhos de Clayton et al. (2012) e Wang e Lei (2013).

No sistema híbrido, o novo incremento de análise denotado por $\delta\mathbf{x}'$ é a soma de dois termos, definida como:

$$\delta\mathbf{x}' = \delta\mathbf{x} + \sum_{k=1}^K (\mathbf{x}_k^e \circ \mathbf{a}_k) \quad (4.1)$$

onde:

- $\delta\mathbf{x}'$: é o novo incremento de análise híbrido (dimensão: $n \times 1$);
- $\delta\mathbf{x}$: é o incremento de análise original (i.e., $\mathbf{x}^a - \mathbf{x}^b$; $n \times 1$);
- \mathbf{a}_k : é a extensão da variável de controle a ser analisada para cada membro do conjunto de K previsões ($n \times k$);
- \mathbf{x}_k^e : é a k -ésima previsão do conjunto ($n \times 1$).

Na Eq. 4.1, o primeiro termo do lado direito da equação ($\delta\mathbf{x}$), é o incremento de análise associado com a covariância estática do erro de previsão (\mathbf{B}). O segundo termo ($\sum_{k=1}^K \mathbf{x}_k^e \circ \mathbf{a}_k$), é o incremento associado as covariâncias (\mathbf{P}^b) provenientes do filtro de Kalman por conjunto. Os vetores \mathbf{a}_k (com $k = 1, \dots, K$), denotam as extensões das variáveis de controle para cada membro (os quais serão explicados com mais detalhes adiante). O símbolo \circ representa o produto Schur (i.e., elemento a elemento) entre os elementos dos vetores \mathbf{a}_k e \mathbf{x}_k^e . Em outras palavras, o segundo termo na Eq. 4.1, representa a combinação linear local das perturbações do conjunto com o incremento de análise original ($\delta\mathbf{x}$). O coeficiente \mathbf{a}_k para cada membro varia espacialmente, o qual determina a escala espacial de localização horizontal do conjunto. Ainda na Eq. 4.1, o vetor \mathbf{x}_k^e representa a k -ésima perturbação do conjunto normalizada por $\sqrt{K-1}$, onde K é o tamanho do conjunto de previsões. Este vetor é determinado pela Eq. 4.2:

$$\mathbf{x}_k^e = \frac{(\mathbf{x}_k^b - \bar{\mathbf{x}}^b)}{\sqrt{K-1}} \quad (4.2)$$

onde:

- \mathbf{x}_k^e : é a k -ésima perturbação do conjunto de K membros de previsão (dimensão: $n \times k$);
- \mathbf{x}_k^b : é o k -ésimo membro do conjunto de previsões ($n \times 1$);
- $\bar{\mathbf{x}}^b$: é a média do conjunto de K previsões ($n \times 1$).

No sistema híbrido 3DVar, a solução da análise segue a abordagem variacional do 3DVar, com a análise sendo calculada a partir da minimização da função-custo (um passo a passo da minimização da função custo variacional tridimensional é apresentada no Apêndice A, entre as Eqs. A.1 e A.20). Com o novo incremento de análise definido para o híbrido, a função custo representada pela Eq. 2.1 (na forma incremental na Eq. 4.3) pode ser reescrita considerando-se uma combinação linear entre o termo da função custo referente ao incremento de análise variacional do 3DVar denotada por $J_{3dvar,1}$ e o termo referente a extensão da variável de controle a denotada por J_e , que é proveniente do Filtro de Kalman por ensemble, além do termo referente as inovações denotado aqui por $J_{3dvar,2}$:

$$J_{3dvar}(\delta\mathbf{x}) = \underbrace{\frac{1}{2}(\delta\mathbf{x})^T \mathbf{B}^{-1}(\delta\mathbf{x})}_{J_{3dvar,1}} + \underbrace{\frac{1}{2}[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)]}_{J_{3dvar,2}} \quad (4.3)$$

A Eq. 4.4 abaixo, representa a base sobre a qual o sistema híbrido ensemble-variacional é fundamentado:

$$J(\delta\mathbf{x}, \mathbf{a}) = \alpha_1 J_{3dvar,1} + \alpha_2 J_e + J_{3dvar,2} \quad (4.4)$$

onde:

- $J_{3dvar,1}$: é o termo da função custo variacional do 3DVar referente ao incremento de análise original;
- $J_{3dvar,2}$: é o termo da função custo variacional do 3DVar referente a inovação original;

- J_e : é o novo termo adicionada a função custo variacional, referente a aplicação da extensão da variável de controle (indicado na Eq. 4.5 a seguir);
- $\delta\mathbf{x}$: é o incremento de análise devido a contribuição da matriz de covariâncias estática;
- α_1 e α_2 : são os pesos empíricos atribuídos as contribuições das matrizes de covariâncias estática e fluxo-dependente, respectivamente ($\alpha_n \in [0, 1]$, $n = 1, 2$).

Wang et al. (2008a), Wang et al. (2008b) utilizam da igualdade $\mathbf{y}'^o = \mathbf{y}^o - H(\mathbf{x}^b)$ para escrever a forma incremental da função custo híbrida em termos da extensão da variável de controle (Eq. 4.5), e em sua forma final, Eq. 4.6:

$$J(\delta\mathbf{x}', \mathbf{a}) = \alpha_1 \frac{1}{2} (\delta\mathbf{x}')^T \mathbf{B}^{-1} (\delta\mathbf{x}') + \alpha_2 \frac{1}{2} (\mathbf{a})^T \mathbf{A}^{-1} (\mathbf{a}) + \frac{1}{2} [\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})] \quad (4.5)$$

$$J(\delta\mathbf{x}') = \frac{1}{2} (\delta\mathbf{x}')^T (\alpha_1 \mathbf{B} + \alpha_2 \mathbf{P}^b \circ \mathbf{A})^{-1} (\delta\mathbf{x}') + \frac{1}{2} [\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x}')]^T \mathbf{R}^{-1} [\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x}')] \quad (4.6)$$

onde:

- \mathbf{P}^b : é a matriz de covariâncias dos erros do conjunto de previsões, tal como definida nas Eqs. 2.5 e 2.9;
- \mathbf{A} : é uma matriz de correlação que faz a localização horizontal das covariâncias dos erros do conjunto. Ela é formada pelos vetores \mathbf{a}_k e, tal como a matriz \mathbf{B} , sua aplicação se dá através de um filtro recursivo, como o que está apresentado no Apêndice D.

Para conservar a variância do erro da previsão na Eq. 4.6, os parâmetros α_1, α_2 são restringidos pela relação $\alpha_1 + \alpha_2 = 1$. Logo uma contribuição de, eg., 75% das covariâncias do conjunto de previsões (i.e., através de \mathbf{P}^b) implica em $\alpha_2 = 0,75$ e consequentemente $\alpha_1 = 0,25$.

4.2 Ajustes dos Parâmetros Principais

Nesta seção são apresentados os aspectos principais relacionados com as diversas componentes que compõem o sistema híbrido 3DVar. Algumas destas componentes devem ser configuradas e outras estabelecidas para a realização cíclica do sistema híbrido. Em alguns casos, as opções padrão foram adotadas e em outras, soluções a parte tiveram que ser criadas.

4.2.1 Matriz de Covariâncias Estática (B)

Nesta subseção, são apresentadas as principais componentes e opções relacionadas com a determinação e aplicação da matriz de covariâncias estática dos erros de previsão.

Aplicação das Covariâncias Estáticas

Diferentemente da matriz idealizada apresentada na Figura 3.1, a matriz de covariâncias do GSI contém as seguintes estruturas: amplitudes (representadas pelos desvios-padrão e variâncias), distâncias (representadas pelos comprimentos de escalas horizontais e verticais) e matrizes de projeção de balanço (representadas por coeficientes de regressão). Nesta versão da matriz de covariâncias utilizada pelo GSI (independente da versão do modelo de circulação geral do CPTEC) as amplitudes são representadas por médias zonais (as quais podem ser ponderadas durante o processo de minimização da função custo), variando apenas nas latitudes e na vertical. As partes desbalanceadas das variáveis velocidade potencial (χ), temperatura virtual (T) e pressão em superfície (ps), possuem matrizes de projeção (\mathbf{G} e \mathbf{W} , nas Eqs. 4.7 e 4.8 respectivamente) que são responsáveis por projetar o incremento de análise da função de corrente (ψ) no perfil vertical da parte balanceada do incremento de cada uma destas variáveis. Para a velocidade potencial, uma matriz de correlação (\mathbf{C} na Eq. 4.9) é utilizada para contabilizar a correlação positiva entre divergência (D) e vorticidade (ζ).

$$T_b = \mathbf{G}\psi \quad (4.7)$$

$$P_b = \mathbf{W}\psi \quad (4.8)$$

$$\chi_b = \mathbf{C}\psi \quad (4.9)$$

onde, nas Eqs. 4.7, 4.8 e 4.9 o subscrito b indica a parte balanceada das variáveis.

Esta é a forma utilizada pelo GSI para aplicar os incrementos de análise de forma filtrada, i.e., decompõem-se as variáveis de controle em parte balanceada (referente a parte lenta do fluxo atmosférico) e em parte não balanceada (referente a parte rápida e ruidosa do fluxo atmosférico) e então os incrementos de análise são realizados apenas na parte balanceada.

Filtro Recursivo

O filtro recursivo utilizado para a aplicação das covariâncias no GSI tem a importante propriedade de ajustar as amplitudes calculadas de forma que seu aspecto possua algum grau de anisotropia (i.e., variação do aspecto representado em todas as direções) permitindo que as covariâncias se adaptem ao fluxo representado pela previsão a ser corrigida pelas observações. Os parâmetros de escala horizontais utilizados pelo filtro recursivo neste processo, são tabelados e uma discussão sobre a sua determinação pode ser encontrada em [Wu et al. \(2002\)](#).

No GSI, a matriz de covariância não é explicitamente construída e, ao invés disso, o filtro recursivo ([PURSER et al., 2003a](#); [PURSER et al., 2003b](#)) é utilizado para realizar a aplicação das covariâncias. Segundo [Wu et al. \(2002\)](#), a matriz \mathbf{B} é então decomposta da seguinte forma:

$$\mathbf{B} = \mathbf{B}_z(V^1\mathbf{B}_x^1\mathbf{B}_y^1\mathbf{B}_y^1\mathbf{B}_x^1V^1 + V^2\mathbf{B}_x^2\mathbf{B}_y^2\mathbf{B}_y^2\mathbf{B}_x^2V^2)\mathbf{B}_z \quad (4.10)$$

onde:

- V^1 e V^2 são os desvios-padrão de cada variável de controle (obtidos a partir da própria matriz de covariâncias calculada);
- \mathbf{B}_x , \mathbf{B}_y e \mathbf{B}_z representam a aplicação do filtro recursivo nas direções x (direção Oeste-Leste), y (direção Sul-Norte) e z (na vertical);
- \mathbf{B}^1 e \mathbf{B}^2 representam a aplicação do filtro recursivo nas escalas horizontais indicadas (direções x e y).

Exemplos de filtros recursivos unidimensional de grau 1 e 2, são apresentados nos Apêndices C e D, respectivamente.

Correção de Viés

Um outro aspecto tratado durante o cálculo da matriz de covariância é o viés do modelo numérico. O viés (β^t) é uma medida do erro sistemático do modelo de PNT e pode ser representada pelo modelo (ω^f) subtraído do erro real (ω^t) e de um erro aleatório, cuja esperança é zero, ie., $\langle \epsilon^t \rangle = 0$ (CHEPURIN et al., 2005). A remoção de viés (Eq. 4.11) é realizada com a finalidade de se eliminar os erros mais grosseiros, e é aplicada em toda a coluna vertical do modelo. Este procedimento é realizado para as seguintes variáveis: parte balanceada da função de corrente, parte desbalanceada da velocidade potencial, temperatura virtual, umidade relativa normalizada, razão de mistura de ozônio, razão de mistura de condensação de água em nuvens e pressão em superfície.

$$\beta^t = \omega^f - \omega^t - \epsilon^t \quad (4.11)$$

Comprimentos de Escala

Os comprimentos de escala horizontais são importantes estruturas da matriz \mathbf{B} e são dependentes dos valores de variâncias calculados - Eq. 4.12. Os comprimentos de escala verticais podem ser estimados através da correlação vertical de cada variável de controle (WU et al., 2002).

$$L = \left(\frac{8V_\psi}{V_\zeta} \right)^{\frac{1}{4}} \quad (4.12)$$

onde:

- V_ψ : representa a variância da função de corrente;
- V_ζ : representa a variância da vorticidade.

Os comprimentos de escala descrevem o quão distantes as covariâncias entre os erros das variáveis de controle podem ser relacionadas espacialmente. É através deles que os elementos da diagonal principal da matriz \mathbf{B} são transportados até uma determinada distância para fora da diagonal principal. Portanto, a variação do comprimento de escala descreve o modo como a covariância varia com a distância. Além disso, é através dos comprimentos de escala que o filtro recursivo é aplicado para a modelagem das matrizes de correlação do GSI.

4.2.2 Assimilação de Dados Variacional

Nesta seção são apresentados os aspectos principais e alguns ajustes da componente de assimilação variacional que foram feitos e utilizados entre todos os experimentos realizados com o sistema GSI. O sistema GSI possui vários parâmetros e opções que podem ser ajustados permitindo que o sistema seja realizado com determinadas funções e procedimentos. Três destas opções são apresentadas: configurações relacionadas com a minimização da função custo variacional, a escolha da variável de controle da umidade e a injeção de umidade.

Minimização da Função Custo

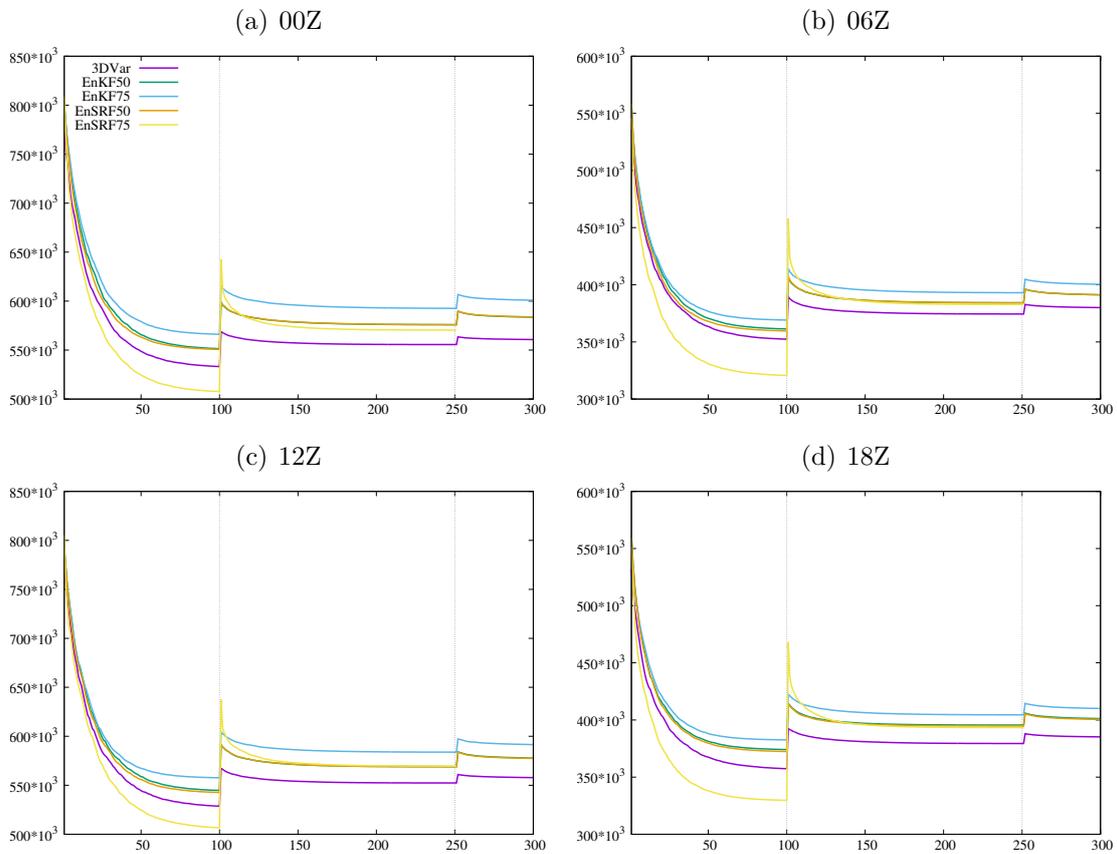
Dois aspectos importantes estão relacionados com a minimização da função custo variacional tridimensional, utilizada no sistema GSI: o ajuste do número de iterações externas e internas e o método de minimização da função custo. As iterações internas e externas estão relacionadas a forma como o operador observação será aplicado e que tipos de relações ele irá tratar (NICHOLS, 2010) além do controle de qualidade das observações. Nas iterações externas, são tratadas as relações não lineares entre as observações não convencionais e o campo de previsão através da aplicação do operador observação não linear (H). Nas iterações internas, são tratadas relações lineares com a aplicação do operador observação linear (\mathbf{H}). O método de minimização da função custo, é um método iterativo (cujo número de iterações respeita um critério pré-estabelecido de parada ou o número de iterações definido) é aplicado dentro das iterações internas e é utilizado para se atingir um estado que, ao final da última iteração externa, é o estado de análise.

Em todos os experimentos, o sistema GSI foi configurado para ser realizado com 3 iterações externas (ou *outer loops*). A primeira iteração externa foi ajustada com um número máximo de 100 iterações internas (ou *inner loops*); a segunda iteração externa, foi ajustada com 150 iterações internas e a terceira iteração externa, com 50 iterações internas. Como nos experimentos realizados não se fez nenhum compromisso com o desempenho computacional, as opções para a configuração do número de iterações do procedimento de minimização da função custo variacional foram escolhidas com valores relativamente grandes para se garantir a convergência da função custo variacional (o que inclui a minimização das análises realizadas no experimento com o 3DVar puro e com os experimentos com as análises híbridas 3DVar).

A Figura 4.1 mostra um exemplo das curvas de minimização da função custo variacional

dos experimentos realizados, para cada um dos horários sinóticos padrão, para o dia 20130101. Além do decaimento suave e característico, destaca-se também as diferenças em relação aos valores: para os horários das 00 e 12Z (Figuras 4.2(a) e 4.2(c)) há mais observações (principalmente observações não convencionais de radiancias) do que nos horários das 06 e 18Z (Figuras 4.2(b) e 4.2(d)). Por esta razão é que o valor inicial e referente a retomada das iterações externas são mais altos nos horários das 00 e 12Z.

Figura 4.1 - Exemplos das curvas de minimização da função custo variacional para os experimentos realizados (20130101).



Em (a) para as 00Z; (b) 06Z; (c) 12Z e (d) 18Z. No eixo x estão representadas as iterações da minimização e estão destacados (linha pontilhada cinza) a parada das iterações internas e reinício das iterações externas.

Fonte: Produção do autor.

O método de minimização da função custo utilizado é o método do Gradiente

Conjugado. Este método é o método padrão, mas há outras opções que podem ser testadas. O pré-condicionamento da função custo em aplicações operacionais é necessário para minimizar o custo computacional do cálculo da análise. Um opção é realizar o pré-condicionamento da minimização da função custo com base na matriz de covariâncias estática completa, ou com base na sua raiz quadrada. Em todos os experimentos realizados, foi utilizado, portanto, o método de minimização do gradiente conjugado com pré-condicionamento em \mathbf{B} completa. No Apêndice B é apresentada uma visão geral sobre o pré-condicionamento com base em \mathbf{B} completa.

Variável de Controle da Umidade

A variável de controle de umidade no GSI pode ser escolhida de duas formas distintas: utilizando-se a pseudo-umidade relativa ou a umidade relativa normalizada. A escolha da variável de controle da umidade foi a umidade relativa normalizada porque esta representação permite que a umidade varie durante as iterações internas de acordo com as variações na pressão em superfície, temperatura e umidade específica.

Injunção de Umidade

A implementação prática da função custo variacional, pode incluir outros termos além daqueles mostrados, e.g., na Eq. 4.4. Um termo importante é o termo de injunção de umidade. Este termo é responsável por contabilizar e controlar a umidade negativa e supersaturada, que surgem devido aos modos computacionais e físicos envolvidos no processo de modelagem (i.e., prognóstico) e de análise. No GSI, dois parâmetros podem ser ajustados a fim de que a umidade não física possa ser controlada. O primeiro teste com a matriz de covariâncias na resolução TQ0062L028, mostrou-se instável (após poucos ciclos de análise). Vários testes foram realizados e, com base no trabalho de Rodrigues (2017), foi possível ajustar os seguintes valores: 5.0 para a umidade supersaturada e 0,005 para a umidade negativa. Nos testes realizados, notou-se que a magnitude destes valores podem ser relacionadas também com a resolução da grade da análise e modelo (neste caso, ~200 km).

4.2.3 Assimilação de Dados por Conjuntos

Esta seção é dedicada a apresentar os principais parâmetros e aspectos relacionados com a aplicação do filtro de Kalman por conjunto, nas duas versões testadas (EnKF e EnSRF). Para ambos os algoritmos testados, as opções de localização do conjunto de previsões bem como a inflação (i.e., o *inflation*) das covariâncias são as mesmas. Esta medida foi tomada para facilitar a comparação dos resultados obtidos. Entretanto,

isso não significa que os sistemas devem ser realizados com as mesmas opções ou que não há necessidade de se ajustar adequadamente estas.

Determinação do Conjunto de Previsões Inicial

Para a realização cíclica do sistema híbrido 3DVar utilizando os algoritmos EnKF ou EnSRF, um conjunto de previsões inicial é necessário, a partir do qual o filtro de Kalman por conjunto irá realizar a atualização das covariâncias a serem utilizadas em combinação com as covariâncias estáticas na minimização das função custo variacional. A metodologia empregada aqui é aquela semelhante ao *poor man's ensemble* (e.g., Ebert (2001), Arribas et al. (2005)), em que previsões independentes previamente realizadas por centros operacionais são utilizadas para se compor um conjunto de previsões para aplicações em previsões de curto a médio prazo. Entretanto, uma ideia semelhante foi aplicada aqui para se gerar um conjunto de previsões inicial para uso com o sistema híbrido 3DVar, o qual foi atualizado e propagado no tempo utilizando o filtro de Kalman por conjunto.

As etapas necessárias para se gerar um conjunto de previsões iniciais são as seguintes:

- a) Parte-se de uma análise determinística de resolução maior ou igual a resolução de interesse (i.e., pode-se partir de uma análise na resolução TQ0299L064 e realizar as previsões para a resolução TQ0062L028), realizam-se previsões a cada 12 horas para um período de 30 dias;
- b) Do conjunto de 60 previsões (2 previsões por dia), seleciona-se um subconjunto de previsões que representará o tamanho do conjunto;
- c) Renomeia-se os arquivos de forma que todos sejam válidos para o mesmo horário de previsão;
- d) Altera-se a data interna dos arquivos de forma que todos sejam efetivamente válidos para o horário de previsão desejado.

Nas etapas acima, é necessário alterar-se a data interna dos arquivos de previsão, o que requer um programa que seja capaz de ler o arquivo de entrada, alterar a informação da data e horário da previsão no cabeçalho do arquivo e salvar estas informações junto com os coeficientes espectrais representados.

Para os experimentos realizados neste trabalho (Seção 4.5), foi utilizado um conjunto de 40 previsões iniciais, os quais se mostraram suficientes para a realização dos experimentos.

Localização do Conjunto de Análises

A aplicação do filtro de Kalman por conjunto utilizando-se um conjunto de previsões relativamente pequeno para realizar a propagação dos estados e das suas covariâncias, pode fazer com que o filtro seja ineficiente e colapse. Isso pode ocorrer porque, à medida em que as observações são assimiladas no conjunto, o seu espalhamento tende a diminuir devido as inovações (e correções trazidas pelas observações). Isso implica em uma estimativa pobre das covariâncias dos erros de análise devido a um conjunto pequeno. Um conjunto pequeno contém alguns poucos membros (na ordem de 10 membros). Os 40 membros utilizados nos experimentos deste trabalho representam um número relativamente pequeno, mas suficiente para a realização dos experimentos propostos na resolução escolhida, sem se comprometer severamente a disponibilidade de tempo computacional e espaço para armazenamento.

A localização do conjunto de previsões é então utilizada para se compensar os efeitos da atualização cíclica de um conjunto que possui um espalhamento relativamente pequeno, devido ao seu tamanho. A implementação do EnKF no GSI, permite que sejam ajustados os seguintes parâmetros para se realizar a localização do conjunto de previsões:

- Escala de localização horizontal do conjunto: 800 km;
- Escala de localização vertical do conjunto: 0,5 (unidades de grade).

Outra opção disponível para a localização do conjunto de previsões, é dada em função da razão entre a variância da análise no espaço das observações e a variância das previsões. Esta opção representa um valor limite para que as observações reduzam a variância do conjunto de previsões (e conseqüentemente manter parte do espalhamento do conjunto). O valor utilizado nesta opção foi de 98%.

Inflação das Covariâncias

Com a aplicação de um conjunto relativamente pequeno para a aplicação do filtro de Kalman por conjunto, as covariâncias derivadas do conjunto serão deficientes em posto (i.e., *rank deficient*). Isso significa que, sendo a matriz de covariâncias do conjunto dimensionada no espaço do conjunto (i.e., uma de suas dimensões é a dimensão do conjunto, o seu tamanho), então não haverão elementos suficientes para uma adequada amostragem das variâncias e covariâncias dos erros das variáveis de estado. A inflação pode então ser utilizada (em diferentes regiões) com o objetivo de

se inflar artificialmente estas covariâncias.

Na implementação do EnKF dentro do GSI, estão disponíveis as seguintes opções para a realização da inflação das covariâncias, as quais foram mantidas nos experimentos:

- Parâmetro de inflação das covariâncias sobre o Hemisfério Norte: 85%;
- Parâmetro de inflação das covariâncias sobre os Trópicos: 85%;
- Parâmetro de inflação das covariâncias sobre o Hemisfério Sul: 85%;
- Limite máximo de inflação (total): 100%;
- Limite mínimo de inflação (total): 1%.

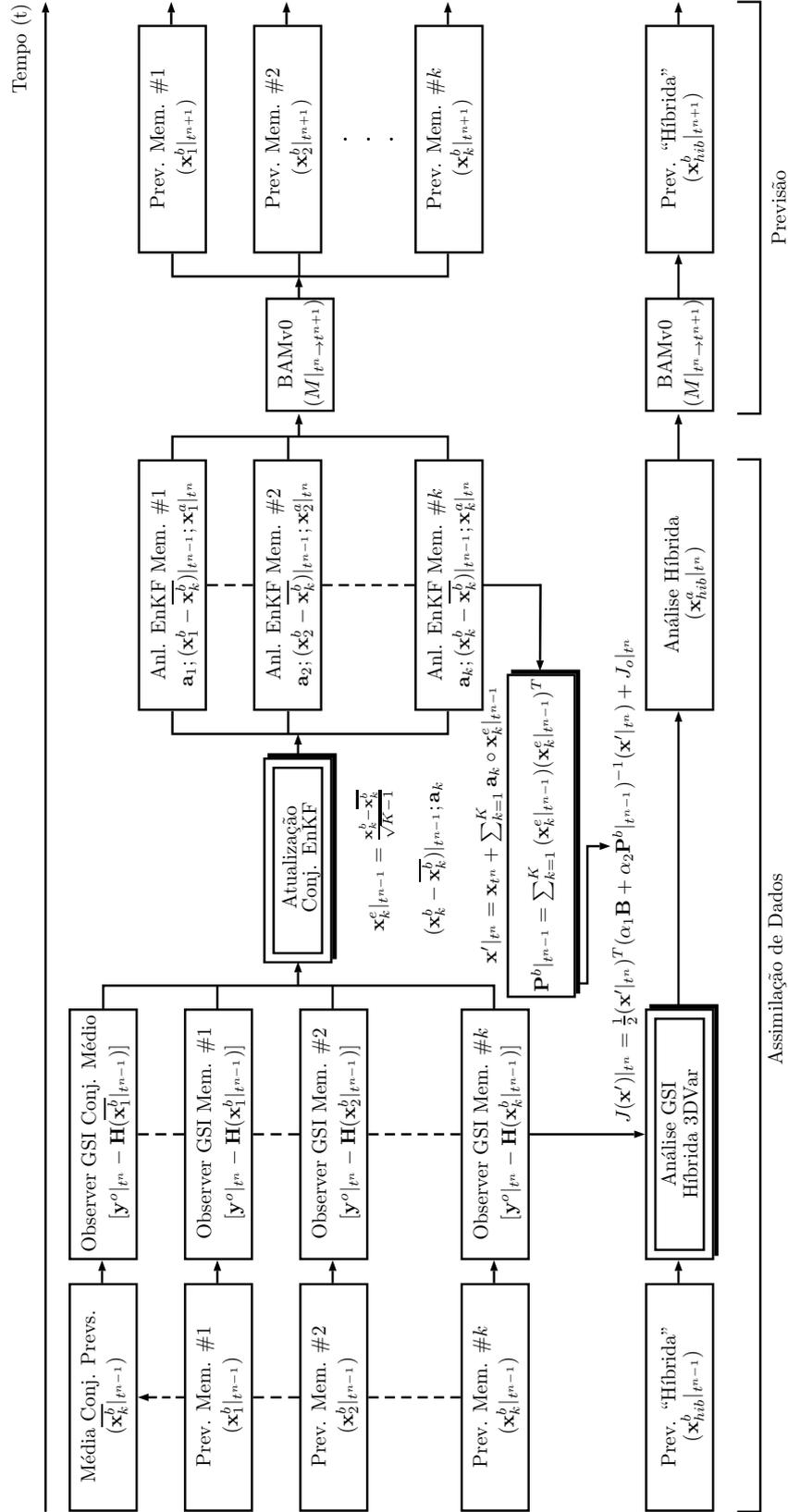
4.3 Ciclo de Assimilação de Dados do Sistema Híbrido Ensemble-Variacional

O ciclo do sistema de assimilação de dados GSI é realizado alternando-se análises e previsões. Em modo de pesquisa, antes de se iniciar o ciclo, são gerados o primeiro conjunto de previsões e *restarts* (que incluem a convecção, radiação e nuvens, dinâmica, - tempo corrente e passado - e superfície) do modelo de circulação geral do CPTEC, a partir dos quais, o ciclo é iniciado. Para isto, antes são preparados os campos fixos e a primeira condição inicial do modelo utilizando-se o pré-processamento. No ciclo, a primeira etapa, consiste do processamento das observações em um procedimento em que os arquivos de observações convencionais e não convencionais no formato PrepBUFR são preparados para leitura pelo GSI. Em seguida é executada a interface que converte a previsão de 9 horas do modelo do formato espectral para ponto de grade, preparando também para a leitura pelo GSI. Com as observações e previsões preparados, o GSI é executado e então a análise é determinada. Com a análise pronta, o modelo de circulação geral é então executado, lendo o *restarts* gerado antes do ciclo, atualizando-se a Temperatura da Superfície do Mar (TSM) e a máscara de neve, prognosticando a dinâmica e diagnosticando a física do modelo conforme as configurações da Tabela 2.1. Depois que o modelo escreve o resultado das previsões e escreve os *restarts* para o próximo ciclo, o ciclo é reiniciado para o horário seguinte, realizando as mesmas etapas.

No diagrama da Figura 4.2 é mostrado o ciclo de análises e previsões do sistema híbrido ensemble-variacional proposto. As equações em cada passo do ciclo são aquelas utilizadas pelo sistema de assimilação de dados híbrido para a determinação

da análise. Neste diagrama, não estão representados o pré-processamento, interfaces, conversão das observações, mas apenas as partes principais. Para o ciclo de assimilação de dados híbrido, destacam-se as Eqs. 2.5, 4.1, 4.2 e a função custo da Eq. 4.6.

Figura 4.2 - Diagrama esquemático do sistema híbrido 3DVar.



As janelas de assimilação e previsão estão indicadas junto com o tempo de ocorrência das previsões, observações e análise (i.e., t^{n-1} , t^n , t^{n+1}). As equações referentes as etapas principais estão indicadas em suas respectivas posições.

Fonte: Produção do autor.

A Figura 4.2 apresenta o ciclo de geração da análise híbrida 3DVar e a geração de um conjunto de K membros. A análise híbrida é atualizada com a estimativa de covariância do erro de previsão estimada pela combinação entre a covariância estática do erro de previsão e a covariância do conjunto de K membros. A perturbação do conjunto é então atualizada com inovações baseadas na inflação do conjunto aplicado.

No sistema GSI, utilizando-se ou não a matriz de covariâncias estacionária, é dado o exemplo de como é calculada a análise do campo horizontal do vento. Este é um exemplo interessante, porque mostra o tratamento feito dentro do sistema em relação a variável de controle (divergência e vorticidade), a variável observada (componentes zonais do vento horizontal) e as quantidades tratadas dentro da matriz de covariâncias, importantes para a análise do vento (função de corrente e velocidade potencial).

4.4 Exemplo da Análise do Campo de Vento Horizontal

O modelo de circulação geral do CPTEC (seja o MCGAv4 ou o BAMv0, descritos na Seção 2.1) prognostica a divergência (D) e a vorticidade (ζ) para representar as componentes divergente e não divergente, respectivamente, do fluxo atmosférico. Na matriz de covariância, estão representadas as covariâncias relacionadas ao vento em termos da função de corrente (ψ) e velocidade potencial (χ). Em relação as observações de vento, diversos instrumentos são utilizados para a coleta de informações, mas são utilizadas pelo sistema de assimilação de dados (e também na matriz de covariância dos erros de observação) as componentes do vento horizontal, sendo representadas por meio das componentes zonal (u) e meridional (v) do vento.

Tomando-se por base a Eq.2.1 (desconsiderando os termos de injeção), temos:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (4.13)$$

O problema de análise, seguindo a função custo da Eq. 2.1, pode ser dividido em duas partes: a primeira, relacionada ao termo de inovação da função custo (i.e., o termo $\mathbf{y}^o - H(\mathbf{x}^b)$), e a segunda parte, relacionada ao termo referente ao incremento de análise (i.e., o termo $\mathbf{x} - \mathbf{x}^b$).

Considerando que o incremento de análise é obtido quando o gradiente de J é igual a zero, ie., $\nabla_{\mathbf{x}} J = 0$ quando $\mathbf{x} = \mathbf{x}^a$, temos que:

$$0 = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (4.14)$$

A Eq. 4.14 acarreta na Equação do Incremento de Análise (vide Apêndice A):

$$\mathbf{x}^a - \mathbf{x}^b = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \{ \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)] \} \quad (4.15)$$

Dentro do GSI, esta equação não é explicitamente resolvida, mas os termos são todos calculados. As etapas relacionadas ao cálculo do incremento de análise, são as seguintes:

- a) Utilizando as Eqs. 3.10 e 3.11, \mathbf{H} transforma D e ζ do modelo para u e v , e interpola para o ponto de observação;
- b) Com u e v do modelo interpolados para o ponto da observação, é calculada a inovação $\mathbf{y}^o - H(\mathbf{x}^b)$;
- c) A inovação é então dividida pelo erro da observação, $\mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)]$;
- d) Esta inovação normalizada pelo erro da observação, é então transformada novamente para D e ζ (ie., as inovações das componentes u e v são transformadas novamente para D e ζ) por \mathbf{H}^T , $\mathbf{H}^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)]$. Este termo está representado no segundo coeficiente do lado direito da Eq. 4.14.

Esta é a primeira parte do cálculo da análise do vento horizontal, referente ao termo de inovação. A segunda parte, é referente ao termo de incremento de análise e, além de envolver a matriz de covariâncias dos erros de observação (\mathbf{R} , que neste caso representa os erros das observações de u e v), envolve também a matriz de covariâncias dos erros de previsão (\mathbf{B} , representando portanto, os erros associados as quantidades de ψ e χ). O coeficiente $(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$, lado direito da Eq. 4.14), indica a matriz inversa da soma das matrizes \mathbf{B} e \mathbf{R} , no espaço do modelo (ponto de grade). A matriz \mathbf{B} utiliza a Função de Corrente (ψ) e a Velocidade Potencial (χ). As Eqs. 3.12 e 3.13, são então utilizadas para calcular ψ e χ a partir de D e ζ , respectivamente, e então \mathbf{B} é aplicada para o cálculo final do incremento de análise de T , ps e χ em termos de ψ . Finalmente, os incrementos de ψ e χ são transformados de volta em incrementos de D e ζ .

A equação a seguir organiza as informações apresentadas de acordo com a natureza das quantidades avaliada para a análise do vento horizontal.

$$\mathbf{x}^a - \mathbf{x}^b = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \{ \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)] \}$$

D, ζ
 ψ, χ
 u, v
 u, v
 u, v
 D, ζ

4.5 Experimentos com o Sistema Híbrido 3DVar

Uma das principais características do sistema híbrido 3DVar, é a capacidade de combinar covariâncias estáticas e covariâncias provenientes de um conjunto de previsões. Tal como foi apresentado nas seções anteriores, os coeficientes α_1 e α_2 podem ser utilizados para se ponderar a quantidade de contribuição destas covariâncias que serão utilizadas na minimização da função custo variacional. Com o objetivo de se exercitar e avaliar o sistema híbrido 3DVar em ciclos de assimilação de dados utilizando os dados de observações sumarizados na Tabela 2.2, foram elaborados dois grupos de experimento. O primeiro grupo é composto por dois experimentos que serão utilizados como controle; o segundo grupo é composto por 4 experimentos que representam o sistema híbrido 3DVar.

Na Tabela 4.1 são apresentados os experimentos dos grupos 1 e 2, e apresenta as características principais de cada um deles.

Tabela 4.1 - Experimentos com o sistema híbrido 3DVar.

Experimento	Ajuste das Covariâncias	Descrição
REF	–	–
3DVar	–	B (BAMv0, 730 pares)
EnKF50	50% Estático ($\alpha_1 = 0, 5$)	B (BAMv0, 40 membros)
EnKF75	75% Conjunto ($\alpha_1 = 0, 25$)	B (BAMv0, 40 membros)
EnSRF50	50% Estático ($\alpha_1 = 0, 5$)	B (BAMv0, 40 membros)
EnSRF75	75% Conjunto ($\alpha_1 = 0, 25$)	B (BAMv0, 40 membros)

O objetivo do experimento REF é servir como referência para todos os experimentos que envolvem ciclos de assimilação de dados, de forma que se possa conhecer as diferenças entre realizar o modelo de circulação geral do CPTEC com e sem os ciclos de assimilação de dados. Para que seja, portanto, possível se comparar o desempenho dos experimentos híbridos, realizou-se também um experimento cíclico com o sistema 3DVar puro, de forma que este sirva como base para as comparações com as contribuições das covariâncias dos conjuntos de previsões. Os experimentos

EnKF50 (EnKF75) e EnSRF50 (EnSRF75), são os experimentos que utilizarão o conjunto de 40 membros de previsões com o EnKF e EnSRF, com 50% e 75% de contribuição das covariâncias do conjunto, as covariâncias estáticas.

5 RESULTADOS

Neste capítulo são apresentados os resultados obtidos a partir dos experimentos realizados com o sistema híbrido 3DVar. Nesses experimentos, buscou-se testar os dois algoritmos de assimilação de dados EnKF e EnSRF, além dos efeitos da contribuição do conjunto de análises na determinação da matriz de covariâncias híbrida. A Seção 5.1 traz os resultados relacionados a aplicação do sistema híbrido 3DVar. A verificação dos resultados é feita com base na avaliação da inovação dos conjuntos de análises e a sua contribuição na determinação das análises híbridas determinísticas. Do ponto de vista das previsões numéricas, é avaliada a habilidade das previsões de até 5 dias a partir das análises variacionais, obtidas com a aplicação da matriz de covariâncias híbrida.

A Seção 5.2, contextualiza e valida a aplicação do sistema híbrido 3DVar através de um estudo de caso da Zona de Convergência do Atlântico Sul (ZCAS), em Janeiro de 2013.

5.1 Assimilação de Dados Híbrida 3DVar

Nesta seção são apresentados os resultados obtidos com os experimentos do sistema híbrido 3DVar utilizando a matriz estática de covariâncias dos erros de previsão do modelo BAMv0. Nos experimentos avaliados, foram verificadas as contribuições das covariâncias dos erros do conjunto de previsões, derivadas a partir dos algoritmos EnKF e EnSRF. Os resultados desta seção, estão organizados da seguinte forma: na Seção 5.1.1, são mostradas as estatísticas de inovação dos conjuntos de análise dos experimentos híbridos. As estatísticas de inovação apresentadas foram calculadas em relação ao erro das observações convencionais assimiladas e podem ser interpretadas como importantes indicadores do condicionamento do conjunto de análises, indicando deficiências no espalhamento do conjunto e na inflação das covariâncias. Na Seção 5.1.3, são apresentados os resultados obtidos com a integração por até 120 horas das análises oriundas dos experimentos. A avaliação destes resultados é feita com base na correlação de anomalias das previsões contra suas próprias análises. Por fim, na Seção 5.1.4, são apresentados os resultados da avaliação das previsões de precipitação de 24 horas do modelo BAMv0, calculadas a partir das análises dos experimentos.

5.1.1 Viés Ponderado dos Conjuntos de Análises

Para uma avaliação do conjunto de análises produzido a partir da aplicação do sistema híbrido 3DVar utilizando filtro de Kalman por conjunto (EnKF/EnSRF),

uma avaliação das inovações do conjunto é feita com o objetivo de se diagnosticar as deficiências no espalhamento do conjunto devido aos ajustes na inflação das covariâncias e a localização do conjunto.

Para então se verificar a forma como o espalhamento do conjunto é representado na presença dos erros das observações, as inovações do conjunto são mostradas como uma medida do quão bem representado é o espalhamento do conjunto devido as inovações trazidas pelas observações. A comparação entre as estatísticas de inovação do EnKF e EnSRF são apresentadas para três regiões diferentes: Hemisfério Norte, Trópicos e Hemisfério Sul, e somente para as observações convencionais. Para tornar o entendimento dos gráficos mais fácil, no eixo x estão representadas as razões entre o desvio padrão das inovações dos *priors* (i.e., o conjunto de previsões) e a raiz quadrada do espalhamento total do conjunto (Eq. 5.1).

$$VI = \frac{\delta(\mathbf{y}^o - \mathbf{H}\mathbf{x}_k^b)}{\sigma(\mathbf{y}^o - \mathbf{H}\mathbf{x}_k^b)} \quad (5.1)$$

onde:

- δ : é o viés;
- σ : é o desvio-padrão;
- \mathbf{y}^o : é vetor observação;
- \mathbf{x}_k^b : é o k -ésimo membro do conjunto de K previsões;

Como o desvio padrão das inovações dos *priors* é normalizada pela raiz quadrada do espalhamento total, quanto menores os valores determinados, melhor é a inovação do conjunto representado. Idealmente, o desvio padrão dos *priors* deve se equiparar com os valores do espalhamento total se o conjunto de análises estiver bem condicionado, i.e., com uma boa quantidade de espalhamento e valores de localização e inflação apropriadamente ajustados. No caso dos experimentos realizados, como já foi mencionado anteriormente, estes valores foram mantidos iguais entre todos os experimentos que envolveram o filtro de Kalman por conjunto.

As figuras estão apresentadas para o horário das 00Z para reduzir o tamanho das séries temporais, uma vez que não foram verificadas grandes diferenças com os demais horários. Nos painéis de cima, as linhas vermelhas (pontilhadas e contínuas)

referem-se os experimentos híbridos 3DVar utilizando o EnKF, em que foi considerado 50% de contribuição das covariâncias estáticas (\mathbf{B}), enquanto que as linhas azuis referem-se aos mesmos experimentos, mas denotando os casos em que 75% de contribuição das covariâncias do conjunto de previsões foi considerado. Nos painéis de baixo, os experimentos utilizando o EnSRF estão representados pelas linhas amarela (representando a contribuição de 50%) e verde (representando a contribuição de 75% das covariâncias).

A Figura 5.1 apresenta as estatísticas de inovação do conjunto dos experimentos híbridos 3DVar utilizando o EnKF e EnSRF para o horário das 00Z. Em geral, não foram observadas grandes diferenças com os demais horários, seja na avaliação da magnitude e da variação das amplitudes das curvas. Entretanto, pode-se afirmar que as amplitudes dos *priors* (i.e., do conjunto de previsões) e dos *posteriors* (i.e., do conjunto de análises) variam discretamente entre si. Isto pode ser devido as variações na disponibilidade das observações nos horários das 06 e 18Z. Para algumas variáveis, o desvio padrão dos *priors* parece muito distante de concordar com o espalhamento total do conjunto. Por exemplo, a temperatura (T , Figuras 5.2(d), 5.2(e) e 5.2(f)), em todas as regiões avaliadas, saturam o desvio padrão dos *priors* e se mantém com alta amplitude até o final da simulação (de 1 a 31 de Janeiro de 2013). Na região TR, T (Figura 5.2(e)) leva em torno de 20 ciclos (i.e., 5 dias) para estabilizar as estatísticas de inovação do conjunto. Uma situação semelhante ocorre na região HS (Figura 5.2(f)), também para T e também na região TR para a variável q (Figura 5.2(h)), em que o desvio padrão dos *priors* é maior do que o espalhamento total do conjunto, além de aumentar com o tempo. Em outras palavras, isto significa que o erro das previsões tende a dominar o espalhamento, o que mostra a falta de um ajuste de inflação adequado no conjunto de previsões destas variáveis para estas regiões.

A pressão em superfície (ps) nas regiões HN (Figura 5.2(m)), mostra que o desvio padrão dos *priors* tende a crescer com o tempo, mas que isso é compensado pelo espalhamento do conjunto indicando que o conjunto de previsões tenta se ajustar as observações ainda com algum espalhamento. Uma situação diferente nas regiões TR e HS onde o desvio padrão dos *priors* para ps (Figuras 5.2(n), 5.2(o)) também é maior do que o espalhamento do conjunto, mas negativo. O sinal negativo do desvio padrão dos *priors*, neste caso, indica que o desvio padrão da inovação do conjunto (i.e., $\sigma(\mathbf{y}^o - \mathbf{H}\mathbf{x}_k^e)$) é negativo, como se as previsões tentassem corrigir as observações.

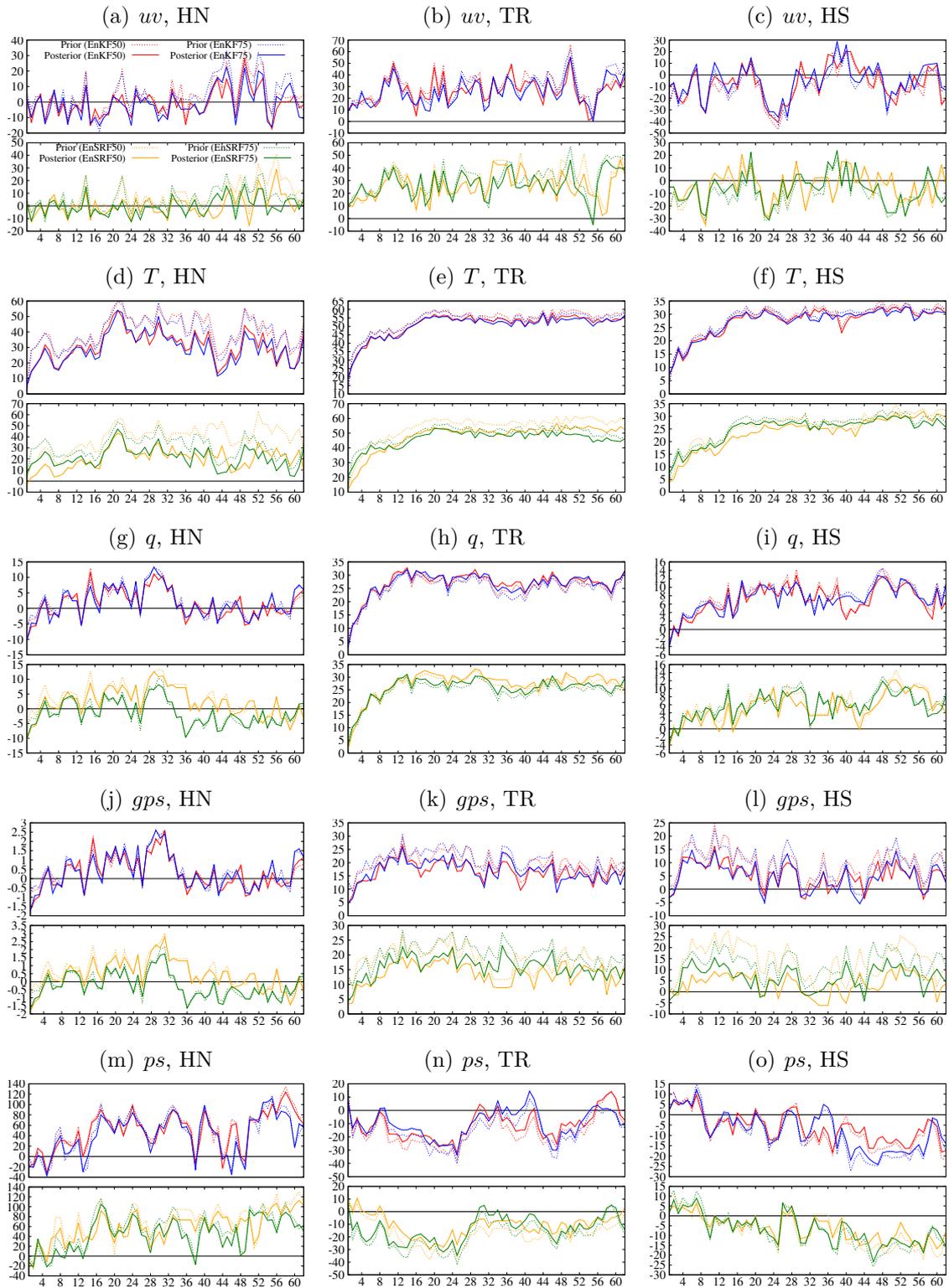
Diferenças entre a quantidade relativa de contribuição do conjunto (i.e., as diferenças

entre as contribuições de 50% e 75%, independente do tipo de filtro de Kalman utilizado), parecem ser mais ou menos sensíveis para algumas variáveis. Para ps na região TR e HS (Figuras 5.2(n), 5.2(o)), está claro que com 50% de contribuição das covariâncias fluxo-dependentes do EnKF as covariâncias estáticas, é um valor razoável se comparado com os 75% e contribuição do EnSRF, independente do horário sinótico (se 00 ou 12Z). Isto pode ser devido ao fato de a perturbação das observações serem realizadas no algoritmo do EnKF, quando o conjunto de previsões tem um papel mais importante na determinação dos desvios padrão nas estatísticas de inovação.

Do ponto de vista das diferenças entre os experimentos com o EnKF e o EnSRF, em geral, os resultados mostram que utilizando o sistema híbrido com o EnSRF, as diferenças entre os *priors* e *posteriors* são maiores do que o que foi encontrado com o EnKF. As Figuras 5.2(d), 5.2(e) e 5.2(f) com as estatísticas de inovação do conjunto para a variável T nas regiões HN, TR e HS (independentemente da contribuição relativa das covariâncias do conjunto), mostram que estas diferenças são mais evidentes no HN, em que T é mais próxima de zero com o EnSRF do que com o EnKF. Este é um resultado desejável, uma vez que o desvio padrão dos da inovação dos *priors* se equipara com o espalhamento total do conjunto.

O desvio padrão do vento horizontal (uv) sobre a região HN (Figura 5.2(a)), é muito próximo do espalhamento total do conjunto, indicando que o ajuste do conjunto de previsões as observações foi bom e que a quantidade de espalhamento parece ser razoável. Na região TR (Figura 5.2(b)), entretanto, foi encontrado o contrário, em que durante todo o período de simulações, o desvio-padrão dos *priors* dominou a relação com o espalhamento total do conjunto. Na região HS (Figura 5.2(c)), ocorreu uma situação semelhante ao que foi encontrado na região HN em que a amplitude do sinal da inovação do conjunto aumenta, indicando, possivelmente, deficiências na inflação no conjunto devido ao reduzido número de observações na região.

Figura 5.1 - Viés ponderado da inovação dos conjuntos de análises e previsões dos experimentos, válido para as 00Z.



Em vermelho (azul) estão representadas as inovações com 50% (75%) de contribuição das covariâncias do EnKF. Em laranja (verde), as inovações com 50% (75%) de contribuição do EnSRF. As linhas pontilhadas representam os *priors* e as linhas sólidas, os *posteriors*. A esquerda, os resultados para o Hemisfério Norte (HN); meio para os Trópicos (TR) e a direita, Hemisfério Sul (HS).

Fonte: Produção do autor.

5.1.2 Inovação dos Conjuntos de Análises

Outra forma de se diagnosticar o condicionamento dos conjuntos de análises e previsões utilizadas no processo de assimilação de dados do sistema híbrido 3DVar, é a partir da verificação da razão entre o desvio-padrão das inovações e a raiz quadrada do espalhamento total, como mostrado na Eq. 5.2. Neste caso, como o desvio padrão das inovações tem a mesma ordem de grandeza que a raiz quadrada do espalhamento total (representado pela soma entre os escalares, espalhamento do conjunto S e erro das observações R), logo, espera-se que as curvas representando as inovações do conjunto se aproximem da unidade, indicando que o conjunto de análises (representado pelos *posteriors*) está bem condicionado (i.e., à medida em que o sistema cíclico de análises e previsões evolui no tempo, a razão entre as inovações e o espalhamento deve se equilibrar). Desta forma, espera-se também observar a redução das amplitudes das curvas dos *posteriors* em relação aos *priors*, sendo este um indicativo de que as observações corrigiram as previsões do conjunto, enquanto a incerteza inerente ao conjunto de previsões também é mantida.

$$IC = \frac{\sigma(\mathbf{y}^o - \mathbf{H}\mathbf{x}_k^b)}{\sqrt{S + R}} \quad (5.2)$$

onde:

- σ : é o desvio-padrão;
- \mathbf{y}^o : é vetor observação;
- \mathbf{x}_k^b : é o k -ésimo membro do conjunto de K previsões;
- S : é o espalhamento do conjunto (escalar);
- R : é o erro (escalar) da observação, proveniente da matriz de covariâncias dos erros de observação (\mathbf{R}).

Na Figura 5.2 estão indicados os resultados obtidos a partir dos experimentos com o sistema híbrido 3DVar, na avaliação das estatísticas de inovação. Para o vento horizontal (uv), o EnSRF foi melhor (em geral) no ajuste sobre as regiões HN (Figura 5.3(a)) e HS (Figura 5.3(c)). Na região TR (Figura 5.3(b)), embora a amplitude das curvas tenham menor magnitude, ambos os algoritmos EnKF e EnSRF apresentaram resultados semelhantes, porém mostrando que a incerteza associada à previsão e ao

erro na observação é maior. Neste aspecto, sobre o SH (Figura 5.3(c)), encontra-se o melhor ajuste entre a inovação das observações e o espalhamento total. Na avaliação desta variável, encontrou-se também que o algoritmo EnSRF com 75% de contribuição das covariâncias do conjunto para as covariâncias estáticas, mostrou maiores diferenças entre *priors* e *posteriors*.

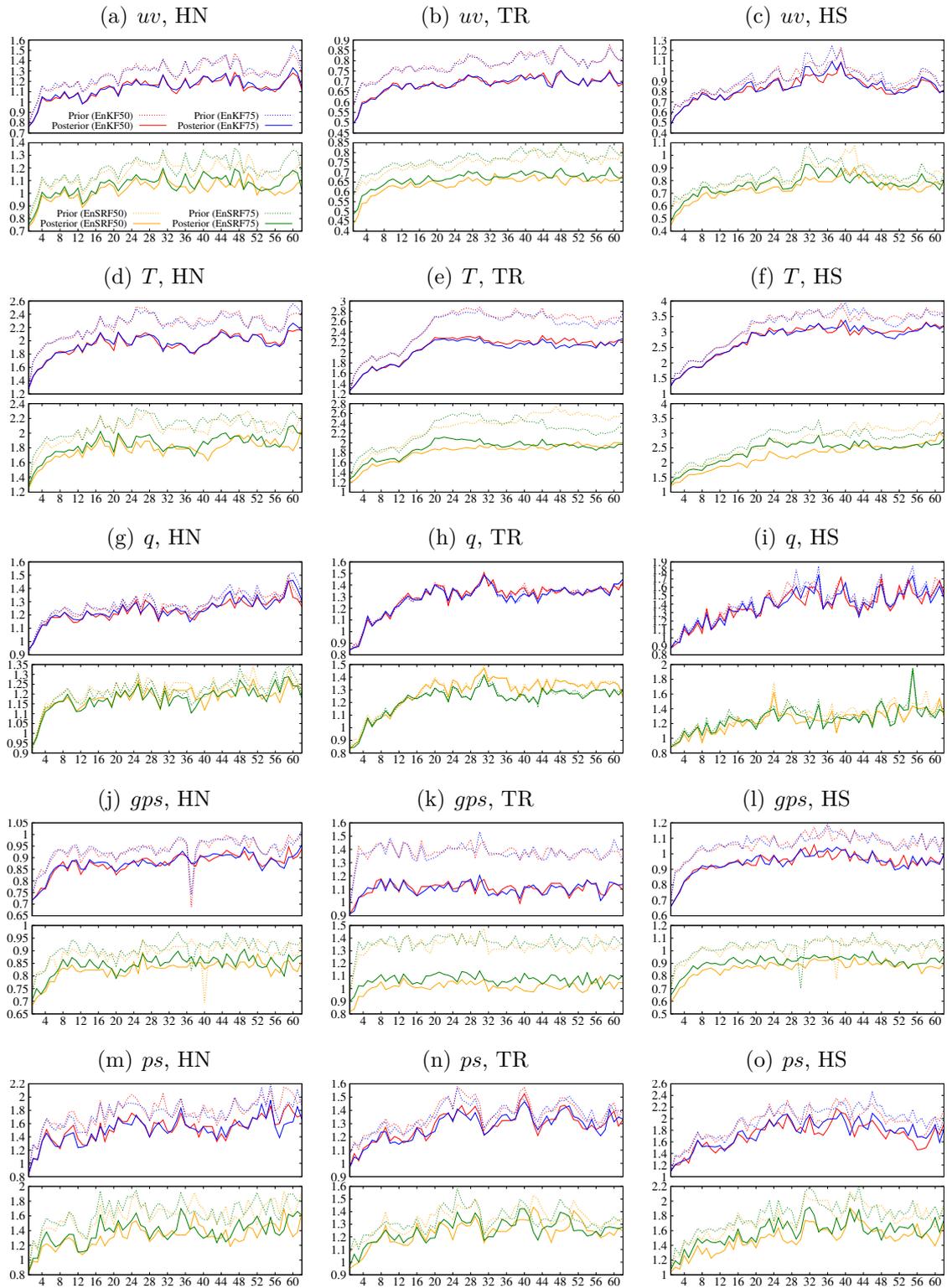
Para a temperatura (T), sobre todas as regiões os altos valores representados pela variação de amplitude das curvas (todas elas iniciando-se acima de 1) mostram que a inovação das observações foi maior do que o espalhamento do conjunto durante todo o período do experimento. Sobre a região HS (Figura 5.3(f)), além de se observar maior diferenças entre as curvas representando as contribuições de 50 e 75%, observa-se também uma tendência positiva, a qual pode ser um indicativo de deficiências na manutenção do espalhamento dos conjuntos dos *priors* e *posteriors*). De forma semelhante ao que se encontrou para uv , o algoritmo EnSRF com 75% de contribuição das covariâncias, foi o que apresentou melhores resultados na relação entre as inovações e o espalhamento do conjunto.

Os dados assimilados de rádio ocultação GPS (representados por *gps*) trazem contribuições para a análise dos campos de umidade do modelo. De forma geral, em relação a esta variável a partir deste tipo de observação, observa-se o ajuste entre as inovações e o espalhamento dos conjuntos, foi melhor nas regiões HN e HS (Figuras 5.3(j) e 5.3(l)) do que na região TR (Figura 5.3(k)). Nesta região, encontrou-se diferenças maiores entre *priors* e *posteriors*, indicando maiores correções nos campos de previsão de umidade. Estas correções não indicam, necessariamente, que o processo de assimilação de dados foi deficiente, mas que as observações de rádio ocultação GPS podem ser bastante importantes sobre a região. Além disso, o algoritmo EnSRF foi mais eficiente na balanceamento entre as inovações e o espalhamento total do conjunto, uma vez que este foi capaz de reduzir mais a amplitude da curva dos *posteriors*, sobretudo com 75% de contribuição das covariâncias do conjunto.

O ajuste entre as inovações e o espalhamento dos conjuntos de previsões e análises em relação às observações de pressão em superfície (ps), assim como para a temperatura (T), indicam deficiências no espalhamento do conjunto para estas variáveis, pois a razão entre as quantidades consideradas é alta logo no início dos ciclos e tende a aumentar até o final dos experimentos. As diferenças entre *priors* e *posteriors*, entretanto, são bastante discretas (tal como para a umidade (q)), mas com a diferença de que sobre a região HS (Figura 5.3(o)), o EnSRF com 75% de contribuição foi melhor, principalmente nos primeiros 30 dias de ciclos de análises e previsões.

Para todas as variáveis avaliadas, encontrou-se que com o algoritmo EnKF não houve diferenças entre realizar o sistema híbrido com 50% ou 75% de contribuição das covariâncias do conjunto nas covariâncias estáticas.

Figura 5.2 - Estatísticas de inovação dos conjuntos de análises dos experimentos, válido para as 00Z.



Em vermelho (azul) estão representadas as inovações com 50% (75%) de contribuição das covariâncias do EnKF. Em laranja (verde), as inovações com 50% (75%) de contribuição do EnSRF. As linhas pontilhadas representam os *priors* e as linhas sólidas, os *posteriors*. A esquerda, os resultados para o Hemisfério Norte (HN); meio para os Trópicos (TR) e a direita, Hemisfério Sul (HS).

Fonte: Produção do autor.

5.1.3 Habilidade das Previsões até 120 horas

Os resultados apresentados a seguir referem-se a habilidade da previsão do modelo BAMv0 para até 120 horas (5 dias) a partir das análises dos experimentos REF, 3DVar puro e híbridos 3DVar (conforme discutido na Seção 4.5). Em todos os experimentos, foram utilizadas as mesmas configurações do modelo BAMv0 na resolução TQ0062L028 e, nos experimentos com ciclos de assimilação de dados, a versão da matriz de covariâncias estática é a mesma.

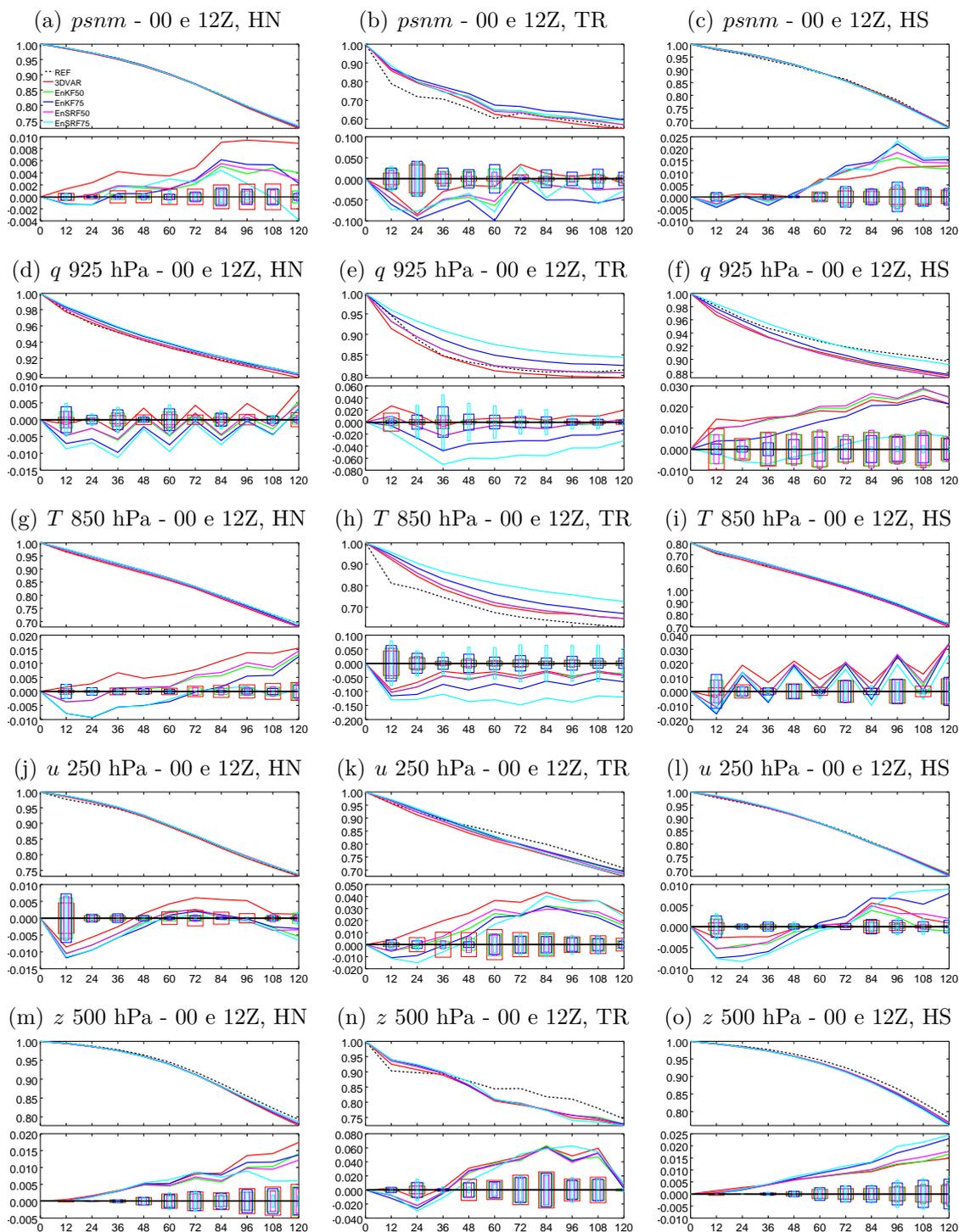
A habilidade de previsão do modelo BAMv0 para até 5 dias foi avaliada em termos da Correlação de Anomalia (CA). Como os desenvolvimentos do CPTEC tem foco sobre as regiões Tropical e América do Sul, a habilidade de previsão do modelo de circulação geral do CPTEC foi feita para as seguintes regiões: Globo (GL), Hemisfério Norte (HN), Trópicos (TR), Hemisfério Sul (HS) e América do Sul (AS). Para se verificar o quão diferentes são os experimentos entre si, um teste de significância t-Student foi aplicado para cada caso com um nível de confiança de 95%. Para cada figura de correlação de anomalia (painéis de cima), tem-se portanto, uma figura em anexo (painéis de baixo) mostrando os resultados do teste. No teste de significância, há uma curva correspondente a cada experimento (exceto o experimento de referência, REF). Quando as curvas cruzam suas respectivas caixas (cada caixa e curva possui uma cor distinta), há a indicação de que a diferença entre o experimento (e.g., experimento EnKF50) e a referência (e.g., experimento REF) não é significativa, e que, portanto, o teste da hipótese nula falha. Nesta avaliação, a hipótese nula estabelece que a média do experimento (híbridos 3DVar ou 3DVar puro) verificado é estatisticamente indistinguível da média do experimento de referência.

Na Figura 5.3 são apresentados os resultados da CA para as previsões de até 5 dias com o teste t-Student dos experimentos para as regiões HN, TR e HS. A Figura 5.4 apresenta os resultados para as regiões GL e AS. As variáveis avaliadas são a pressão em superfície $psnm$, a umidade específica em 925 hPa (q_{925}), a temperatura do ar em 850 hPa (T_{850}), a componente zonal do vento em 250 hPa (u_{250}) e a altura geopotencial em 500 hPa (z_{500}). Em geral, os experimento EnKF75 e EnSRF50 indicam as análises que produziram as previsões com maiores valores de correlação, para todas as regiões e variáveis avaliadas. A $psnm$ sobre o HN (Figura 5.4(a)) não apresenta diferenças distinguíveis entre os experimentos EnKF75 e EnSRF50, para as primeiras 24 horas de previsão. Este resultado indica que na análise da pressão em superfície do sistema 3DVar, pouco da contribuição dos 50% das covariâncias do conjunto de previsões foi utilizada ou que realmente modificou as covariâncias

estáticas. Isto pode ser devido ao fato de que as observações utilizadas pelo EnSRF não foram perturbadas, o que por sua vez pode não ter sido benéfico para a análise híbrida deste sistema.

Na região TR, - uma região particularmente de difícil previsão - as previsões de $psnm$, $q925$ e $T850$ (Figuras 5.4(b), 5.4(e) e 5.4(h)), para ambos os experimentos EnKF75 e EnSRF50, mostraram-se com boa performance em relação aos experimentos REF e 3DVar. As melhorias em relação do experimento REF para a $psnm$ (Figura 5.4(b)), são de quase 24 horas (considerando o limite de 80% de CA). A Figura 5.4(e) mostra as melhorias de $q925$ apresentadas pelo experimento EnSRF75 em relação ao experimento 3DVar com praticamente 4 dias de antecedência (considerando-se 85% de CA). Uma melhoria semelhante foi também encontrada para a $T850$ (Figura 5.4(h)), em comparação com o experimento 3DVar, em que o experimento EnSRF75 permanece com 80% de CA na previsão de 72 horas, enquanto o experimento 3DVar limita sua habilidade de previsão em apenas 36 horas, considerando-se o mesmo nível de habilidade de previsão que o experimento EnSRF75. Nas Figuras 5.4(b), 5.4(h) e 5.4(n) pode-se observar o tempo que o modelo BAMv0 leva para se estabilizar quando iniciado com as análises do NCEP (experimento REF). No entanto, este efeito não foi observado para todas as variáveis. Na Figura 5.4(g), por exemplo, indica que há não diferenças práticas entre os experimentos com CA semelhante. O teste t-Student, entretanto, revela que o experimento 3DVar não falha o teste da hipótese nula para as 120 horas de previsão, enquanto que o resto dos experimentos falham no teste em até 36 horas (experimentos EnKF50 e EnSRF50) e 72 horas (EnKF75 e EnSRF75). Por outro lado, os experimentos com 75% de contribuição das covariâncias do conjunto, são estatisticamente diferentes do experimento REF até 72 horas de previsão.

Figura 5.3 - Correlação de Anomalia com teste de significância t-Student (95%) das previsões para até 120 horas para os horários das 00 e 12Z, sobre as regiões Hemisfério Norte (HN), Trópicos (TR) e Hemisfério Sul (HS).



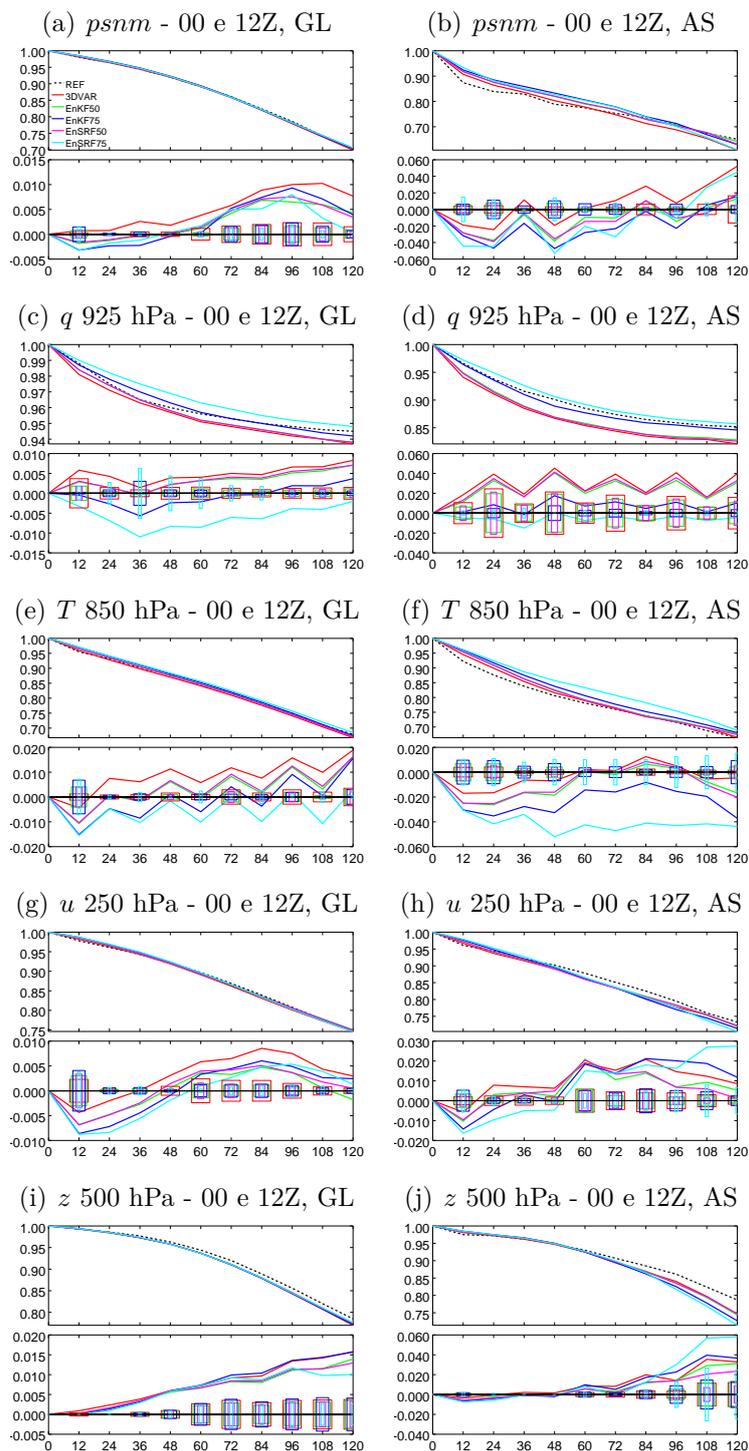
As variáveis avaliadas são pressão em superfície ($psnm$), umidade em 925 hPa (q_{925}), temperatura em 850 hPa (T_{850}), vento zonal em 250hPa (u_{250}) e altura geopotencial em 500 hPa (z_{500}). As curvas pretas pontilhadas, indicam o experimento REF; em vermelho está representado o experimento 3DVar; em verde, o experimento EnKF50; em azul, EnKF75; em rosa, EnSRF50 e em verde água, EnSRF75.

Fonte: Produção do autor.

A Figura 5.4 apresenta a mesma avaliação da Figura 5.3, mas para as regiões GL e AS. A região AS é de grande interesse para o CPTEC por ser a região alvo de previsões do Brasil. Nestas regiões, a mesma resposta dos experimentos EnKF75 e EnSRF50 foi encontrada, indicando que na região AS, as covariâncias do conjunto de previsões também tem um papel importante na determinação da habilidade de previsão do modelo BAMv0. Para a região GL, os índices de CA são semelhantes ao que foi encontrado para a região HN. Além disso, a CA para a variável $q925$ na região GL (Figura 5.5(c)) mostra que as previsões obtidas com o experimento EnSRF75 (e também com o experimento EnKF75) foram boas para 120 horas, ficando próximo de 95%. Neste caso, entretanto, o teste t-Student revelou que o experimento EnKF75 falha no teste da hipótese nula em 72 horas de previsão, quando a análise do experimento REF fez com que o modelo BAMv0 produzisse previsões com melhor correlação com a climatologia. As demais variáveis (i.e., $psnm$, $T850$, $u250$ e $z500$ (Figuras 5.5(a), 5.5(e), 5.5(g) e 5.5(i) respectivamente) mostram índices de CA semelhantes em comparação com os resultados obtidos sobre a região HN. Para a região AS, bons resultados foram obtidos para $psnm$ (Figura 5.5(d)), para até 84 horas de previsão; $q925$ (Figura 5.5(b)) com CA maior do que 85% para 120 horas e $T850$ (Figura 5.5(f)) com CA de 80% com os experimentos EnSRF75 até 72 horas de previsão e EnKF75 para até 60 horas de previsão. O vento zonal em 250 hPa ($u250$) e, especialmente, a altura geopotencial em $z500$ (Figuras 5.5(h) e 5.5(j), respectivamente) não apresentaram bom desempenho. $z500$ nos experimentos EnKF75 e EnSRF75 não desempenharam como as demais variáveis. Para além de 84 horas de previsão, elas foram piores e os experimentos EnKF50, EnSRF50 e mesmo o experimento 3DVar, foram melhores.

Uma comparação entre os esquemas de análise utilizados nos experimentos, mostram que a análise da umidade dos experimentos EnKF75 e EnSRF50 apresentaram melhora em relação do experimento 3DVar. As Figuras 5.4(c), 5.4(f), 5.4(i), 5.5(c) e 5.5(f), mostram que as previsões de umidade foram piores com os ciclos de assimilação de dados se comparados com o experimento REF. Para os experimentos com o sistema híbrido 3DVar, a matriz estática dos erro de previsão é a mesma utilizada também no experimento 3DVar. Isto pode ser um indicativo de que as covariâncias dos conjuntos de previsões tem um papel importante também na definição de como os incrementos de análise são aplicados durante a minimização da função custo variacional.

Figura 5.4 - Idem Figura 5.3, mas para as regiões Global (GL) e América do Sul (AS).



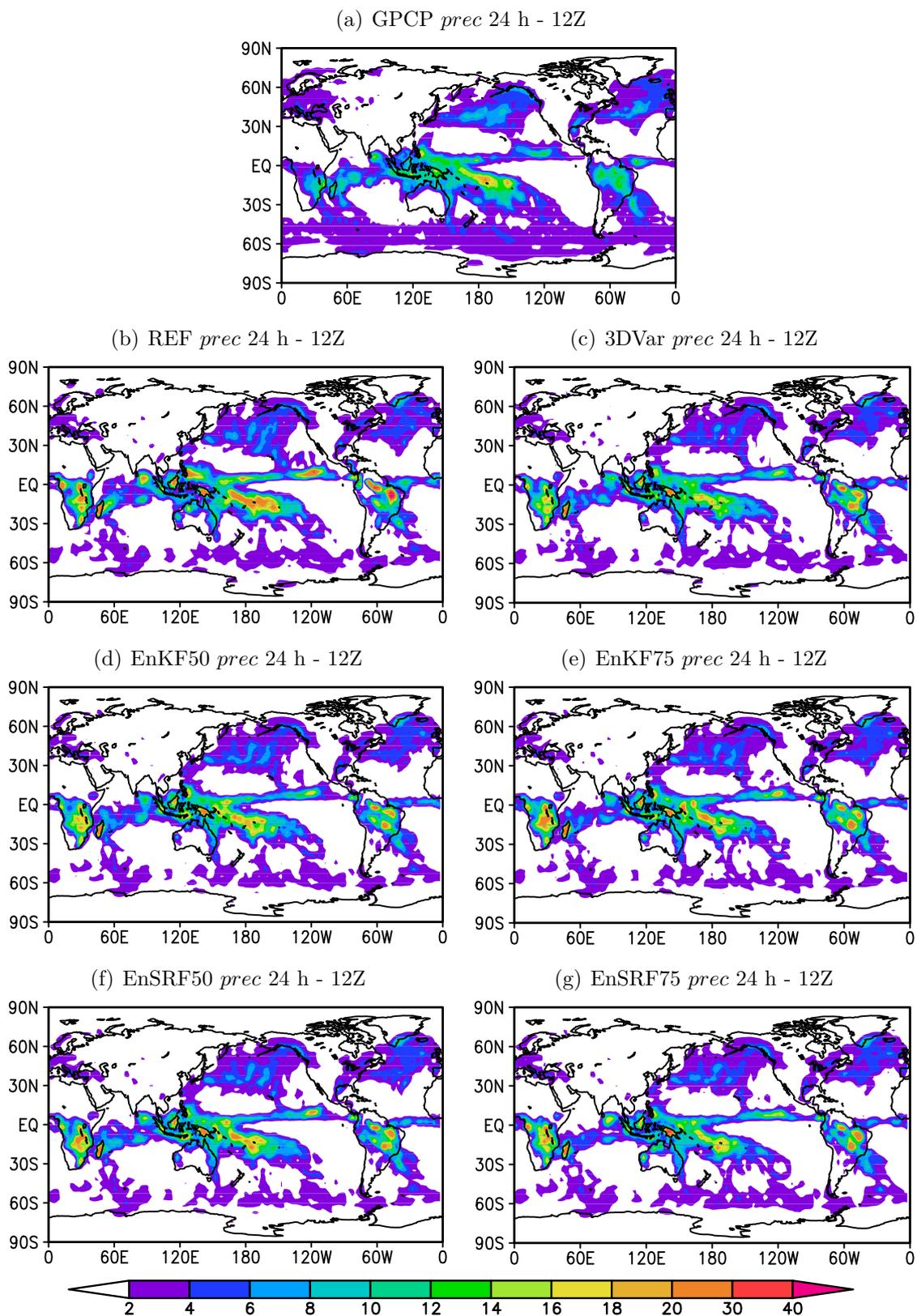
Fonte: Produção do autor.

5.1.4 Avaliação das Previsões de 24 horas de Precipitação

As previsões de precipitação de 24 horas produzidas a partir das análises dos experimentos REF, 3DVar puro e híbridos 3DVar foram avaliadas. A Figura 5.6(a) mostra a média mensal para Janeiro de 2013 da precipitação observada do *Global Precipitation Climatology Project v2.2* (GPCPv2.2) com resolução espacial de $2,5^\circ$ (ADLER et al., 2003) e as Figuras 5.6(b), 5.6(c), 5.6(d), 5.6(e), 5.6(f) e 5.6(g) mostram a precipitação de 24 horas prevista - representada como média mensal, dos experimentos realizados REF, 3DVar puro e híbridos 3DVar. A comparação entre as médias mensais das precipitações de 24 horas e o GPCPv2.2 é válida no horário das 12Z e os dados em ponto de grade dos modelos foram interpolados para $2,5^\circ$ para coincidir com a resolução dos dados observados do GPCPv2.2. O campo da média mensal de precipitação do GPCPv2.2 mostra as principais características da precipitação tropical e subtropical, com a maior parte da precipitação distribuída sobre a região tropical (entre 30S e 30N).

Na Figura 5.5, com excessão do experimento REF, todos os experimentos foram realizados com ciclos de assimilação de dados, utilizando todos os dados observados da Tabela 2.2, a mesma resolução e com as mesmas configurações do modelo de circulação geral (tal como definido na Tabela 2.1). Todos os experimentos mostram as características da precipitação convectiva e de larga escala e a Figura 5.6(a) é usada como referência para uma avaliação da intensidade e distribuição das médias espacial e temporal. A Tabela 5.1 apresenta os valores das médias espaciais das médias do período, para o GPCPv2.2 e os experimentos realizados.

Figura 5.5 - Médias espaciais das previsões de precipitação de 24 horas (em mm/dia) válido para Janeiro de 2013 às 12Z.



Em (a) precipitação do GPCPv2.2; (b) experimento REF; (c) 3DVar; (d) EnKF50; (e) EnKF75; (f) EnSRF50; (g) EnSRF75.

Fonte: Produção do autor.

As previsões de precipitação dos experimentos, apresentam distribuição espacial coerente em comparação com a média mensal do GPCPv2.2 (Figura 5.6(a)). Ressalta-se que a configuração do modelo para a condensação de larga escala, convecção *Cumulus* e difusões horizontal e vertical foram mantidas iguais, portanto, as diferenças observadas são devidas apenas as análises utilizadas. As principais diferenças entre as previsões de precipitação dos experimentos estão mais relacionadas a intensidade da precipitação. A Figura 5.6(a), mostra a média mensal da precipitação de 24 horas resultante do modelo BAMv0 inicializado com a análise do NCEP (experimento REF). Em comparação com os demais experimentos (incluindo a precipitação de referência do GPCPv2.2), o modelo BAMv0 apresentou tendência a produzir mais precipitação convectiva sobre a região tropical. A média espacial da média temporal resulta em 2,9718 mm/mês, contabilizando a maior quantidade de precipitação entre os experimentos realizados, sendo maior até do que a precipitação de referência do GPCPv2.2 (2,7197 mm/mês). Por outro lado, todos os experimentos com assimilação de dados, produziram precipitação com distribuição menos concentrada e médias espaciais mais razoáveis. A média espacial do experimento EnSRF50 (Figura 5.6(d)) é a que mais se aproxima com o valor apresentado pelo GPCPv2.2, embora os demais experimentos com assimilação de dados tenham distribuído a precipitação convectiva de forma mais coerente (em relação ao GPCPv2.2).

Tabela 5.1 - Médias espaciais sobre a região Global (GL) para a referência (GPCPv2.2) e os experimentos realizados das previsões de 24 horas da precipitação total às 12Z (em mm/mês). Em azul, estão indicados os experimentos cujas médias são mais próximas a média de precipitação da referência.

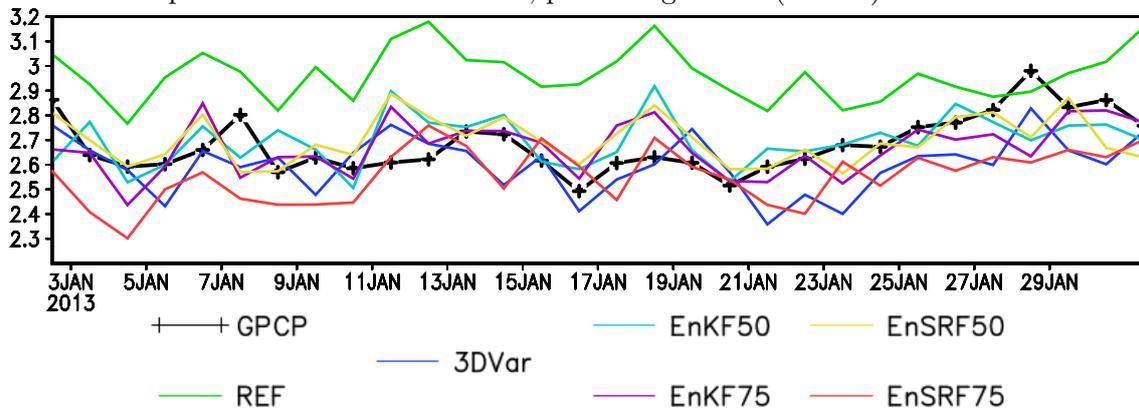
	μ
GPCPv2.2	2,7197
REF	2,9718
3DVar	2,6098
EnKF50	2,7034
EnKF75	2,6730
EnSRF50	2,7045
EnSRF75	2,5618

A quantidade exagerada de precipitação produzida pelo experimento REF (Figura 5.6(b)) pode ser devida ao fato de que este experimento foi feito simplesmente integrando-se o modelo BAMv0 com as análises do NCEP, embora a suavização

da topografia e inicialização com modos normais para cada análise no período de avaliação tenham sido feitas neste experimento em específico. Os experimentos com ciclos de assimilação de dados, com exceção do experimento EnSRF75 (Figura 5.6(g)) - o qual resultou na menor média mensal, mostrou-se estar bem condicionado as opções escolhidas para as previsões.

Para se tentar entender e identificar as regiões onde os experimentos acumularam mais ou menos a precipitação, as Figuras 5.6 e 5.7 apresentam as séries temporais das médias espaciais para cada experimento, incluindo o GPCP sobre as regiões GL, HN, TR, HS e AS. Nestas figuras, o GPCPv1.2 1DD (*1 Degree Daily* - Huffman et al. (2001)) foi utilizado como dado observado e este foi interpolado para a grade das análises dos experimentos (com $1,875^\circ$. Em adição as Figuras 5.6 e 5.7, a Tabela 5.2 mostra as médias espaciais (μ) e os desvios-padrão (σ) para cada experimento (incluindo o dado interpolado do GPCPv1.2) nas regiões avaliadas indicadas.

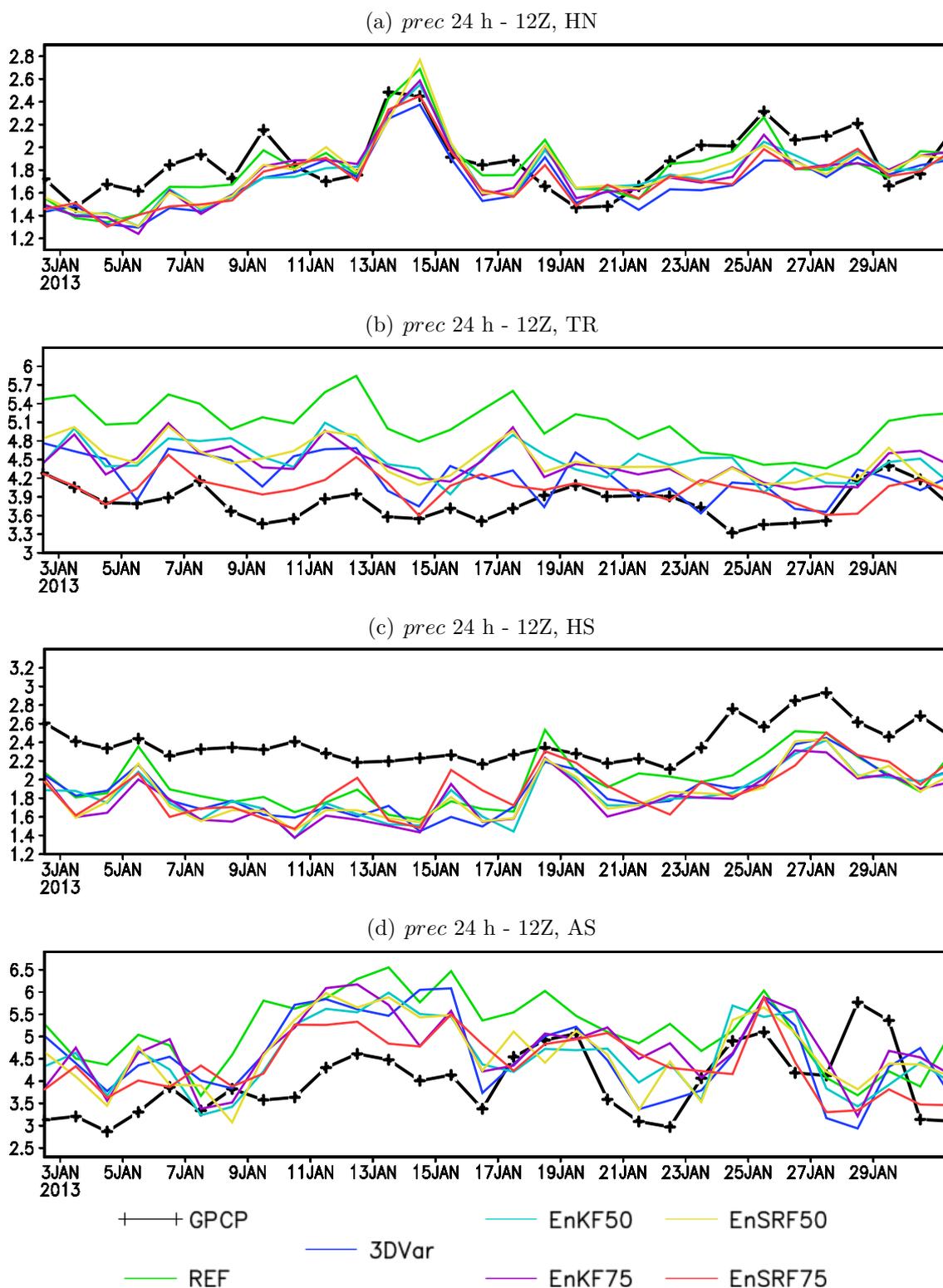
Figura 5.6 - Médias espaciais das previsões de precipitação de 24 horas (em mm/dia) válido para Janeiro de 2013 às 12Z, para a região GL (Global).



A precipitação dos experimentos é comparada com a precipitação observada do GPCPv1.2.

Fonte: Produção do autor.

Figura 5.7 - Idem Figura 5.6, para as regiões (a) precipitação para todo o Hemisfério Norte (HN); (b) Trópicos (TR); (c) Hemisfério Sul (HS); (d) América do Sul (AS).



A precipitação dos experimentos é comparada com a precipitação observada do GPCPv1.2.

Em contraste com o que foi encontrado com a análise da distribuição espacial dos campos de precipitação dos experimentos na Figura 5.5, a série temporal com as médias espaciais mostradas nas Figuras 5.6 e 5.7, indicam que - com excessão da região HN (Figura 5.8(a)), todos os experimentos apresentaram dificuldades em reproduzir a precipitação observada do GPCPv1.2 em previsões de 24 horas. Deficiências podem ser notadas em relação a amplitude (máximos e mínimos) e em relação a média. Em alguns casos específicos, a precipitação prevista foi praticamente igual aos valores médios observados - e.g., o experimento EnSRF75 na região TR (Figura 5.8(b)) se igualou ao GPCPv1.2 durante alguns períodos do mês. Este resultado está de acordo com o que foi encontrado na análise da habilidade do modelo na previsão de até 120 horas (Figuras 5.3 e 5.4), mostrando que entre os experimentos híbridos 3DVar realizados, a contribuição de 75% das covariâncias do EnSRF foi benéfica na maioria dos casos. Por outro lado, o experimento REF exagerou na precipitação ao longo do tempo, principalmente sobre as regiões GL e TR (Figuras 5.6 e 5.8(b)). Este é um resultado esperado, porque este experimento utilizou uma análise independente e o modelo não foi configurado para gerar o melhor resultado a partir desta análise. Para as regiões TR e AS (Figuras 5.8(b), 5.8(d), respectivamente), as previsões de precipitação foram todas praticamente superestimadas em relação aos valores observados do GPCPv1.2. Nestas regiões, a precipitação convectiva tem um papel importante e os modelos, em geral, tendem a exagerar na previsão de precipitação. Se considerarmos que a região HS (Figura 5.8(c)) - a qual contém mais partes oceânicas (em oposição a região HN), os modelos tenderão a produzir mais precipitação de larga escala enquanto que a precipitação convectiva tende a ser menos importante nos resultados encontrados. Conseqüentemente, a previsão de precipitação tende a ser subestimada em relação ao dado observado.

Tabela 5.2 - Médias espaciais, desvios-padrão e erro médio absoluto (MAE, em relação ao GPCP) das previsões de 24 horas da precipitação total às 12Z (em mm/mês) para as regiões Global (GL), Hemisfério Norte (HN), Trópicos (TR), Hemisfério Sul (HS) e América do Sul (AS). Em azul, estão indicados os experimentos cujas médias são mais próximas a média de precipitação do GPCPv1.2 (considerando o desvio-padrão do GPCP). Em verde, estão indicados os valores de MAE mais baixos em relação ao GPCP.

	GL			HN			TR			HS			AS		
	μ	σ	MAE	μ	σ	MAE	μ	σ	MAE	μ	σ	MAE	μ	σ	MAE
GPCPv1.2	2,8635	0,1123	-	1,7228	0,2653	-	4,2820	0,2694	-	2,6068	0,2053	-	3,1321	0,7743	-
REF	3,0478	0,1036	0,2824	1,5547	0,2842	0,0575	5,4655	0,3733	1,2610	2,0758	0,2685	0,4111	5,2812	0,7786	1,2175
3DVar	2,7601	0,1118	0,0799	1,4344	0,2438	0,1775	4,7657	0,3420	0,4142	2,0499	0,2512	0,5186	5,0152	0,8779	0,6738
EnKF50	2,6107	0,0999	0,0161	1,4799	0,2576	0,1165	4,4379	0,2849	0,6743	1,8863	0,2459	0,5572	4,3304	0,7679	0,6704
EnKF75	2,6620	0,1057	0,0130	1,4948	0,2744	0,1162	4,4412	0,2835	0,6256	2,0280	0,2502	0,5959	3,8494	0,7618	0,8231
EnSRF50	2,8103	0,0935	0,0188	1,5605	0,2779	0,0960	4,8385	0,2990	0,6637	1,9999	0,2546	0,5579	4,6480	0,7768	0,7018
EnSRF75	2,5745	0,1089	0,1250	1,4566	0,2520	0,1428	4,2710	0,2209	0,2306	1,9756	0,2619	0,4955	3,8128	0,6702	0,4778

Todos os experimentos com a matriz de covariâncias híbrida mostraram resultados satisfatórios (i.e., para estes caso, em relação a média do GPCP, as médias dos experimentos com as matrizes híbridas possui desvio padrão que varia entre 0,0432 para o EnSRF50 e 0,289 para EnSRF75) na previsão de precipitação de 24 horas, se comparado com os experimentos 3DVar e REF (com desvios-padrão de 0,1034 e 0,1843, respectivamente em relação ao GPCP). Apesar de haver maior variância entre os resultados com as matrizes híbridas, há que se considerar também o fato de que dentro deste intervalo de variação dos desvios-padrão, o experimento EnSRF50 esteve mais próximo da média. No entanto, a precipitação convectiva parece não ser tão sensível as mudanças na matriz de covariâncias, apesar de que os experimentos com maior contribuição das covariâncias do conjunto (i.e., EnKF75 e EnSRF75 sobre a AS - Tabela 5.2) produzem melhores precipitações do que os demais experimentos.

Na Tabela 5.2 estão também indicados os valores do Erro Médio Absoluto da precipitação dos experimentos em relação ao GPCP. Considerando estas informações, sobre todo o globo (região GL), o experimento EnKF75 foi aquele que produziu precipitação mais próxima do observado, diferentemente do que pôde-se concluir em relação às médias consideradas e o desvio-padrão do GPCP. Para a região HN, dentre todos os experimentos que foram destacados em azul, apenas os experimentos REF e EnSRF50 são destacados, considerando-se os menores valores de MAE. Na região TR, o experimento EnSRF75 gerou o menor MAE e foi também o que apresentou a melhor média (em comparação com o GPCP). Sobre a região HS, embora nenhum experimento tenha aproximado a média do GPCP (considerando o desvio-padrão de 0,2053), os experimentos REF e EnSRF75 produziram precipitação sobre a região

com os menores erros. Considerando-se apenas a América do Sul (região AS), entre os experimentos destacados em azul, apenas o experimento EnSRF75 produziu precipitação mais próxima do GPCP.

5.2 Estudo de Caso

Segundo [Cavalcanti et al. \(2013\)](#) - doravante referenciado como “Boletim Climanálise V28 N01”, a banda característica de nebulosidade associada à ZCAS, esteve ativa em três períodos distintos de Janeiro de 2013 (Tabela 5.3), dos quais o primeiro e o último foram mais significativos, produzindo acumulados de chuva mais expressivos. Considerando a Figura 5.8(d), nestes três períodos verificou-se que os experimentos apresentaram tendências no acumulado de chuva um pouco mais elevados do que o representado pelo GPCPv1.2, especialmente no período PI. Nos períodos PII e PIII, entretanto, os acumulados apresentados pelos experimentos não foram tão proeminentes quanto aqueles apresentados no período PI, e além disso, no período PIII, a precipitação do GPCPv1.2 foi maior do que a dos modelos. Diante deste fato, o período escolhido para a aplicação das análises e previsões produzidas nos experimentos com o sistema híbrido 3DVar foi o PI, em que as precipitações obtidas nos experimentos foram maiores do que o observado pelo GPCPv1.2.

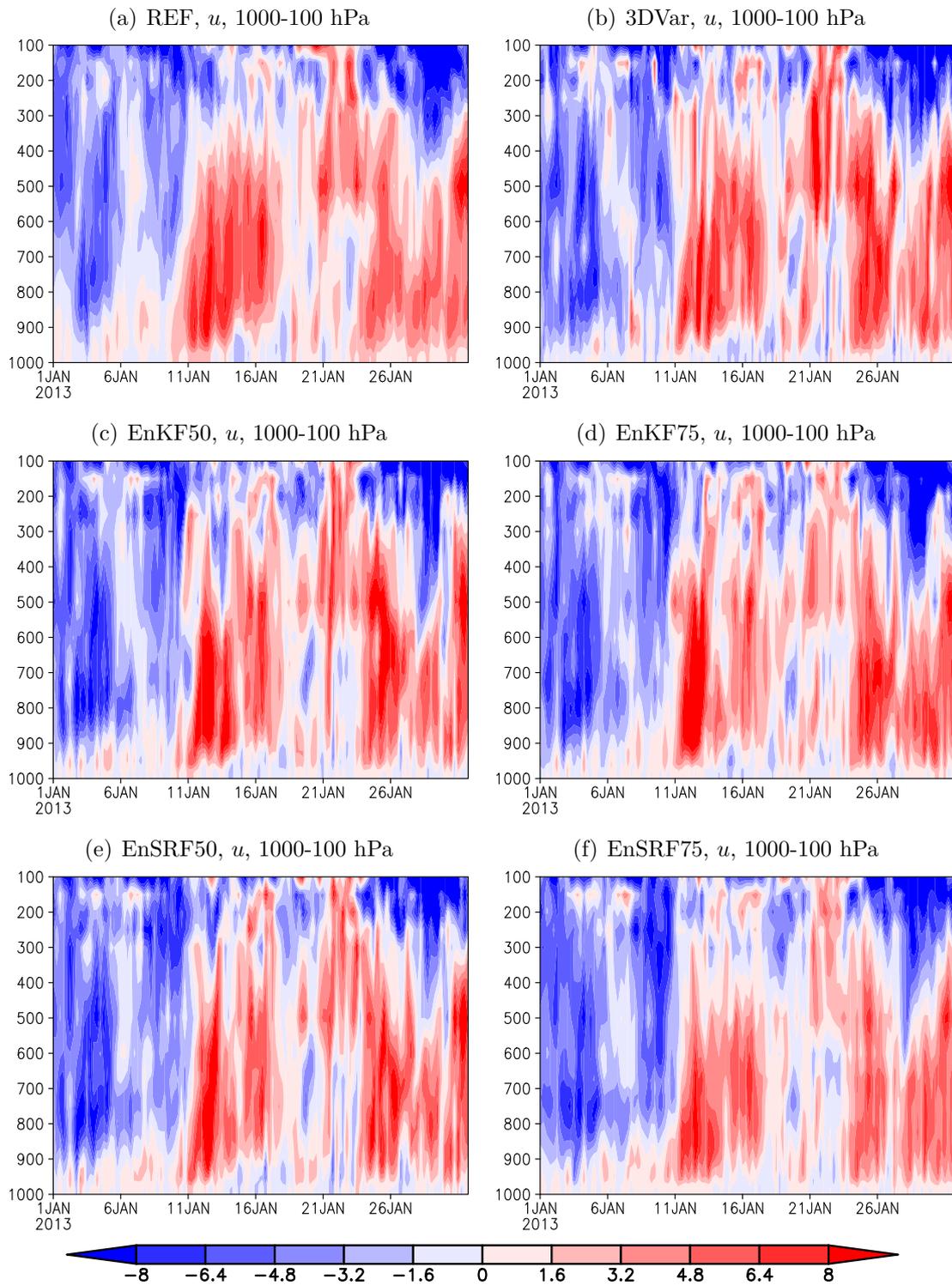
Tabela 5.3 - Períodos de atividade e duração das ZCAS em Janeiro de 2013.

	Janeiro de 2013	Duração
PI	09 a 14	6 dias
PII	19 a 23	5 dias
PIII	26 a 31	6 dias

Para se verificar como e se os experimentos com o sistema 3DVar foram capazes de detectar o sinal característico da ZCAS nos períodos considerados, na Figura 5.8, são apresentadas as séries temporais dos perfis verticais da componente zonal do vento horizontal (u) durante Janeiro de 2013. O critério utilizado para a identificação dos episódios de ZCAS, segue aquele utilizado por [Herdies et al. \(2002\)](#), em que a componente u do vento horizontal sendo positiva ao nível de 850 hPa, indica a atividade de um episódio de ZCAS, enquanto que valores negativos desta variável, indicam a sua ausência. A partir deste critério, verificou-se que todos os experimentos apresentaram a característica de aceleração da componente u no nível referenciado. Entretanto, destacam-se as seguintes características: o experimento REF (Figura

5.9(a)), mostra a atividade da componente u em 850 hPa no período PI (09 a 14 de Janeiro) com maior persistência, isto é, iniciando-se um pouco antes do dia 11 de Janeiro e estendendo-se para além do dia 16 de Janeiro; os demais experimentos apresentam características semelhantes entre si, mas diferenciam-se do experimento REF quanto ao início, fim e intensidade do vento avaliado. Os experimentos EnKF50 e EnKF75 (Figuras 5.9(c) e 5.9(d), respectivamente), mostram que no período PI, a intensidade da componente u do vento é mais intensa e é limitada entre os dias 11 e 16 de Janeiro. O experimento EnSRF50 (Figura 5.9(e)), também apresenta esta característica, mas com menor intensidade. Já o experimento EnSRF75 (Figura 5.9(f)), é o experimento que apresenta atividade da componente u do vento com menor intensidade, apesar de também identificar a atividade da ZCAS segundo o critério adotado.

Figura 5.8 - Séries temporais do perfil vertical da componente zonal (u) do vento horizontal durante Janeiro de 2013, para o experimentos realizados.



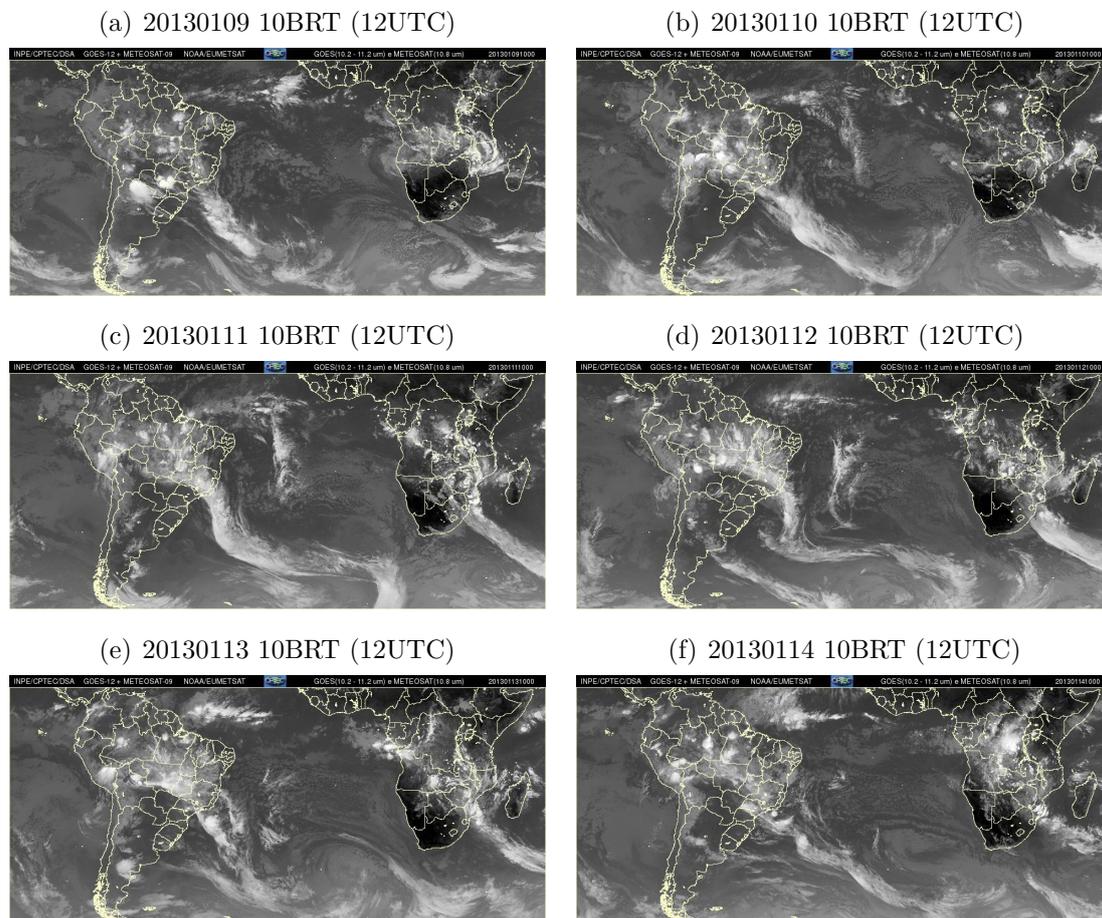
Fonte: Produção do autor.

Para o período PII (19 a 23 de Janeiro), conforme descrito no “Boletim Climanálise”, a atividade da banda de nebulosidade associada à ZCAS foi menos intensa, e seguindo o critério adotado, os experimentos também evidenciam isto. Para este período, fazendo-se uso apenas do critério do vento, apenas os experimentos 3DVar, EnKF50 e EnKF75 (Figuras 5.9(b), 5.9(c) e 5.9(d), respectivamente) apontaram atividade mais intensa no dia 21 de Janeiro. No período PIII (26 a 31 de Janeiro), os experimentos 3DVar, EnKF50, EnKF75 e EnSRF50 (Figuras 5.9(b), 5.9(c), 5.9(d) e 5.9(e), respectivamente) mostram maior intensidade da componente u do vento, embora o experimento REF (Figura 5.9(a)) apresente intensidade da componente mais consistente dentro do período avaliado.

A partir dos perfis verticais para o mês de Janeiro de 2013 apresentados na Figura 5.8, verificou-se que a componente zonal do vento horizontal esteve mais acelerada (intensa) durante os períodos PI e PIII. As Figuras 5.9 e 5.10 a seguir, apresentam as sequências de imagens infravermelho do satélite GOES 12 (em composição com infravermelho do satélite METEOSAT 9) para os períodos PI e PIII, respectivamente. As imagens são mostradas no horário das 10BRT (ou 12UTC) para cada dia dos períodos indicados.

Na sequência de imagens da Figura 5.9, observa-se uma banda de umidade se organizando entre os dias 09 e 10 de Janeiro, com intensa atividade sobre o continente nos dias 12 e 13 de Janeiro e sua desintensificação a partir do dia 14 de Janeiro. Segundo o boletim climanálise, esta banda de nebulosidade atuou principalmente sobre o centro das regiões sudeste e centro-oeste do Brasil. As situações dinâmicas que deram suporte a este episódio de ZCAS, incluem uma circulação ciclônica na baixa troposfera atuando no interior do continente e um vórtice ciclônico na alta troposfera, atuando sobre o Oceano Atlântico.

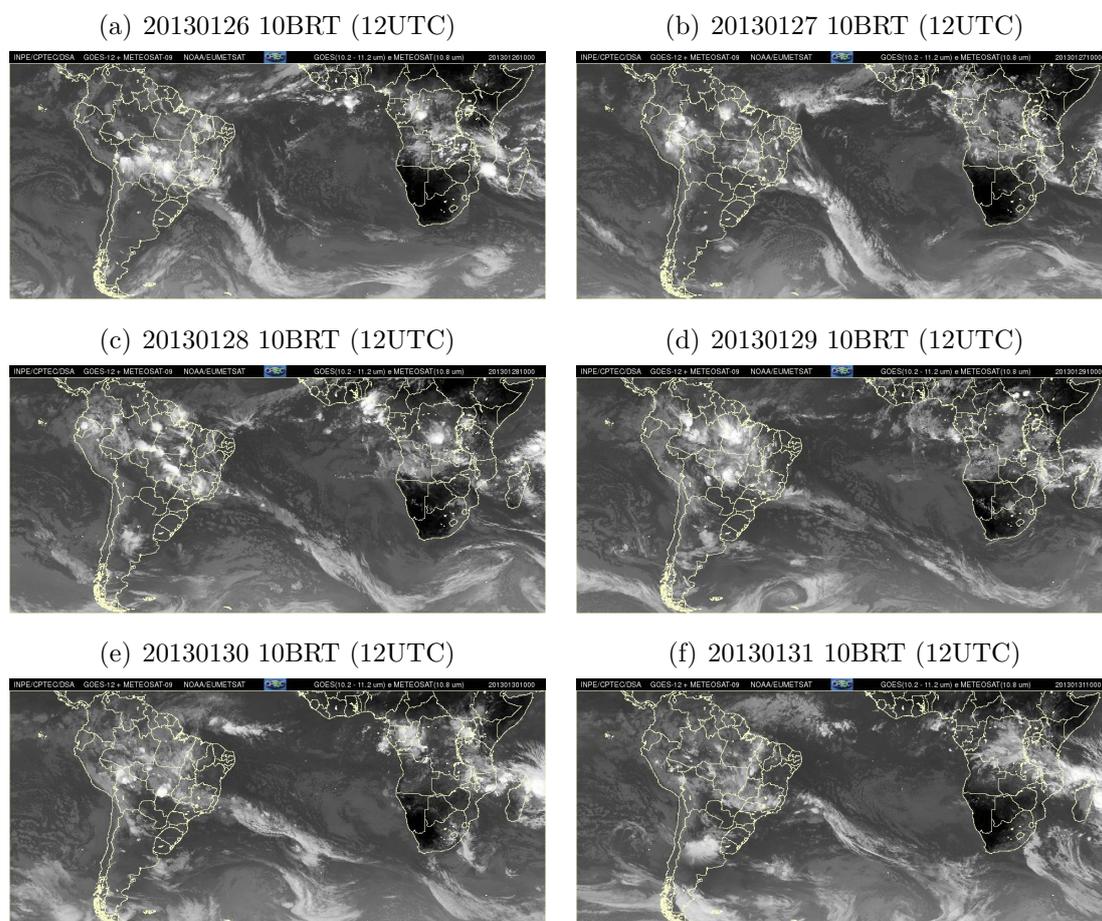
Figura 5.9 - Imagens infravermelho dos satélites GOES 12 (10.2 - 11.2 μm) em composição com o METEOSAT 9 (10.8 μm), entre os dias 09 e 14 de Janeiro de 2013 às 10BRT (12UTC) mostrando o continente Sul-Americano, parte do continente Africano, partes do Oceano Pacífico e o oceano Atlântico Sul.



Fonte: <http://satelite.cptec.inpe.br/acervo/>

Semelhante ao que foi apresentado na Figura 5.9, a sequência de imagens da Figura 5.10, apresenta a organização da banda de nebulosidade e a evolução do fenômeno ZCAS ocorrido no período PIII. Neste período, a banda de nebulosidade tem sua organização a partir do dia 26 de Janeiro, e sua maior intensidade sobre o continente Sul-Americano se dá entre os dias 28, 29 e 30 de Janeiro. Neste segundo episódio de ZCAS, o suporte dinâmico foi semelhante ao do primeiro episódio, com a diferença de que houve a atuação da Alta Bolívia sobre a região central do continente Sul-Americano que, em associação com a circulação ciclônica na baixa troposfera, deu suporte para a manutenção do sistema neste período.

Figura 5.10 - Idem Figura 5.9, mas entre os dias 26 e 31 de Janeiro de 2013.

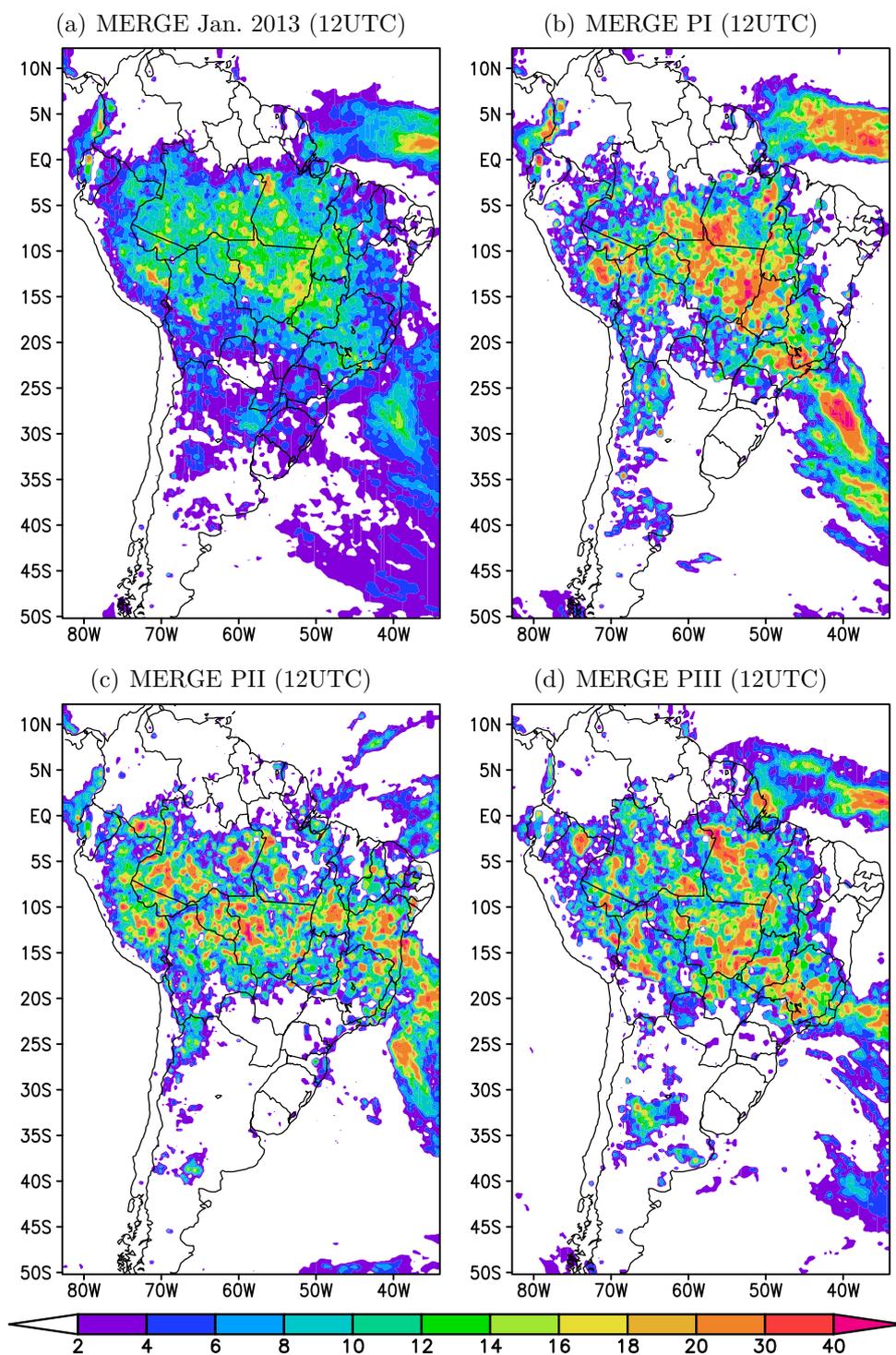


Fonte: <http://satelite.cptec.inpe.br/acervo/>

Na Figura 5.11 estão representadas as médias da precipitação observada (provenientes do produto MERGE do CPTEC), para todo o mês de Janeiro e os períodos individuais PI, PII e PIII. A atividade convectiva no período PI, em comparação com os demais períodos, mostra que a precipitação esteve mais distribuída pela região central do país, envolvendo boa parte da região centro-oeste e sudeste, conforme mostra a Figura 5.12(b). Neste mesmo período, a precipitação esteve também mais concentrada na porção central do continente, provocando acumulados significativos sobre a região centro-oeste. Nos períodos PII e PIII (Figuras 5.12(c) e 5.12(d), respectivamente), a atividade convectiva resultou em uma distribuição da precipitação que alcançou parte da região nordeste, deixando a porção sul das regiões centro-oeste e sudeste livres. No terceiro período, entretanto, a banda de nebulosidade voltou a atuar sobre a porção sul das regiões centro-oeste, mas não atuando sobre o estado do Paraná

como no período PI.

Figura 5.11 - Média da precipitação MERGE para Janeiro de 2013 às 12UTC (em mm/dia).



Nas figuras, estão representadas: (a) média para todo o mês; (b) média para o período PI; (c) média para o período PII; (d) média para o período PIII.

Fonte: <http://ftp.cptec.inpe.br/modelos/io/produtos/MERGE/2013/>

Para se validar a aplicação do sistema híbrido 3DVar na simulação da atmosfera real, elegeu-se o experimento EnKF75, a partir do qual foram avaliadas as análises e as previsões na simulação de dois episódios da Zona de Convergência do Atlântico Sul (ZCAS) durante o mês de Janeiro de 2013. O experimento EnKF75 foi escolhido porque foi o experimento que mais aproximou a média da precipitação sobre a América do Sul, em relação ao dados do GPCPv1.2 (Tabela 5.2).

A seguir, são apresentados os resultados das simulações dos experimentos realizados com o sistema híbrido 3DVar (na versão EnKF75) para o episódio de ZCAS ocorrido no período PI. Para o episódio destacado, são verificadas nas análises dos experimentos as características principais da circulação atmosférica sobre o continente Sul-Americano que deram suporte dinâmico para o desenvolvimento e a manutenção dos episódios de ZCAS. Para que seja possível fazer comparações com as representações dos fenômenos, foram consideradas também as análises dos experimentos REF e 3DVar puro.

Episódio de ZCAS de 09 a 14 de Janeiro de 2013

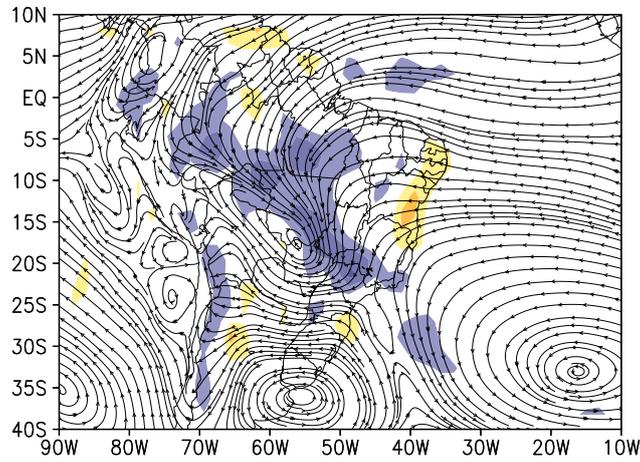
No primeiro episódio de ZCAS em Janeiro de 2013, duas situações dinâmicas na atmosfera ocorreram para que a banda de umidade característica da zona de convergência se formasse: o escoamento de uma circulação ciclônica na baixa troposfera no interior do continente Sul-Americano e um vórtice ciclônico na alta troposfera sobre o Oceano Atlântico.

As Figuras 5.12, 5.13 e 5.14 apresentam as condições simuladas sobre o continente Sul-Americano, dos padrões sinóticos do tempo nos níveis de 850, 200 e 500 hPa. O desenvolvimento do episódio das ZCAS se deu a partir da atuação da Alta da Bolívia em altos níveis, que em associação com a circulação promovida pelo Cavado do Nordeste, provocou divergência de massa neste nível. Na Figura 5.13, nota-se que todos os três experimentos considerados, representaram a atuação destes dois sistemas. Na mesma figura, em sombreado, está representada a divergência do vento. Na Figura 5.14(b), nota-se que o experimento 3DVar apresentou um alto valor de divergência do vento sobre a região sul da Amazônia. Em baixos níveis, sobre esta região, o experimento 3DVar apresenta uma região de convergência de massa (Figura 5.13(b)), ligeiramente mais intensa do que aquela apresenta pelo experimento EnKF75 (Figura 5.13(c)). Entretanto, a Figura 5.13(a), mostra que o experimento REF, também representou nesta região, convergência de massa um pouco mais intensa. Apesar disso, a região de maior divergência em altos níveis não foi representada pelo experimento REF (Figura 5.14(a)), tal como foi representada pelo experimento 3Dvar (puro).

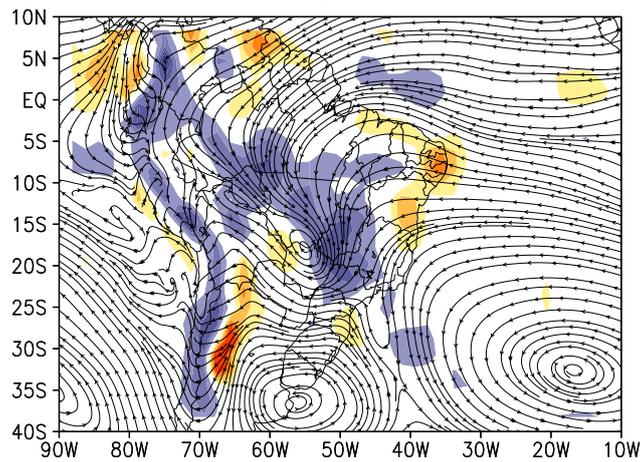
No caso do experimento EnKF75, tanto a convergência de massa em baixos níveis (Figura 5.13(c)), quanto a divergência do vento em altos níveis, foram representadas de forma razoável (Figura 5.14(c)).

Figura 5.12 - Médias da divergência de umidade (D_q - sombreado, $\times 10^{-7} Kgs^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 850 hPa para o período de 09 a 14 de Janeiro de 2013, dos experimentos REF, 3DVar e EnKF75.

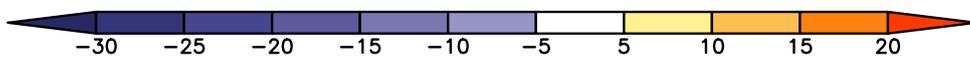
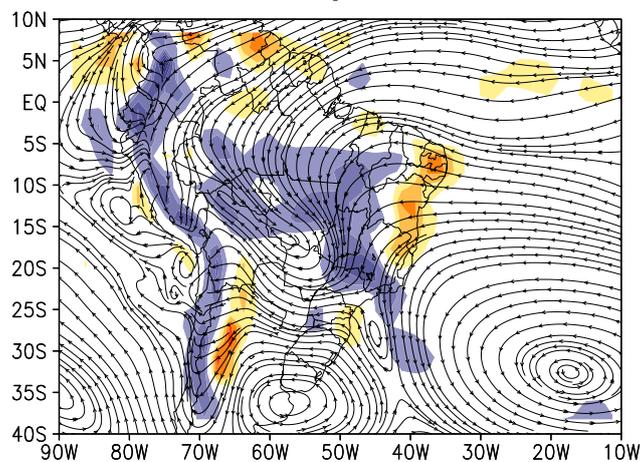
(a) REF, D_q e w , 850 hPa



(b) 3DVar, D_q e w , 850 hPa



(c) EnKF75, D_q e w , 850 hPa



Fonte: Produção do autor.

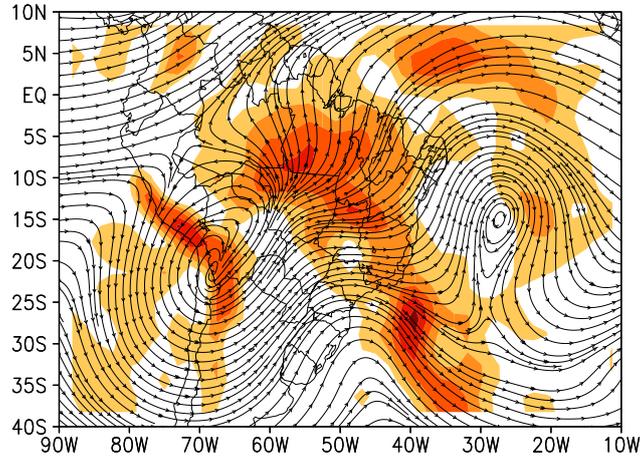
Além das circulações associadas a Alta da Bolívia e ao Cavado do Nordeste, representados pelos experimentos na Figura 5.13, nota-se também a atuação de um cavado ao sul do continente. Este cavado também está representado nos três experimentos e contribuiu para o levantamento de massa na região, dando suporte termodinâmico para a ZCAS.

Em níveis médios, a Figura 5.14 apresenta o campo de velocidade vertical em 500 hPa, e vento (linhas de corrente). Em termos da circulação, as análises dos experimentos representaram um cavado ao sul do continente, sendo este um reflexo do cavado observado em 200 hPa (Figura 5.13). Este cavado observado em 200 e 500 hPa, foi um dos mecanismos que ajudaram a manter a ZCAS ativa.

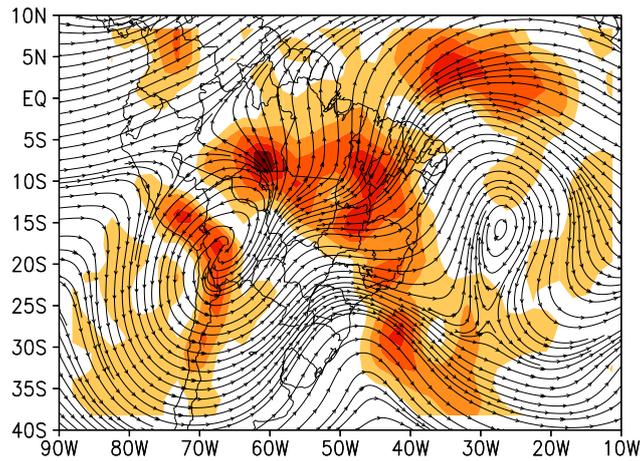
Em altos níveis, os experimentos identificaram um padrão de divergência do vento que se refletiu em convergência de massa em baixos níveis. Na média troposfera, o campo de velocidade vertical mostra bastante atividade ao longo da banda de orientação NO-SE associada a ZCAS. Neste caso, as diferenças entre os experimentos foram mais sensíveis. No caso do experimento REF (Figura 5.15(a)), nota-se maior intensidade do campo de velocidade vertical do que nos demais experimentos. Além disso, observa-se também, atividade sobre a região dos Andes, o que não é representado da mesma maneira pelos demais experimentos. O experimento EnKF75 (Figura 5.15(c)), foi o que representou o campo de velocidade vertical de forma melhor distribuída em relação a orientação da banda de nebulosidade das ZCAS e em comparação com a imagem de satélite da Figura 5.10(d).

Figura 5.13 - Idem Figura 5.12, mas para a divergência do vento horizontal (D_w - sombreado, $\times 10^{-5} s^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 200 hPa.

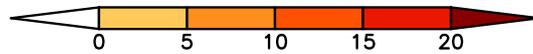
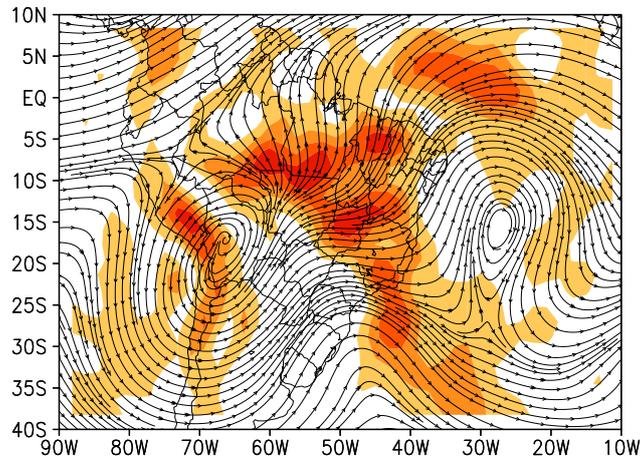
(a) REF, D_w e w , 200 hPa



(b) 3DVar, D_w e w , 200 hPa



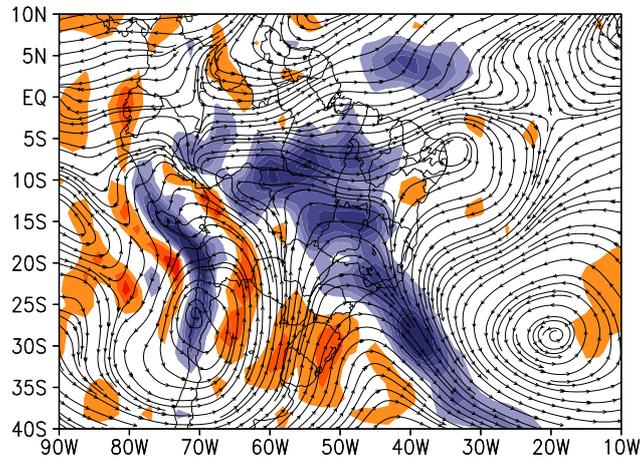
(c) EnKF75, D_w e w , 200 hPa



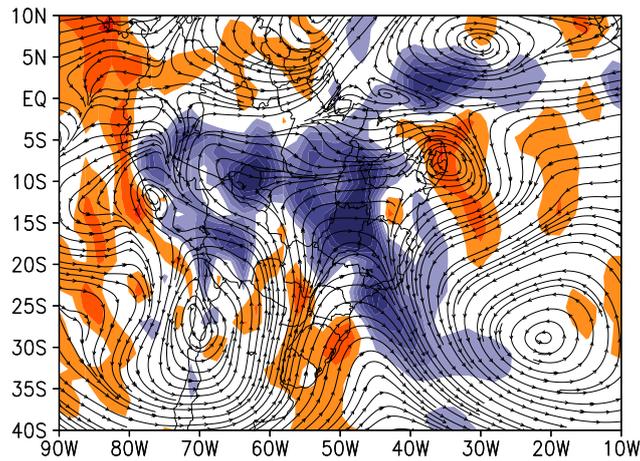
Fonte: Produção do autor.

Figura 5.14 - Idem Figura 5.12, mas para a velocidade vertical (Ω - sombreado, x $10^{-1} Pa s^{-1}$) e vento horizontal (w - linhas de corrente, ms^{-1}) em 500 hPa.

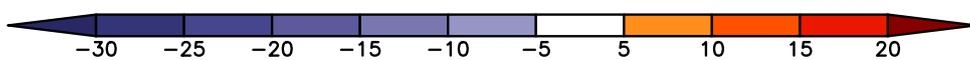
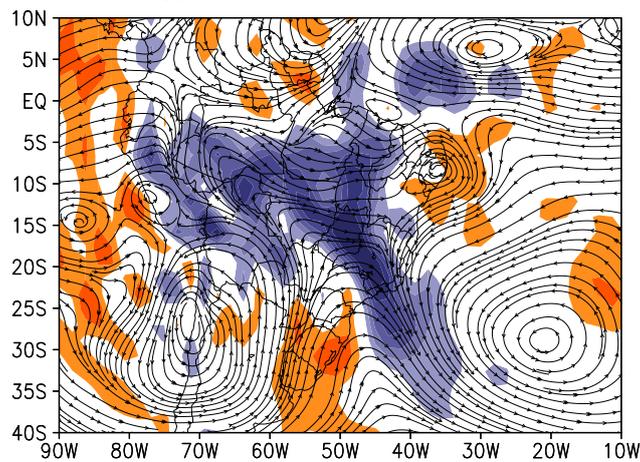
(a) REF, Ω e w , 500 hPa



(b) 3DVar, Ω e w , 500 hPa



(c) EnKF75, Ω e w , 500 hPa



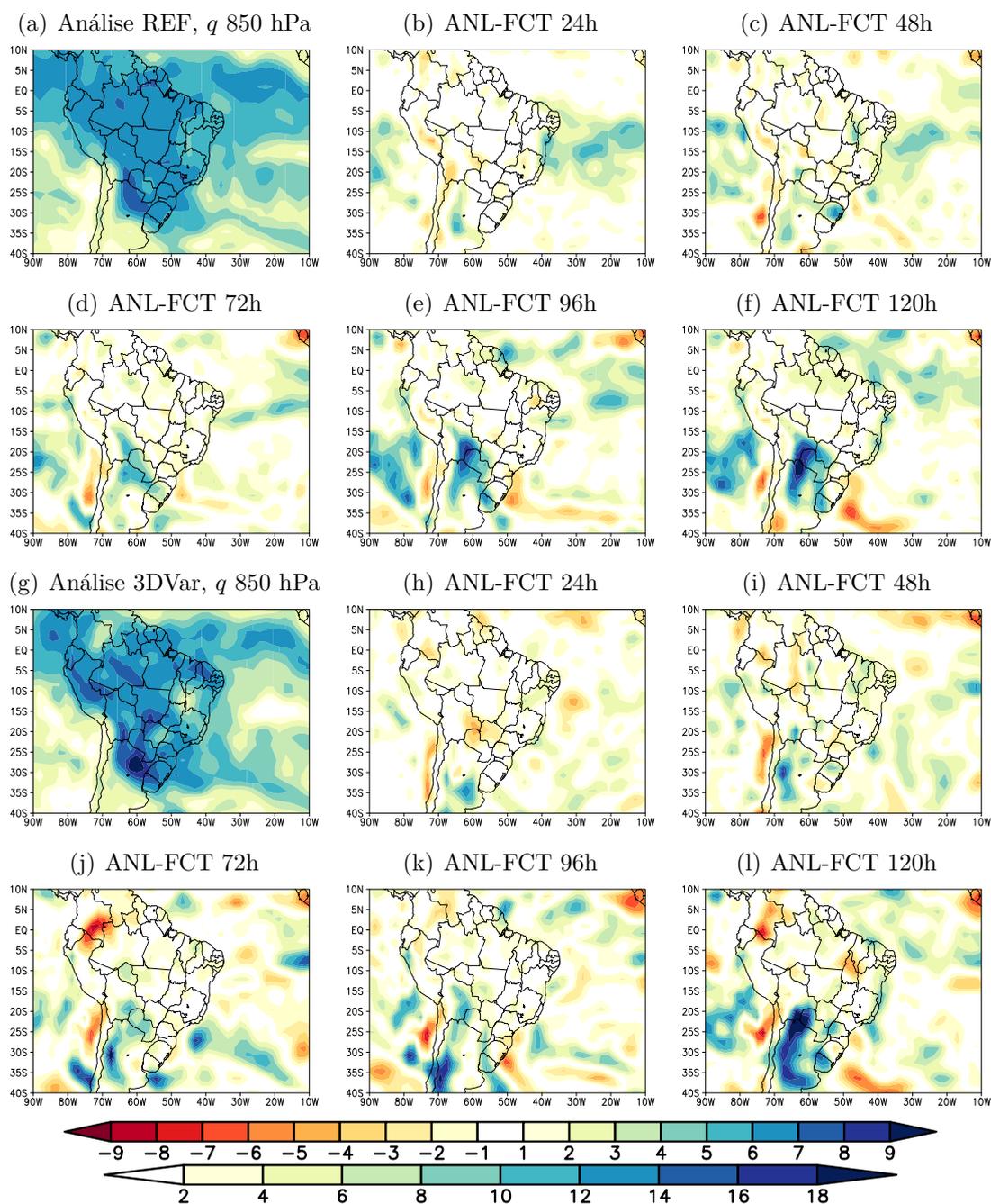
Fonte: Produção do autor.

Com a finalidade de se verificar o erro associado as previsões produzidas pelos experimentos, as Figuras 5.15 e 5.16 apresentam as diferenças entre as análises e as previsões para até 120 horas dos experimentos REF, 3DVar e EnKF75, para os dias 12 de Janeiro, dia em que a ZCAS esteve mais ativa. As Figuras 5.16(a), 5.16(g) e 5.17(a) apresentam os campos de umidade em 850 hPa produzido pelos experimentos no dia 2013011200. Em comparação com os demais experimentos, o campo de umidade do experimento REF (Figura 5.15) apresentam maior distribuição da umidade sobre a região continental da América do Sul e também sobre o Oceano Atlântico. Sobre o norte da Argentina, observam-se valores acumulados que nos experimentos 3DVar (Figura 5.16(g)) e EnFK75 (Figura 5.17(a)) são mais evidentes. Na média entre os dias 9 e 14 de Janeiro, na Figura 5.15(a) notam-se valores positivos de velocidade vertical nesta região para o experimento REF. Nas Figuras 5.15(b) e 5.15(c), entretanto, o sinal associado com a subsidência de massa não é tão proeminente.

As diferenças entre a análise e as previsões do experimento REF (Figuras 5.16(b), 5.16(c), 5.16(d), 5.16(e) e 5.16(f)) o erro associado ao desvio da previsão do modelo BAMv0 em relação a análise do NCEP, concentrou-se mais sobre a região do Oceano Atlântico e na região central da Argentina com valores positivos indicando subestimativa da previsão em relação aos valores analisados. Em 48 horas de previsão, o erro da previsão da umidade aparece também em outras regiões, principalmente sobre o estado do Rio Grande do Sul e a partir da previsão de 72 horas (Figura 5.16(d)) até 120 horas (Figura 5.16(f)), a subestimativa é maior concentrada - sobre a região continental, principalmente sobre a região norte da Argentina, sul da Bolívia e Uruguai.

No caso do experimento 3DVar, as diferenças entre a análise e as previsões mostram que para 24 horas (Figura 5.16(h)), a subestimativa apresentada pela previsão a partir da análise do NCEP, é posicionada mais ao sul da Argentina, enquanto que o modelo BAMv0 tende também a localizar a região entre o Mato Grosso do Sul e o Paraguai, como sendo mais úmida. Em 48 horas de previsão (Figura 5.16(i)), o modelo BAMv0 deslocou a região com subestimativa de umidade para o oeste da Argentina. Esta situação ocorreu também com a análise do NCEP (Figura 5.16(c)), porém com menor intensidade. A partir da previsão de 72 horas, o modelo BAMv0 com a análise do 3DVar (puro), a subestimativa da umidade nas previsões se intensifica até que em 120 horas, o erro é maior ($\geq 9 \times 10^5 gKg^{-1}$) e espacialmente mais concentrado do que o que foi apresentado a partir da análise do NCEP. Especula-se que este erro possa estar relacionado com a representação da topografia da região, considerando a proximidade com a cordilheira dos Andes e a baixa resolução dos experimentos.

Figura 5.15 - Análises dos experimentos REF e 3DVar para a umidade em 850 hPa no dia 2013011200 e diferenças entre as análises e as suas respectivas previsões para até 120 horas.

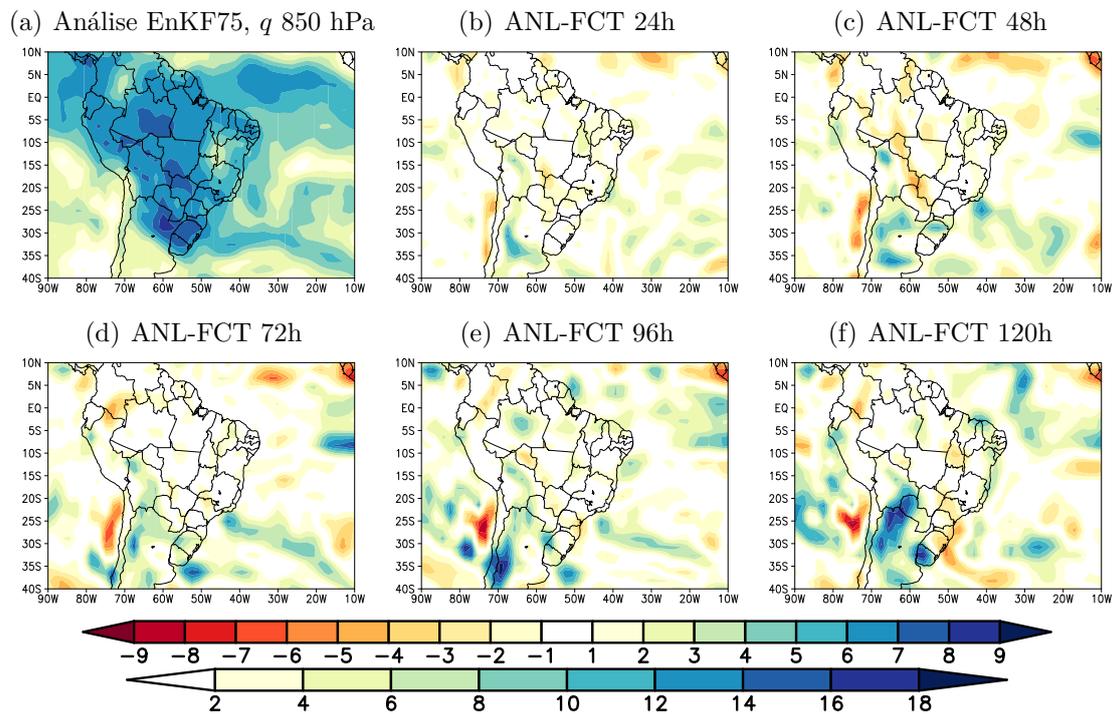


Entre as Figuras (a) e (f), estão representadas as análises e diferenças em relação ao experimento REF; entre as Figuras (g) e (l), estão representadas as análises e diferenças em relação ao experimento 3DVar puro. A primeira barra de cores (no intervalo de -9 a 9), refere-se as diferenças entre as análises e as previsões; a segunda barra de cores (no intervalo de 2 a 18), refere-se as análises. A umidade está representada na unidade $10^5 gK g^{-1}$.

Fonte: Produção do autor.

A Figura 5.16 mostra os campos de diferenças entre a análise do EnKF75 e as previsões até 120 horas. Na previsão de 24 horas, o sinal da subestimativa do modelo BAMv0 apresentado nas Figuras 5.16(b) e 5.16(h), também é apresentado quando o modelo é realizado a partir da análise do experimento EnKF75. A diferença está na posição do sinal que aparece sobre a região central da Argentina. Na previsão de 48 horas (Figura 5.17(c)), o sinal de subestimativa da umidade se amplifica sobre a região da Argentina e é um pouco mais intenso do que foi encontrado com a análise do NCEP. Para a previsão de 72 horas (Figura 5.17(d)), entretanto, a análise do EnKF75 permitiu o modelo BAMv0 prever o campo de umidade com menor intensidade do erro, e seguindo o que foi encontrado no campo de previsão de 96 horas (Figura 5.17(f)) com a análise do 3DVar (Figura 5.16(k)), o erro em comparação com a análise do NCEP é também menor. Em 120 horas de previsão (Figura 5.17(f)) o erro da previsão do campo de umidade a partir da análise do EnKF75 foi menor em relação as outras duas análises.

Figura 5.16 - Idem Figura 5.15, mas para a análise do experimento EnKF75.



Fonte: Produção do autor.

6 CONCLUSÕES

A matriz de covariâncias dos erros de previsão é um das mais importantes componentes de um sistema de assimilação de dados. Os últimos desenvolvimentos da área demonstram a preocupação na forma como estes erros são especificados e aplicados dentro dos sistemas de assimilação de dados. Novas técnicas tem surgido, inclusive, possibilitando que a matriz estática estacionária seja combinada com uma matriz evolutiva no tempo, de forma que ambas as informações se complementem e que os erros do dia possam ser contabilizados pelos sistemas ao se calcular a análise dos modelos.

Neste trabalho foram estudados os efeitos da aplicação de uma nova matriz de covariâncias dos erros de previsão em combinação com covariâncias oriundas de um filtro de Kalman por conjunto. Esta combinação foi aplicada em um sistema de assimilação de dados com bases operacionais, e foi executado por um período de dois meses em um experimento cíclico de assimilação de dados. O sistema construído para esta aplicação, corresponde a um sistema híbrido 3DVar. A característica híbrida desse sistema corresponde as combinações lineares possíveis entre as covariâncias consideradas. A estrutura variacional do sistema GSI foi utilizada para permitir que as covariâncias de um filtro de Kalman por conjunto pudessem ser inseridas através de uma extensão da variável de controle variacional. Essa extensão, além de ter a característica de levar os estados do conjunto de previsões para os estados determinísticos, tem também a importante propriedade de localizar as covariâncias atualizadas.

No Capítulo 1, foi feita a introdução ao problema principal da assimilação de dados, que é a determinação das covariâncias dos erros de previsão. Foram apresentados os fatos científicos que possibilitaram o desenvolvimento da previsão numérica de tempo, bem como dos métodos utilizados atualmente para a assimilação de observações meteorológicas. A motivação principal para a realização deste trabalho foi também apresentada e justificada. O CPTEC é o maior centro de previsão numérica do tempo sobre o Hemisfério Sul. A sua missão é a de prover o país com as melhores previsões numéricas de tempo sobre a América do Sul e sobre a região Tropical. Durante algum tempo, o CPTEC operou um sistema de assimilação de dados utilizando uma matriz de covariâncias estática, proveniente de um outro centro, calculada com base em um outro modelo de previsão numérica. Isto sempre representou uma limitação para o desenvolvimento da assimilação de dados no centro, mas é apenas parte do problema. A medida em que os desenvolvimentos com a assimilação de dados

começam a serem impulsionadas, outros aspectos do problema também podem ser abordados. Um exemplo disso, é o controle de qualidade das observações, a qual poderá se beneficiar com previsões numéricas mais consistentes, feitas a partir de uma análise mais consistente e oriunda de um sistema de assimilação de dados também mais consistente. As bases que sustentam o argumento principal desta tese, a de que as covariâncias de um filtro de Kalman por conjunto são importantes fontes de informações sobre as variações do fluxo atmosférico, puderam ser testadas com os experimentos realizados.

No Capítulo 2, inicia-se a descrição das componentes da metodologia. São apresentadas as descrições do modelo de circulação geral e as suas principais opções, o sistema de assimilação de dados GSI e os dados de observação utilizados nos experimentos com assimilação. Os Capítulos 3 e 4, fazem parte da metodologia do trabalho, e foram destacados devido a sua importância. Estes capítulos trazem informações novas acerca da natureza da matriz de covariâncias. Foram descritos em detalhes a sua importância, a metodologia para o seu cálculo, as suas estruturas principais e o aspecto do incremento de análise da matriz calculada. Através de um exemplo simples, foi mostrado que o incremento de análise variacional é proporcional as colunas da matriz de covariâncias. Logo, uma matriz de covariâncias calculada com base nas previsões do próprio modelo, se for bem ajustada, é um passo importante para a determinação de boas análises e, conseqüentemente, boas previsões. Outras metodologias podem também ser utilizadas para o cálculo de covariâncias e estas podem ser testadas no futuro, embora o método NMC pareça ser o método de escolha da maioria dos centros de previsão numérica de tempo, devido a sua simplicidade. Por outro lado, o cálculo de uma matriz de covariâncias é algo bastante específico e inerente ao sistema de assimilação de dados empregado e também a estrutura de assimilação de dados disponível. Em relação aos sistemas de assimilação de dados por conjunto, neste trabalho foram testados dois tipos de filtro de Kalman por conjunto (EnKF e EnSRF). Foram apresentadas as suas formulações básicas e foram comparadas as formas como ambos calculam as perturbações na atualização das covariâncias.

No Capítulo 4, foi apresentada a estrutura variacional disponível para a incorporação das covariâncias do conjunto e como a sua combinação com as covariâncias estáticas são feitas. O estabelecimento de um ciclo de assimilação de dados, aproveitando a estrutura previamente estabelecida do sistema G3DVAR, permitiu que o sistema híbrido 3DVar pudesse ser exercitado. A principal vantagem desta implementação está no fato de que o sistema, apesar das limitações computacionais, pode ser utilizado em

diversos tipos de estudos, pois utiliza um modelo de previsão de tempo operacional em baixa resolução (para o qual já se tem a estrutura necessária para calcular a matriz de covariâncias), observações meteorológicas provenientes de um fluxo contínuo e, portanto, também operacional. A grande limitação, entretanto, está na capacidade computacional que o centro atualmente possui. A resolução escolhida para a realização dos experimentos reflete não apenas os aspectos do conjunto de previsões em si, i.e, um conjunto de previsões com 40 membros (que pode ser considerado um conjunto grande, dependendo da resolução escolhida), mas também das limitações computacionais envolvidas no processo de desenvolvimento deste trabalho. Seria muito importante validar os resultados encontrados para outras épocas do ano e com resoluções maiores, focando mais os sistemas convectivos típicos dos meses de verão sobre a América do Sul.

No Capítulo 5, são apresentados os resultados obtidos com os experimentos propostos com o sistema híbrido 3DVar. Foram realizados 6 experimentos numéricos, sendo 2 deles experimentos controle (com e sem a assimilação de dados 3DVar pura) e 4 deles com a assimilação de dados híbrida 3DVar. Entre estes experimentos, buscou-se variar a quantidade de contribuição das covariâncias do conjunto as covariâncias estáticas. Como um dos experimentos controle foi com a assimilação de dados 3DVar pura, definiu-se que os demais experimentos deveriam conter as contribuições do conjunto em 50% e 75%. Não foram realizados experimentos com 100% de contribuição das covariâncias por motivo de tempo de processamento e espaço disponível para armazenamento. Um único experimento híbrido (para o período de 2 meses na resolução TQ0062L028) ocupa aproximadamente 6TB, sem contar as previsões de 5 dias que foram realizadas na avaliação dos resultados. Além disso, pôde-se verificar que quanto maior a contribuição do conjunto, não necessariamente será melhor o resultado. Esta equação, depende também de outros fatores, como a configuração do próprio filtro de Kalman por conjuntos. A avaliação das análises e previsões do sistema híbrido foram organizadas de forma a avaliar o conjunto de análises por meio de estatísticas de inovação, a qual compreende uma razão entre o desvio padrão das inovações do conjunto em relação a raiz quadrada do espalhamento total. Neste caso, o espalhamento total foi considerado como sendo a soma do espalhamento do conjunto com o erro da observação. Nesta avaliação, encontrou-se que as análises do EnSRF responderam melhor as configurações padrão de inflação e localização das covariâncias. Isto pôde ser verificado através das diferenças entre os *priors* (previsões) e os *posteriors* (as análises). No caso dos conjuntos de análises do EnKF, as curvas representando estas quantidades estiveram muito próximas entre si ao longo do tempo de simulação.

A avaliação dos experimentos foi feita também em relação a habilidade das previsões de até 5 dias em relação as análises dos experimentos. Nesta avaliação, foi empregada a correlação de anomalia junto com um teste t-Student, a fim de se verificar a significância dos resultados. A partir destes resultados, pôde-se perceber que as análises dos sistema híbrido 3DVar foram, para muitas variáveis e, em várias das regiões avaliadas, superiores as análises do 3DVar puro e NCEP. Vale ressaltar que as análises do sistema 3DVar puro também utilizaram a mesma matriz de covariâncias estática que os experimentos com o método híbrido 3DVar. Os principais resultados obtidos nesta avaliação mostram que sobre a região tropical, o experimento EnKF75 na previsão da pressão em superfície apresentou ganho de 12 horas com destreza de 80%; ganhos expressivos com o experimento EnSRF75 para as previsões da umidade em 925 hPa, em que o experimento 3DVar puro apresenta 80% de destreza com 96 horas de previsão (o experimento EnSRF75, mantém essa destreza para previsões mais longas); o mesmo desempenho foi observado para as previsões da temperatura em 850 hPa, com destreza de 80% em 72 horas de previsão com o experimento EnSRF75. Sobre a América do Sul, os destaques desta avaliação referem-se a umidade em 925 hPa, onde o experimento EnSRF75 apresentou o melhor resultado e a temperatura em 850 hPa, com destreza de 80% para as previsões de 72 horas também com o experimento EnSRF75. Estes resultados mostram ganhos bastante importantes para o modelo de circulação geral do CPTEC, principalmente sobre as variáveis representadas mais próximas a superfície e sobre as regiões de interesse, ou seja, a região tropical e a América do Sul.

Uma avaliação das previsões de 24 horas a partir dos experimentos também foi feita, muito embora seja bastante difícil representar a precipitação convectiva em baixa resolução (TQ0062L028, ~200 km sobre a região equatorial). Nesta avaliação, foram utilizados os dados do GPCP na avaliação das médias temporais para o mês de Janeiro de 2013 e na avaliação das médias espaciais para o mesmo período. Apesar da boa destreza na previsão das variáveis de estado do modelo (como temperatura, pressão e umidade), a representação da precipitação pelos experimentos não foi muito satisfatória. Vários elementos podem contribuir para o desempenho obtido: escolha das parametrizações *Cumulus*, difusões horizontal e vertical, esquemas de radiação e nuvens. Além disso, pode-se especular também sobre a sensibilidade da física do modelo em relação aos ajustes na matriz de covariâncias.

Além da avaliação das previsões de precipitação em 24 horas, escolheu-se um dos experimentos híbridos 3DVar para se fazer um estudo de caso sobre um episódio das ZCAS, em Janeiro de 2013. O experimento escolhido foi o experimento EnKF75

pelo fato de ter representado, na média espacial, o acumulado de precipitação mais próximo do GPCP. Além deste experimento, decidiu-se também incluir os experimentos REF e 3DVar (puro) para comparação, uma vez que estes experimentos são os controles. Neste estudo de caso, identificaram-se 3 casos de ZCAS, conforme documentado no boletim Climanálise, os quais puderam ser identificados em todos os experimentos realizados. Para isso, utilizou-se como critério de identificação da atividade das ZCAS, a componente zonal do vento em 850 hPa. As diferenças entre os experimentos na identificação, foi em relação a intensidade, duração (início e fim) do escoamento do vento seguindo o critério. O caso escolhido foi o primeiro, entre 9 e 14 de Janeiro, para o qual, foram verificadas as condições dinâmicas representadas pelos experimentos REF, 3DVar e EnKF75. Os experimentos 3DVar e EnKF75 não diferiram muito entre si, uma vez que estes experimentos foram realizados com ciclos de assimilação, utilizando o mesmo conjunto de observações e matriz de covariâncias. Em relação ao experimento REF, este mostrou que as previsões de 48 horas do campo de umidade são subestimadas em relação ao que foi encontrado com os experimentos com assimilação de dados. O experimento EnKF75, por outro lado, mostrou potencial para amenizar a subestimativa do campo de umidade em relação aos demais.

Ainda no escopo desta pesquisa, em estudos preliminares a aplicação das covariâncias de um filtro de Kalman por conjunto e sua combinação com covariâncias estáticas, foram verificadas as diferenças entre as matrizes de covariâncias calculadas para os horários sinóticos individuais. Por exemplo, uma matriz de covariâncias calculada apenas com os pares de previsões de 48 e 24 horas (utilizando-se o método NMC) válidos para as 00, 06, 12 e 18Z. Estas matrizes individuais foram testadas em experimentos de assimilação de dados de duas formas distintas: realizando-se os ciclos de assimilação com cada matriz e realizando-se ciclos de assimilação em que as matrizes de covariâncias são alternadas, cada uma sendo utilizada em seu respectivo horário sinótico. Para esta verificação, encontrou-se que a alternância dos pares de previsões, em relação a matriz completa pode, em alguns casos, permitir que mais observações sejam assimiladas em cada ciclo. Este resultado está relacionado ao fato de que uma matriz de covariâncias própria permite que a análise utilize a informação de referência da superfície a partir da previsão de curto prazo para aplicar o operador observação e calcular o perfil do modelo relativo aos perfis das radiâncias a serem assimiladas. Neste caso, as diferenças na quantidade total de dados assimilados está justamente no aumento de dados de radiâncias assimilados.

Buscou-se verificar também, o condicionamento das previsões para até 15 dias a partir dos conjuntos de análises realizados com os 40 membros, utilizados nos experimentos

híbridos 3DVar. Este experimento não pôde ser trazido para o escopo da tese por duas razões: 1) os resultados de destreza das previsões em termos do “Continuous Rank Probability Skill Score” (CRPSS) não foram satisfatórios em relação ao que se já tem com o Sistema de Previsão por Conjuntos (SPCON) do CPTEC, ou seja, não houve vantagem em se realizar o sistema híbrido para esta finalidade (previsão estendida para até 15 dias), na forma como ele está configurado; 2) o sistema híbrido 3DVar não é necessariamente desenhado para previsões até 15 dias, mas para previsões numéricas até 5 ou 7 dias. Isto é consequência da finalidade a que o sistema híbrido 3DVar se presta, ou seja, assimilação de dados. Isso quer dizer que previsões de curto prazo são utilizadas para atualização e então são integradas por um modelo de previsão de tempo configurado para prever com uma determinada destreza para até 5 ou 7 dias e que as perturbações realizadas, mediante as configurações de inflação e localização, não são adequadas para previsões de mais longo prazo. A aplicação de um sistema híbrido para a melhoria do sistema de previsões do conjunto do CPTEC, irá requerer uma reconfiguração do sistema para que este possa ser utilizado, por exemplo, em dupla resolução. Neste caso, uma resolução mais alta poderá ser utilizada para as previsões até 5 dias e utilizando-se a análise híbrida variacional e, a partir disso, uma outra resolução (mais baixa) para as previsões estendidas em conjunto com uma configuração mais adequada do modelo poderá ser utilizada para esta finalidade.

Este trabalho experimentou uma nova matriz de covariâncias dos erros de previsão, aplicada a uma versão modificada do sistema global de assimilação de dados do CPTEC. A realização deste sistema se apresenta como um resultado positivo para o CPTEC e o desenvolvimento de suas rotinas. Mais testes de validação precisam ser feitos, mas os resultados aqui apresentados são promissores.

Diante dos objetivos iniciais proposto para o desenvolvimento desta tese, fez-se primeiro o seguinte questionamento: **qual é a contribuição das covariâncias do filtro de Kalman por conjunto na determinação das análises e previsões de um sistema de assimilação de dados global?**

A resposta para esta questão está nos resultados encontrados em que as análises que utilizaram as covariâncias do filtro de Kalman por conjunto, permitiram que o modelo de circulação geral gerasse previsões para algumas variáveis e em algumas regiões, mais bem representadas do que aquelas que consideram apenas as covariâncias estáticas. Isto mostra que se as correlações espaciais dos erros do modelo são semelhantes as correlações entre diferentes previsões (próximas, assim como preconiza o método NMC), então podemos afirmar também que as correlações entre previsões

diferentes, - ou seja, os membros de um mesmo conjunto - são também semelhantes e representativas do que chamamos de erros do dia. Logo, o uso das covariâncias provenientes de um filtro de Kalman por conjunto, ajudaram a melhorar a qualidade das análises e previsões através da representação nas análises, das variações espaço-temporais do fluxo atmosférico.

Em relação aos objetivos específicos elencados para o desenvolvimento da tese, foi possível calcular uma nova matriz de covariâncias, a partir da qual foram identificadas e determinadas as estruturas principais e qual o seu impacto nas previsões do modelo de previsão. O sistema GSI foi habilitado para ler as covariâncias de um filtro de Kalman por conjunto, a partir do qual foi estabelecida uma rotina cíclica para que este sistema pudesse combinar as covariâncias estáticas previamente calculadas e as dinâmicas, derivadas do filtro de Kalman por conjunto, com diferentes porcentagens de contribuições. Os efeitos desta atualização foram investigados nas análises e previsões do sistema. Este trabalho mostrou, do ponto de vista teórico e prático, que a matriz de covariâncias dos erros de previsão é uma componente fundamental e determinante do processo de assimilação de dados. A sua melhor representação dentro do sistema de assimilação de dados, tem um impacto positivo e incrementa a qualidade das análises e conseqüentemente, das previsões.

6.1 Sugestões para Trabalhos Futuros

Durante o desenvolvimento deste trabalho, diversos aspectos do sistema híbrido 3DVar não puderam ser completamente testados e acessados. Estes aspectos já tem sido trabalhos em outros centros operacionais e permitem que outras funcionalidades dos sistema sejam acessadas. Além disso, por questões práticas, não foi possível realizar o sistema em resolução mais alta. Inicialmente tinha-se como objetivo realizar o sistema híbrido para a resolução TQ0299L064 (~45 km), com uma nova matriz de covariâncias. Porém, os testes iniciais mostraram que o custo computacional é muito alto.

Dentre os aspectos a serem abordados em trabalhos futuros com o sistema híbrido 3DVar, destacam-se:

- Investigar o impacto do procedimento de re-centralização do conjunto de previsões em torno da média do conjunto;
- O uso de dupla resolução para previsões mais longas ou em mais alta resolução (neste caso, realiza-se o conjunto de análises do filtro de Kalman

por conjunto com mais membros, mas em menor resolução);

- Realizar testes de sensibilidade com outros valores de inflação e localização das covariâncias;
- Determinar o tamanho ideal do conjunto de previsões em relação a resolução escolhida em aplicação para previsão de tempo de 5 a 15 dias;
- Investigar as relações entre o papel da localização do filtro de Kalman por conjunto e os comprimentos de escala nos incrementos de análise do GSI;
- Investigar a manutenção do espalhamento do conjunto de análises dentro do sistema híbrido;
- Comparar as análises híbridas 3DVar com um esquema em que a matriz de covariâncias é atualizada a partir de uma média móvel no tempo;
- Testar outras metodologias para o cálculo da matriz de covariâncias;
- Verificar a sensibilidade da aplicação da matriz de covariâncias mediante a modificações na física e dinâmica do modelo de previsão, a partir do qual serão gerados os pares de previsão;
- Validar o sistema híbrido 3DVar para outras épocas do ano, na simulação de fenômenos meteorológicos com estruturas fisicamente dependentes (e.g., frentes frias);
- Estabelecer uma técnica mais robusta para a avaliação das inovações do conjunto, como a utilização de um sistema de pontuação (*score*);

REFERÊNCIAS BIBLIOGRÁFICAS

ADLER, R. F.; HUFFMAN, G. J.; CHANG, A.; FERRARO, R.; XIE, P.-P.; JANOWIAK, J.; RUDOLF, B.; SCHNEIDER, U.; CURTIS, S.; BOLVIN, D.; GRUBER, A.; SUSSKIND, J.; ARKIN, P.; NELKIN, E. The version-2 global precipitation climatology project (gpcp) monthly precipitation Analysis (1979–Present). **Journal of Hydrometeorology**, v. 4, n. 6, p. 1147–1167, 2003.

Disponível em:

<[http://dx.doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2)>.

91

AKKRAOUI, A. E.; TODLING, R. **Hybrid data assimilation without ensemble filtering**. Joint Center for Satellite Data Assimilation, 2013. Disponível em:

<<http://www.jcsda.noaa.gov/documents/seminardocs/2013/todling20130116.Pdf>>.

[//www.jcsda.noaa.gov/documents/seminardocs/2013/todling20130116.Pdf](http://www.jcsda.noaa.gov/documents/seminardocs/2013/todling20130116.Pdf)>.

12

ARAUJO, L. R. P. **Os efeitos das parametrizações do comprimento de rugosidade térmica na simulação da temperatura de superfície terrestre no modelo SSiB do G3DVar**. 136 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2015-10-01 2015.

Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m21b/2015/09.04.15.21>>. Acesso em: 17 jun. 2017. 4

ARAVÉQUIA, J. A. A assimilação de radiâncias no filtro de kalman local por conjunto. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 15., 24-29 ago, São Paulo. 2008. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2008/07.18.13.38>>. Acesso em: 18 nov. 2013. 4

ARRIBAS, A.; ROBERTSON, K. B.; MYLNE, K. R. Test of a poor man's ensemble prediction system for short-range probability forecasting. **Monthly Weather Review**, v. 133, n. 7, p. 1825–1839, 2005. Disponível em:

<<https://doi.org/10.1175/MWR2911.1>>. 67

AVANÇO, L. A.; SAPUCCI, L. F.; SOUZA, D. S. V. d.; DIAS, P. H. D.; REIS, M. A. d. **Assimilação de dados de rádio ocultação GNSS no LETKF: disponibilidade de dados e implementação de um operador**. São José dos Campos: [s.n.], 2013. 92 p. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m18@80/2013/07.18.13.38>>. Acesso em: 18 nov. 2013. 4

<<http://urlib.net/sid.inpe.br/mtc-m19/2013/11.06.18.49>>. Acesso em: 13 set. 2017. 4

AZEVEDO, H. B. d. **Uso da técnica de OSE para avaliar o impacto de observações no sistema de assimilação de dados global variacional tridimensional do CPTEC/INPE.** 142 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2014-05-16 2014. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m18/2014/05.07.17.54>>. Acesso em: 17 jun. 2017. 4

BERGTHÓRSSON, P.; DÖÖS, B. R. Numerical weather map analysis. **Tellus**, Blackwell Publishing Ltd, v. 7, n. 3, p. 329–340, 1955. ISSN 2153–3490. Disponível em: <<http://dx.doi.org/10.1111/j.2153-3490.1955.tb01170.x>>. 8, 9

BONATTI, J. P. **Modelo de circulação geral atmosférico do CPTEC.** Instituto Nacional de Pesquisas Espaciais, 1996. Disponível em: <<http://climanalise.cptec.inpe.br/~rclimanl/boletim/cliEsp10a/bonatti.html>>. 22

BROWN, A.; JAKOB, C. Report of the joint scientific committee and the wmo commission for atmospheric sciences. In: THE 27TH SESSION OF THE CAS/JSC WORKING GROUP ON NUMERICAL EXPERIMENTATION., 17-21 out., Boulder. 2011. Disponível em: <http://www.wcrp-climate.org/documents/wgne-27_report_final.pdf>. Acesso em: 15 set. 2017. 5

BUEHNER, M. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational nwp setting. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd., v. 131, n. 607, p. 1013–1043, 2005. ISSN 1477–870X. Disponível em: <<http://dx.doi.org/10.1256/qj.04.15>>. 12, 14

CAVALCANTI, I. F. A.; SETZER, A. W.; QUADRO, A. F. L. de; MELO, A. B. C. de; BOTTINO, M. J.; CARPENEDO, C. B.; CALBETE, N. O. de; LEMOS, C. F.; SATYAMURTY, P.; AQUINO, F. E.; SIMANOGLU, R. A.; MACHADO, L. H. R.; CARAM, R. O. **Boletim de Monitoramento e Análise Climática para Janeiro de 2013.** [s.n.], 2013. (Boletim Climanálise, 1). ISSN 0103-0019 CDU-555. Disponível em: <<http://climanalise.cptec.inpe.br/~rclimanl/boletim/>>. 98

CAVALCANTI, I. F. d. A.; MARENGO, J. A.; PRAKKI, S.; NOBRE, C. A.; TROSNIKOV, I.; BONATTI, J. P.; MANZI, A. O.; TARASOVA, T.; PEZZI, L. P.; D'ALMEIDA, C.; SAMPAIO, G.; CASTRO, C. A. C.; SANCHES, M. B.; CAMARGO, H. Global climatological features in a simulation using the cptec-cola agcm. **Journal of Climate**, v. 15, n. 21, p. 2965–2988, Nov. 2002. ISSN 0894–8755. Disponível em:

<<http://urlib.net/sid.inpe.br/iris@1915/2005/05.05.11.19>>. Acesso em: 18 nov. 2013. 21

CHEPURIN, G. A.; CARTON, J. A.; DEE, D. Forecast model bias correction in ocean data assimilation. **Monthly Weather Review**, v. 133, n. 5, p. 1328–1342, 2005. Disponível em: <<http://dx.doi.org/10.1175/MWR2920.1>>. 63

CHOU, M. **A solar radiation parameterization for atmospheric studies**. [s.n.], 1999. (NASA technical memorandum). Disponível em:

<https://books.google.com.br/books?id=_gYbvgAACAAJ>. 22

CINTRA, R. S. C. **Assimilação de dados com redes neurais artificiais em modelo de circulação geral da atmosfera**. 200 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010-09-27 2010. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2010/09.20.14.46>>. Acesso em: 17 jun. 2017. 4

CLAYTON, A. M.; LORENC, A. C.; BARKER, D. M. Operational implementation of a hybrid ensemble/4D–var global data assimilation system at the met office. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd., v. 139, n. 675, p. 1445–1461, 2012. ISSN 1477–870X. Disponível em: <<http://dx.doi.org/10.1002/qj.2054>>. 12, 15, 18, 58

CORAZZA, M.; KALNAY, E.; PATIL, D. J.; YANG, S.-C.; MORSS, R.; CAI, M.; SZUNYOGH, I.; HUNT, B. R.; YORKE, J. A. Use of the breeding technique to estimate the structure of the analysis “errors of the day”. **Nonlinear Processes in Geophysics**, v. 10, n. 3, p. 233–243, 2003. Disponível em: <<http://www.nonlin-processes-geophys.net/10/233/2003/>>. 16, 18

CRESSMAN, G. P. An operational objective analysis system. **Monthly Weather Review**, v. 87, n. 10, p. 367–374, 1959. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(1959\)087<0367:A00AS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1959)087<0367:A00AS>2.0.CO;2)>. 9

DAHLQUIST, G.; BJÖRCK, Å. **Numerical methods**. [S.l.]: Dover Publications, 1974. (Dover Books on Mathematics). ISBN 9780486428079. 147

DALEY, R. **Atmospheric Data Analysis**. [S.l.]: Cambridge University Press, 1996. (Cambridge Atmospheric and Space Science Series). ISBN 9780521458252. 8

DESROZIERS, G. A coordinate change for data assimilation in spherical geometry of frontal structures. **Monthly Weather Review**, v. 125, n. 11, p. 3030–3038, 1997. Disponível em: [http://dx.doi.org/10.1175/1520-0493\(1997\)125<3030:ACCFDA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1997)125<3030:ACCFDA>2.0.CO;2). 17

DINIZ, F. L. R. **Impacto das observações nas previsões de curto prazo**. 138 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012-12-14 2012. Disponível em: <http://urlib.net/sid.inpe.br/mtc-m19/2012/11.29.01.27>. Acesso em: 18 nov. 2013. 4

EBERT, E. E. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. **Monthly Weather Review**, v. 129, n. 10, p. 2461–2480, 2001. Disponível em: [https://doi.org/10.1175/1520-0493\(2001\)129<2461:A0APMS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<2461:A0APMS>2.0.CO;2). 67

EPSTEIN, E. S. Stochastic-dynamic prediction. **Tellus**, Blackwell Publishing Ltd, v. 21, n. 6, p. 739–759, 1969. ISSN 2153-3490. Disponível em: <http://dx.doi.org/10.1111/j.2153-3490.1969.tb00483.x>. 10

ETHERTON, B. J.; BISHOP, C. H. Resilience of hybrid ensemble/3DVAR analysis schemes to model error and ensemble covariance error. **Monthly Weather Review**, American Meteorological Society, v. 132, n. 5, p. 1065–1080, 2013/11/18 2004. Disponível em: [http://dx.doi.org/10.1175/1520-0493\(2004\)132<1065:ROHDAS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2004)132<1065:ROHDAS>2.0.CO;2). 12, 13, 14

EVENSEN, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. **Journal of Geophysical Research: Oceans**, v. 99, n. C5, p. 10143–10162, 1994. ISSN 2156-2202. Disponível em: <http://dx.doi.org/10.1029/94JC00572>. 11

_____. The ensemble kalman filter: Theoretical formulation and practical implementation. **Ocean Dynamics**, v. 53, n. 4, p. 343–367, 2003. ISSN 1616-7228. Disponível em: <http://dx.doi.org/10.1007/s10236-003-0036-9>. 26, 27

FIGUEROA, S. N.; BONATTI, J. P.; KUBOTA, P. Y.; GRELL, G. A.; MORRISON, H.; BARROS, S. R. M.; FERNANDEZ, J. P. R.; RAMIREZ, E.;

SIQUEIRA, L.; LUZIA, G.; SILVA, J.; SILVA, J. R.; PENDHARKAR, J.; CAPISTRANO, V. B.; ALVIM, D. S.; ENORÉ, D. P.; DINIZ, F. L. R.; SATYAMURTI, P.; CAVALCANTI, I. F. A.; NOBRE, P.; BARBOSA, H. M. J.; MENDES, C. L.; PANETTA, J. The brazilian global atmospheric model (BAM): Performance for tropical rainfall forecasting and sensitivity to convective scheme and horizontal resolution. **Weather and Forecasting**, v. 31, n. 5, p. 1547–1572, 2016. Disponível em: <<http://dx.doi.org/10.1175/WAF-D-16-0062.1>>. 21

FISHER, M. Background error covariance modelling. In: **SEMINAR ON RECENT DEVELOPMENTS IN DATA ASSIMILATION FOR ATMOSPHERE AND OCEAN, 2003, Shinfield Park, Reading. Proceedings...** ECMWF. [S.l.: s.n.], 2003. p. 45–63. 17, 35

GILCHRIST, B.; CRESSMAN, G. P. An experiment in objective analysis. **Tellus**, Blackwell Publishing Ltd, v. 6, n. 4, p. 309–318, 1954. ISSN 2153-3490. Disponível em: <<http://dx.doi.org/10.1111/j.2153-3490.1954.tb01126.x>>. 7

GRELL, G. A. Prognostic evaluation of assumptions used by cumulus parameterizations. **Monthly Weather Review**, v. 121, n. 3, p. 764–787, 1993. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(1993\)121<0764:PE0AUB>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1993)121<0764:PE0AUB>2.0.CO;2)>. 22

GRIFFITH, A. K.; NICHOLS, N. K. **Data assimilation using observers.** [s.n.], 1994. Disponível em: <<https://www.reading.ac.uk/web/files/maths/11-94.pdf>>. Acesso em: 17 Abril 2017. 24

HAMILL, T. M.; SNYDER, C. A hybrid ensemble kalman filter–3D variational analysis scheme. **Monthly Weather Review**, American Meteorological Society, v. 128, n. 8, p. 2905–2919, 2013/11/18 2000. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(2000\)128<2905:AHEKFBV>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2000)128<2905:AHEKFBV>2.0.CO;2)>. 12, 13, 14, 16, 18

HARLIM, J.; HUNT, B. R. **Local ensemble transform Kalman filter: An efficient scheme for assimilating atmospheric data.** 2005. Disponível em: <https://www.atmos.umd.edu/~ekalnay/pubs/harlim_hunt05.pdf>. 27

HARSHVARDHAN; DAVIES, R.; RANDALL, D. A.; CORSETTI, T. G. A fast radiation parameterization for atmospheric circulation models. **Journal of**

Geophysical Research: Atmospheres, v. 92, n. D1, p. 1009–1016, 1987. ISSN 2156-2202. Disponível em: <<http://dx.doi.org/10.1029/JD092iD01p01009>>. 22

HAYDEN, C. M.; PURSER, R. J. Recursive filter objective analysis of meteorological fields: Applications to NESDIS operational processing. **Journal of Applied Meteorology**, v. 34, n. 1, p. 3–15, 1995. Disponível em: <<http://dx.doi.org/10.1175/1520-0450-34.1.3>>. 17

HERDIES, D. L.; SILVA, A. da; DIAS, M. A. F. S.; FERREIRA, R. N. Moisture budget of the bimodal pattern of the summer circulation over south america. **Journal of Geophysical Research: Atmospheres**, v. 107, n. D20, p. LBA 42–1–LBA 42–10, 2002. ISSN 2156-2202. 8075. Disponível em: <<http://dx.doi.org/10.1029/2001JD000997>>. 98

HOFFMAN, M.; PEZZI, L.; HERDIES, D. L.; PENNY, S. The CPTEC ocean data assimilation system–CODAS. In: CONGRESSO BRASILEIRO DE METEOROLOGIA 15., 24-29 ago, São Paulo. 2008. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m18@80/2008/09.24.13.27>>. Acesso em: 18 nov. 2013. 3

HOFFMAN, R. N.; KALNAY, E. Lagged average forecasting, an alternative to monte carlo forecasting. **Tellus A**, Blackwell Publishing Ltd, v. 35A, n. 2, p. 100–118, 1983. ISSN 1600-0870. Disponível em: <<http://dx.doi.org/10.1111/j.1600-0870.1983.tb00189.x>>. 11

HOLLINGSWORTH, A.; LÖNNBERG, P. The statistical structure of short-range forecast errors as determined from radiosonde data. part i: The wind field. **Tellus A**, Blackwell Publishing Ltd, v. 38A, n. 2, p. 111–136, 1986. ISSN 1600-0870. Disponível em: <<http://dx.doi.org/10.1111/j.1600-0870.1986.tb00460.x>>. 35

HOLTSLAG, A. A. M.; BOVILLE, B. A. Local versus nonlocal boundary–layer diffusion in a global climate model. **Journal of Climate**, v. 6, n. 10, p. 1825–1842, 1993. Disponível em: <[http://dx.doi.org/10.1175/1520-0442\(1993\)006<1825:LVNBLD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2)>. 22

HUFFMAN, G. J.; ADLER, R. F.; MORRISSEY, M. M.; BOLVIN, D. T.; CURTIS, S.; JOYCE, R.; MCGAVOCK, B.; SUSSKIND, J. Global precipitation at one-degree daily resolution from multisatellite observations. **Journal of Hydrometeorology**, v. 2, n. 1, p. 36–50, 2001. Disponível em:

<[http://dx.doi.org/10.1175/1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2)>. 94

HUNT, B. R.; KOSTELICH, E. J.; SZUNYOGH, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform kalman filter. **Physica D**, p. 112–126, 2007. 3

KALMAN, R. E.; BUCY, R. S. New results in linear filtering and prediction theory. **ASME. J. Basic Eng.**, v. 83, n. 1, p. 95–108, 1961. 11

KALNAY, E. **Atmospheric modeling, data assimilation and predictability**. Cambridge University Press, 2003. ISBN 9780521796293. Disponível em: <<https://books.google.com.br/books?id=Uqc7zC7NULMC>>. 29

KLEIST, D. T.; PARRISH, D. F.; DERBER, J. C.; TREADON, R.; WU, W.-S.; LORD, S. Introduction of the GSI into the NCEP global data assimilation system. **Weather and Forecasting**, American Meteorological Society, v. 24, n. 6, p. 1691–1705, 2009. Disponível em: <<http://dx.doi.org/10.1175/2009WAF2222201.1>>. 4, 23

KUBOTA, P. Y. **Variabilidade da energia armazenada na superfície e o seu impacto na definição do padrão de precipitação na América do Sul**. 309 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012-08-06 2012. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2012/08.02.02.42>>. Acesso em: 28 nov. 2013. 21

LEI, L.; STAUFFER, D.; HAUPT, S. E.; YOUNG, G. A hybrid nudging-ensemble kalman filter approach to data assimilation. part i: Application in the lorenz system. **Tellus A**, v. 64, n. 0, 2012. ISSN 1600-0870. Disponível em: <<http://www.tellusa.net/index.php/tellusa/article/view/18484>>. 12, 15

LEITH, C. E. Theoretical skill of monte carlo forecasts. **Monthly Weather Review**, v. 102, n. 6, p. 409–418, 1974. Disponível em: <[https://doi.org/10.1175/1520-0493\(1974\)102<0409:TSOMCF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1974)102<0409:TSOMCF>2.0.CO;2)>. 10

LORENC, A. C. Analysis methods for numerical weather prediction. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd, v. 112, n. 474, p. 1177–1194, 1986. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1002/qj.49711247414>>. 18, 25

_____. Modelling of error covariances by 4D-Var data assimilation. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd., v. 129, n. 595, p. 3167–3182, 2003. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1256/qj.02.131>>. 14, 16, 57

MEDEIROS, M. d. D. d. S. **Impacto do uso de radiância na assimilação de dados usando 4D-LETKF na região da América do Sul**. 137 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010-12-06 2011. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2010/11.24.18.42>>. Acesso em: 18 nov. 2013. 4

MELLOR, G. L.; YAMADA, T. A hierarchy of turbulence closure models for planetary boundary layers. **Journal of the Atmospheric Sciences**, v. 31, n. 7, p. 1791–1806, 1974. Disponível em: <[http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2)>. 22

MOLTENI, F. Atmospheric simulations using a GCM with simplified physical parametrizations. i: Model climatology and variability in multi-decadal experiments. **Climate Dynamics**, v. 20, n. 2, p. 175–191, 2003. ISSN 1432-0894. Disponível em: <<http://dx.doi.org/10.1007/s00382-002-0268-2>>. 4

NERGER, L. On serial observation processing in localized ensemble kalman filters. **Monthly Weather Review**, v. 143, n. 5, p. 1554–1567, 2015. Disponível em: <<https://doi.org/10.1175/MWR-D-14-00182.1>>. 32

NICHOLS, N. K. Mathematical concepts of data assimilation. In: LAHOZ, W.; KHATTATOV, B.; MENARD, R. (Ed.). **Data assimilation: making sense of observations**. Springer, 2010. p. 13–40. Disponível em: <<http://centaur.reading.ac.uk/2001/>>. 64

O’KANE, T. J.; FREDERIKSEN, J. S. Comparison of statistical dynamical, square root and ensemble kalman filters. **Entropy**, v. 10, n. 4, p. 684–721, 2008. ISSN 1099-4300. Disponível em: <<http://www.mdpi.com/1099-4300/10/4/684>>. 27, 30

PANOFSKY, R. A. Objective weather–map analysis. **Journal of Meteorology**, v. 6, n. 6, p. 386–392, 1949. Disponível em: <[http://dx.doi.org/10.1175/1520-0469\(1949\)006<0386:OWMA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1949)006<0386:OWMA>2.0.CO;2)>. 7

PARRISH, D. F.; DERBER, J. C. The national meteorological center’s spectral statistical-interpolation analysis system. **Monthly Weather Review**, American

Meteorological Society, v. 120, n. 8, p. 1747–1763, 2013/11/18 1992. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2)>. 35, 37

PENNA, B. R.; SILVEIRA, B. B.; GONÇALVES, L. G. G. D.; SOUZA, S. S. D. Avaliação da temperatura de brilho nos canais sensíveis À superfície terrestre no sistema G3DVAR do CPTEC/INPE: Série NOAA. **Revista Brasileira de Meteorologia**, scielo, v. 30, p. 340–350, 09 2015. ISSN 0102-7786. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-77862015000300340&nrm=iso>. 5

PETRIE, R. E. Background error covariance modelling for convective–scale variational data assimilation. 2012. 40

POTTER, J.; STERN, R. **Statistical filtering of space navigation measurements**. Massachusetts Institute of Technology, Experimental Astronomy Laboratory, 1963. (RE (Massachusetts Institute of Technology. Experimental Astronomy Laboratory)). Disponível em: <https://books.google.com.br/books?id=2_kzHQAACAAJ>. 31

PURSER, R. J.; WU, W.-S.; PARRISH, D. F.; ROBERTS, N. M. Numerical aspects of the application of recursive filters to variational statistical analysis. part i: Spatially homogeneous and isotropic gaussian covariances. **Monthly Weather Review**, v. 131, n. 8, p. 1524–1535, 2003. Disponível em: <[https://doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTA0>2.0.CO;2](https://doi.org/10.1175//1520-0493(2003)131<1524:NAOTA0>2.0.CO;2)>. 62, 147, 148, 149, 150, 153, 157

_____. Numerical aspects of the application of recursive filters to variational statistical analysis. part ii: Spatially inhomogeneous and anisotropic general covariances. **Monthly Weather Review**, American Meteorological Society, v. 131, n. 8, p. 1536–1548, 2013/11/18 2003. Disponível em: <<http://dx.doi.org/10.1175//2543.1>>. 62

RABIER, F. Overview of global data assimilation developments in numerical weather-prediction centres. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd., v. 131, n. 613, p. 3215–3233, 2005. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1256/qj.05.129>>. 17

RIISHØJGAARD, L. P. A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. **Tellus A**, Munksgaard

International Publishers, v. 50, n. 1, p. 42–57, 1998. ISSN 1600-0870. Disponível em: <<http://dx.doi.org/10.1034/j.1600-0870.1998.00004.x>>. 17

RIZVI, S. R.; LIU, Z.; HUANG, X.-Y. Generation of WRF-ARW Background errors (BE) for GSI. 2009. 37

RODRIGUES, C. M. **Estudo do impacto da injeção de umidade do G3DVAR na qualidade da análise.** 83 p. Dissertação (Mestrado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2016-10-03 2017.

Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m21b/2016/11.06.07.22>>. Acesso em: 11 abr. 2017. 5, 66

SAPUCCI, L. F.; DINIZ, F. L. R.; BASTARZ, C. F.; AVANÇO, L. A. Inclusion of global navigation satellite system radio occultation data into center for weather forecast and climate studies local ensemble transform kalman filter (LETKF) using the radio occultation processing package as an observation operator.

Meteorological Applications, John Wiley & Sons, Ltd., v. 23, n. 2, p. 328–338, 2016. ISSN 1469-8080. Disponível em: <<http://dx.doi.org/10.1002/met.1559>>. 4

SASAKI, Y. K. An objective analysis based on the variational method. 1958.

Disponível em: <https://books.google.com.br/books?id=1_z8sgEACAAJ>. 10

_____. Some basic formalisms in numerical variational analysis. **Monthly**

Weather Review, v. 98, n. 12, p. 875–883, 1970. Disponível em:

<[https://doi.org/10.1175/1520-0493\(1970\)098<0875:SBFINV>2.3.CO;2](https://doi.org/10.1175/1520-0493(1970)098<0875:SBFINV>2.3.CO;2)>. 10

SELLERS, P. J.; RANDALL, D. A.; COLLATZ, G. J.; BERRY, J. A.; FIELD, C. B.; DAZLICH, D. A.; ZHANG, C.; COLLELO, G. D.; BOUNOUA, L. A revised land surface parameterization (siB2) for atmospheric GCMs. part i: Model formulation. **Journal of Climate**, v. 9, n. 4, p. 676–705, 1996. Disponível em:

<[http://dx.doi.org/10.1175/1520-0442\(1996\)009<0676:ARLSPF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2)>. 22

SHERMAN, J.; MORRISON, W. J. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. **Ann. Math. Statist.**, The Institute of Mathematical Statistics, v. 21, n. 1, p. 124–127, 03 1950. Disponível em:

<<http://dx.doi.org/10.1214/aoms/1177729893>>. 142

SOUZA, S. S. d.; FIGUEROA, S. N.; HERDIES, D. L.; ARAVÉQUIA, J. A.; BONATTI, J. P.; GONÇALVES, L. G.; KUBOTA, P. Y.; SAPUCCI, L. F.;

FERREIRA, S. H.; MATTOS, J. G. Impact of the scheme of deep convection in the global data assimilation system LETKF of the CPTEC/INPE. In: THE MEETING OF THE AMERICAS., 8-12 aug. 2010, Foz do Iguaçu. **Posters**. [S.l.]: AGU, 2010. Acesso em: 26 maio 2016. 4

THÉPAUT, J.-N.; COURTIER, P. Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. **Quarterly Journal of the Royal Meteorological Society**, John Wiley & Sons, Ltd., v. 117, n. 502, p. 1225–1254, 1991. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1002/qj.49711750206>>. 18

TIEDTKE, M. The sensitivity of the time-mean large-scale flow to cumulus convection in the ECMWF model. In: ECMWF. **Workshop on Convection in Large-scale Numerical Models, 28 November to 1 December 1983**. Shinfield Park, Reading: ECMWF, 1983. p. 297–316. 22

TIPPETT, M. K.; ANDERSON, J. L.; BISHOP, C. H.; HAMILL, T. M.; WHITAKER, J. S. Ensemble square root filters. **Monthly Weather Review**, v. 131, n. 7, p. 1485–1490, 2003. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(2003\)131<1485:ESRF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2)>. 31

TODLING, R.; AKKRAOUI, A. E. **GMAO hybrid ensemble 3D-VAR**. Banff International Research Station, 2013. Disponível em: <http://www.birs.ca/workshops/2013/13w5139/files/amal_banff_gmao_hyb3dvar.pdf>. 12, 15

VITART, F.; ARDILOUZE, C.; BONET, A.; BROOKSHAW, A.; CHEN, M.; CODOREAN, C.; DÉQUÉ, M.; FERRANTI, L.; FUCILE, E.; FUENTES, M.; HENDON, H.; HODGSON, J.; KANG, H.-S.; KUMAR, A.; LIN, H.; LIU, G.; LIU, X.; MALGUZZI, P.; MALLAS, I.; MANOUSSAKIS, M.; MASTRANGELO, D.; MACLACHLAN, C.; MCLEAN, P.; MINAMI, A.; MLADEK, R.; NAKAZAWA, T.; NAJM, S.; NIE, Y.; RIXEN, M.; ROBERTSON, A. W.; RUTI, P.; SUN, C.; TAKAYA, Y.; TOLSTYKH, M.; VENUTI, F.; WALISER, D.; WOOLNOUGH, S.; WU, T.; WON, D.-J.; XIAO, H.; ZARIPOV, R.; ZHANG, L. The subseasonal to seasonal (s2s) prediction project database. **Bulletin of the American Meteorological Society**, v. 98, n. 1, p. 163–173, 2017. Disponível em: <<https://doi.org/10.1175/BAMS-D-16-0017.1>>. 2

WANG, X. Incorporating ensemble covariance in the gridpoint statistical interpolation variational minimization: A mathematical framework. **Monthly Weather Review**, v. 138, n. 7, p. 2990–2995, 2010. 57, 58

WANG, X.; BARKER, D. M.; SNYDER, C.; HAMILL, T. M. A hybrid ETKF–3DVAR data assimilation scheme for the WRF Model. part i: Observing system simulation experiment. **Monthly Weather Review**, American Meteorological Society, v. 136, n. 12, p. 5116–5131, 2008. Disponível em: <<http://dx.doi.org/10.1175/2008MWR2444.1>>. 5, 12, 14, 15, 57, 58, 60

_____. A hybrid ETKF–3DVAR Data assimilation scheme for the WRF Model. part II: Real observation experiments. **Monthly Weather Review**, American Meteorological Society, v. 136, n. 12, p. 5132–5147, 2008. Disponível em: <<http://dx.doi.org/10.1175/2008MWR2445.1>>. 5, 12, 14, 15, 57, 58, 60

WANG, X.; HAMILL, T. M.; WHITAKER, J. S.; BISHOP, C. H. A comparison of hybrid ensemble transform kalman filter–optimum interpolation and ensemble square root filter analysis schemes. **Monthly Weather Review**, American Meteorological Society, v. 135, n. 3, p. 1055–1076, 2007. Disponível em: <<http://dx.doi.org/10.1175/MWR3307.1>>. 5, 12, 14, 18

_____. A comparison of the hybrid and enSRF analysis schemes in the presence of model errors due to unresolved scales. **Monthly Weather Review**, v. 137, n. 10, p. 3219–3232, 2009. Disponível em: <<http://dx.doi.org/10.1175/2009MWR2923.1>>. 18

WANG, X.; LEI, T. GSI-based four-dimensional ensemble-variational (4DEnsvar) data assimilation for NCEP Global forecast system: Single resolution experiments with real data. 2013. 15, 58

WANG, X.; PARRISH, D.; KLEIST, D.; WHITAKER, J. GSI 3DVar-based ensemble-variational hybrid data assimilation for NCEP global forecast system: Single-resolution experiments. **Monthly Weather Review**, v. 141, n. 11, p. 4098–4117, 2013. Disponível em: <<http://dx.doi.org/10.1175/MWR-D-12-00141.1>>. 18

WHITAKER, J. S.; HAMILL, T. M. Ensemble data assimilation without perturbed observations. **Monthly Weather Review**, v. 130, n. 7, p. 1913–1924, 2002. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(2002\)130<1913:EDAWPO>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2)>. 31

WOODBURY, M. A. **Inverting modified matrices**. Princeton, NJ: Princeton University, 1950. (Statistical Research Group Memorandum Reports, 42). 142

WU, W.-S.; PURSER, R. J.; PARRISH, D. F. Three-dimensional variational analysis with spatially inhomogeneous covariances. **Monthly Weather Review**, American Meteorological Society, v. 130, n. 12, p. 2905–2916, 2013/11/18 2002.

Disponível em:

<[http://dx.doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2)>. 4, 18, 23, 43, 47, 62, 63

ZHANG, F.; ZHANG, M.; HANSEN, J. Coupling ensemble kalman filter with four-dimensional variational data assimilation. **Advances in Atmospheric Sciences**, SP Science Press, v. 26, n. 1, p. 1–8, 2009. ISSN 0256-1530. Disponível em: <<http://dx.doi.org/10.1007/s00376-009-0001-8>>. 12, 15

ZHANG, Z.; KRISHNAMURTI, T. N. A perturbation method for hurricane ensemble predictions. **Monthly Weather Review**, v. 127, n. 4, p. 447–469, 1999. Disponível em:

<[https://doi.org/10.1175/1520-0493\(1999\)127<0447:APMFHE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0447:APMFHE>2.0.CO;2)>. 6, 30

ZUPANSKI, M. Maximum likelihood ensemble filter: Theoretical aspects. **Monthly Weather Review**, American Meteorological Society, v. 133, n. 6, p. 1710–1726, 2013/11/18 2005. Disponível em: <<http://dx.doi.org/10.1175/MWR2946.1>>. 12, 14, 16

APÊNDICE A - Derivação da Equação de Análise a partir da Função Custo Variacional Tridimensional

Métodos variacionais como o 3DVar e sequenciais como a Interpolação Ótima, resolvem os mesmos problemas de otimização e pode-se mostrar que ambos são equivalentes. A derivação apresentada a seguir mostra que, a partir da função custo variacional tridimensional - em sua forma mais simples (i.e., sem considerar os termos de injeção das aplicações operacionais), é possível chegar à mesma equação de análise utilizada na Interpolação Ótima.

Seja a Eq. A.1 a função custo variacional tridimensional:

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}[\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x})] \quad (\text{A.1})$$

onde:

- \mathbf{x} é o vetor de estado a ser analisado;
- \mathbf{x}^b é o vetor de estado do background;
- \mathbf{B} é a matriz de covariâncias dos erros de background;
- \mathbf{y}^o é o vetor observação;
- \mathbf{R} é a matriz de covariâncias dos erros de observação;
- H é o operador observação não linear.

Na Eq. A.1, o termo $\mathbf{x} - \mathbf{x}^b$ representa o vetor incremento de análise e o termo $\mathbf{y}^o - H(\mathbf{x})$ representa o vetor inovação.

Para derivar a equação de análise a partir da Eq. A.1, consideramos (por simplicidade) apenas uma observação e que o vetor incremento de análise é pequeno o suficiente para que possamos linearizar o operador observação:

$$\mathbf{x} = \mathbf{x}^b + (\mathbf{x} - \mathbf{x}^b) \quad (\text{A.2})$$

Ou seja, na Eq. A.2 expressamos que o vetor de estado a ser analisado (a análise) é igual ao vetor de estado do background somado ao incremento de análise. Substituindo \mathbf{x} da Eq. A.2 na relação do vetor inovação:

$$\mathbf{y}^o - H(\mathbf{x}) = \mathbf{y}^o - H[\mathbf{x}^b + (\mathbf{x} - \mathbf{x}^b)] \quad (\text{A.3})$$

$$\mathbf{y}^o - H(\mathbf{x}) = \mathbf{y}^o - H(\mathbf{x}^b) - H(\mathbf{x} - \mathbf{x}^b) \quad (\text{A.4})$$

Na Eq. A.4 o operador observação não linear H é aplicado sobre o incremento de análise $\mathbf{x} - \mathbf{x}^b$, o qual é muito pequeno. Sendo assim, considera-se que o operador observação a ser aplicado sobre o incremento de análise pode ser linearizado. Isto significa que pequenas variações do vetor incremento de análise fazem com que H varie muito pouco. Portanto, a Eq. A.4 pode ser reescrita como:

$$\mathbf{y}^o - H(\mathbf{x}) = \mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \quad (\text{A.5})$$

onde:

- \mathbf{H} é a versão linearizada do operador observação não linear, H .

Substituindo a Eq. A.5 na Eq. A.1:

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\ &\quad + \frac{1}{2}[\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)] \end{aligned} \quad (\text{A.6})$$

Fatorando e desenvolvendo a Eq. A.6:

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}\{(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\ &\quad + [\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]\} \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} 2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\ &\quad + [\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]^T \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)] \end{aligned} \quad (\text{A.8})$$

$$\begin{aligned} 2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) \\ &\quad + \{[\mathbf{y}^o - H(\mathbf{x}^b)]^T - [\mathbf{H}(\mathbf{x} - \mathbf{x}^b)]^T\} \mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)] \end{aligned} \quad (\text{A.9})$$

$$\begin{aligned}
2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \\
&\quad + \{[\mathbf{y}^o - H(\mathbf{x}^b)]^T - (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T\} \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]
\end{aligned} \tag{A.10}$$

$$\begin{aligned}
2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \\
&\quad + \{[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} - (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1}\} [\mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{H}(\mathbf{x} - \mathbf{x}^b)]
\end{aligned} \tag{A.11}$$

$$\begin{aligned}
2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \\
&\quad + [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)] \\
&\quad - [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \\
&\quad - (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)] \\
&\quad + (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b)
\end{aligned} \tag{A.12}$$

Reorganizando a Eq. A.12 - aqui serão fatorados os termos 1 e 5 do lado direito:

$$\begin{aligned}
2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}^b) \\
&\quad + (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \\
&\quad - [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \\
&\quad - (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)] \\
&\quad + [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)]
\end{aligned} \tag{A.13}$$

Combinando os dois primeiros termos do lado direito da Eq. A.13, obtemos:

$$\begin{aligned}
2J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}^b)^T [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}] (\mathbf{x} - \mathbf{x}^b) \\
&\quad - [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b) \\
&\quad - (\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)] \\
&\quad + [\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)]
\end{aligned} \tag{A.14}$$

A forma da Eq. A.14 é a forma em que será aplicado o operador gradiente. O gradiente é aplicado na função custo variacional para se determinar o vetor de estado de análise \mathbf{x}^a tal que $\nabla_{\mathbf{x}} J(\mathbf{x}) = 0$. Ou seja, o mínimo de $J(\mathbf{x})$ fornece os valores do

vetor de estado analisado.

Teorema 1. Dada a função quadrática $F(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{d} + c$, onde \mathbf{A} é uma matriz simétrica (ou seja, é igual à sua transposta), \mathbf{d} é um vetor e c é um escalar, o gradiente de $F(\mathbf{x})$ é dado por $\nabla_{\mathbf{x}}F(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}$.

Demonstração.

$$\begin{aligned} F(\mathbf{x}) &= \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{d} + c \\ \frac{\partial F}{\partial \mathbf{x}} &= \frac{1}{2}(\mathbf{A}\mathbf{x} + \mathbf{A}^T \mathbf{x}) + \mathbf{d} \end{aligned}$$

Se \mathbf{A} é simétrica, então $\mathbf{A}^T = \mathbf{A}$, logo:

$$\begin{aligned} \frac{\partial F}{\partial \mathbf{x}} &= \frac{1}{2}(\mathbf{A}\mathbf{x} + \mathbf{A}\mathbf{x}) + \mathbf{d} \\ \frac{\partial F}{\partial \mathbf{x}} &= \frac{1}{2}(2\mathbf{A}\mathbf{x}) + \mathbf{d} \\ \frac{\partial F}{\partial \mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{d} \end{aligned}$$

□

Aplicando o operador gradiente na Eq. A.14, obtém-se:

$$\begin{aligned} \nabla_{\mathbf{x}} 2J(\mathbf{x}) &= \frac{\partial\{(\mathbf{x} - \mathbf{x}^b)^T [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}](\mathbf{x} - \mathbf{x}^b)\}}{\partial \mathbf{x}} \\ &\quad - \frac{\partial\{[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x} - \mathbf{x}^b)\}}{\partial \mathbf{x}} \\ &\quad - \frac{\partial\{(\mathbf{x} - \mathbf{x}^b)^T \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)]\}}{\partial \mathbf{x}} \\ &\quad + \frac{\partial\{[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)]\}}{\partial \mathbf{x}} \end{aligned} \tag{A.15}$$

Na Eq. A.15 observamos que o termo 4 do lado direito não depende do vetor de estado analisado \mathbf{x} e que portanto $\frac{\partial\{[\mathbf{y}^o - H(\mathbf{x}^b)]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x}^b)]\}}{\partial \mathbf{x}} = 0$, logo:

$$\begin{aligned}
2\nabla_{\mathbf{x}}J(\mathbf{x}) &= \frac{\partial\{(\mathbf{x} - \mathbf{x}^b)^T[\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}](\mathbf{x} - \mathbf{x}^b)\}}{\partial\mathbf{x}} \\
&\quad - \frac{\partial\{[\mathbf{y}^o - H(\mathbf{x}^b)]^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{x} - \mathbf{x}^b)\}}{\partial\mathbf{x}} \\
&\quad - \frac{\partial\{(\mathbf{x} - \mathbf{x}^b)^T\mathbf{H}^T\mathbf{R}^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)]\}}{\partial\mathbf{x}}
\end{aligned} \tag{A.16}$$

Utilizando o Teorema 1, obtém-se:

$$\begin{aligned}
2\nabla_{\mathbf{x}}J(\mathbf{x}) &= [(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^T(\mathbf{x} - \mathbf{x}^b) + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(\mathbf{x} - \mathbf{x}^b)] \\
&\quad - \frac{\partial\{[\mathbf{y}^o - H(\mathbf{x}^b)]^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{x} - \mathbf{x}^b)\}}{\partial\mathbf{x}} \\
&\quad - (\mathbf{H}^T\mathbf{R}^{-1})^T[\mathbf{y}^o - H(\mathbf{x}^b)] + (\mathbf{H}^T\mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)]
\end{aligned} \tag{A.17}$$

O segundo termo da Eq. A.17 resulta em zero, pois a derivada da constante $[\mathbf{y}^o - H(\mathbf{x}^b)]\mathbf{R}^{-1}\mathbf{H}$ é zero.

Utilizando-se do fato de que as relações entre as matrizes nos termos 1 a 3 do lado direito da Eq. A.16 geram matrizes simétricas (e.g., $(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^T = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})$), a Eq. A.17 reduz-se a:

$$2\nabla_{\mathbf{x}}J(\mathbf{x}) = 2[(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(\mathbf{x} - \mathbf{x}^b)] - 2(\mathbf{H}^T\mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \tag{A.18}$$

$$\nabla_{\mathbf{x}}J(\mathbf{x}) = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(\mathbf{x} - \mathbf{x}^b) - (\mathbf{H}^T\mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \tag{A.19}$$

No início foi feita consideração de que o vetor de estados analisado \mathbf{x} é a própria análise quando $\nabla_{\mathbf{x}}J(\mathbf{x}) = 0$, logo iguala-se a Eq. A.19 a zero, de forma que $\mathbf{x} = \mathbf{x}^a$:

$$0 = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{H}^T\mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \tag{A.20}$$

Reorganizando a Eq. A.20:

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] = 0 \quad (\text{A.21})$$

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) = (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (\text{A.22})$$

$$(\mathbf{x}^a - \mathbf{x}^b) = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (\text{A.23})$$

A partir da Eq. A.23, obtemos a equação de análise:

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{R}^{-1})[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (\text{A.24})$$

O cálculo do termo de peso dado por $(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$ pode ser realizado aplicando-se a identidade de Sherman-Morrison-Woodbury (SHERMAN; MORRISON, 1950; WOODBURY, 1950), que facilita numericamente o cálculo da inversa da soma das matrizes \mathbf{B}^{-1} e $\mathbf{H}\mathbf{B}\mathbf{H}^T$. Para isso, a identidade será usada de maneira que a equação de análise mostrada na Eq. A.24 poderá ser reescrita em sua forma equivalente (apresentada pela Eq. A.26 adiante). A identidade de Sherman-Morrison-Woodbury é demonstrada da seguinte forma:

$$\begin{aligned} \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\ &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\ &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \\ &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} (\mathbf{H}^T \mathbf{R}^{-1}) \end{aligned} \quad (\text{A.25})$$

Utilizando-se do fato de que o termo de peso utilizado na Eq. A.24 é equivalente ao termo $\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$, podemos então reescrever a Eq. A.24 da seguinte maneira:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}[\mathbf{y}^o - H(\mathbf{x}^b)] \quad (\text{A.26})$$

A vantagem da forma da equação de análise apresentada na Eq. A.26 está no custo computacional, sendo menos custoso calcular o termo $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ do que $(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$.

APÊNDICE B - Pré-condicionamento da Função Custo Variacional com base em **B** Completa

O sistema GSI possui algumas metodologias para a minimização da função custo variacional (3D ou 4D). Dado o custo computacional relacionado às dimensões dos vetores de estado e observação, além da matriz de covariâncias dos erros de previsão, técnicas computacionais são introduzidas a fim de minimizar o custo computacional envolvido. Como exemplo, é apresentado a seguir o pré-condicionamento da função custo variacional tridimensional com base na matriz **B** completa.

Seja a Eq. B.1 a função custo variacional tridimensional na forma incremental:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T \mathbf{B}^{-1}(\delta\mathbf{x}) + \frac{1}{2}[\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})] \quad (\text{B.1})$$

Onde:

- $\delta\mathbf{x}$ é o vetor incremento de análise dado por $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^b$, em que \mathbf{x} é o vetor de estado a ser analisado (a verdade) e \mathbf{x}^b é o vetor de estado do background;
- **B** é a matriz de covariâncias dos erros de background;
- \mathbf{y}'^o é o vetor inovação, dado por $\mathbf{y}'^o = \mathbf{y}^o - H(\mathbf{x}^b)$, em que \mathbf{y}^o é o vetor observação e H o operador observação não linear;
- **R** é a matriz de covariâncias dos erros de observação;
- **H** é o operador observação linear.

O cálculo da análise variacional é feito a partir da minimização da função custo (Eq. B.1), ou seja, $\nabla_{\delta\mathbf{x}} J = 0$. Um algoritmo de minimização bastante comum é o Gradiente Conjugado Precondicionado. Este tipo de algoritmo apresenta a vantagem de convergir rapidamente para a solução. Para a aplicação deste algoritmo, algumas abordagens do método variacional consideram, como pré-condicionamento, a matriz de covariância dos erros de previsão (**B**) completa, como é o caso do Gridpoint Statistical Interpolation (GSI). Alguns sistemas variacionais realizam a minimização da função custo considerando a raiz quadrada da matriz **B**.

Para aplicar o algoritmo do Gradiente Conjugado Precondicionado, é necessária a definição de uma nova variável **z**, definida como:

$$\mathbf{z} = \mathbf{B}^{-1}\delta\mathbf{x} \quad (\text{B.2})$$

Esta nova variável será utilizada para calcular $\nabla_{\delta\mathbf{x}}J = 0$. Substituindo-se a Eq. B.2 na Eq. B.1:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T\mathbf{z} + \frac{1}{2}[\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})]^T\mathbf{R}^{-1}[\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})] \quad (\text{B.3})$$

Desenvolvendo a Eq. B.3, obtemos:

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T\mathbf{z} + \frac{1}{2}\{[(\mathbf{y}'^o)^T - (\delta\mathbf{x})^T\mathbf{H}^T]\mathbf{R}^{-1}[\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})]\} \quad (\text{B.4})$$

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T\mathbf{z} + \frac{1}{2}\{[(\mathbf{y}'^o)^T\mathbf{R}^{-1} - (\delta\mathbf{x})^T\mathbf{H}^T\mathbf{R}^{-1}][\mathbf{y}'^o - \mathbf{H}(\delta\mathbf{x})]\} \quad (\text{B.5})$$

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T\mathbf{z} + \frac{1}{2}\{[(\mathbf{y}'^o)^T\mathbf{R}^{-1}(\mathbf{y}'^o) - (\mathbf{y}'^o)^T\mathbf{R}^{-1}\mathbf{H}(\delta\mathbf{x}) - (\delta\mathbf{x})^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}'^o) + (\delta\mathbf{x})^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\delta\mathbf{x})]\} \quad (\text{B.6})$$

$$J(\delta\mathbf{x}) = \frac{1}{2}(\delta\mathbf{x})^T\mathbf{z} + \frac{1}{2}(\mathbf{y}'^o)^T\mathbf{R}^{-1}(\mathbf{y}'^o) - \frac{1}{2}(\mathbf{y}'^o)^T\mathbf{R}^{-1}\mathbf{H}(\delta\mathbf{x}) - \frac{1}{2}(\delta\mathbf{x})^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}'^o) + \frac{1}{2}(\delta\mathbf{x})^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\delta\mathbf{x}) \quad (\text{B.7})$$

Calculando o gradiente da Eq. B.7 em relação a $\delta\mathbf{x}$ (utilizando as ideias apresentadas no desenvolvimento da Eq. A.17), obtemos:

$$\nabla_{\delta\mathbf{x}}J = \mathbf{z} - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}'^o) + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\delta\mathbf{x}) \quad (\text{B.8})$$

$$\nabla_{\delta\mathbf{x}}J = \mathbf{z} + \mathbf{H}^T\mathbf{R}^{-1}[\mathbf{H}(\delta\mathbf{x}) - (\mathbf{y}'^o)] \quad (\text{B.9})$$

Da mesma forma, podemos partir da Eq. B.7 e calcular o gradiente em relação a \mathbf{z} . Mas temos que $\mathbf{z} = \mathbf{B}^{-1}\delta\mathbf{x}$, e portanto:

$$\delta \mathbf{x} = \mathbf{zB} \quad (\text{B.10})$$

Substituindo a Eq. B.10 na Eq. B.7, obtemos:

$$\begin{aligned} J(\mathbf{zB}) &= \frac{1}{2}(\mathbf{zB})^T \mathbf{z} + \frac{1}{2}(\mathbf{y}'^o)^T \mathbf{R}^{-1}(\mathbf{y}'^o) - \frac{1}{2}(\mathbf{y}'^o)^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{zB}) \\ &\quad - \frac{1}{2}(\mathbf{zB})^T \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}'^o) + \frac{1}{2}(\mathbf{zB})^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{zB}) \end{aligned} \quad (\text{B.11})$$

O gradiente $\nabla_{\mathbf{z}} J$ é calculado, portanto, a partir da Eq. B.11:

$$\nabla_{\mathbf{z}} J = \mathbf{Bz} - \mathbf{BH}^T \mathbf{R}^{-1}(\mathbf{y}'^o) + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{HB}(\mathbf{zB}) \quad (\text{B.12})$$

$$\nabla_{\mathbf{z}} J = \mathbf{B}[\mathbf{z} - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}'^o) + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{zB})] \quad (\text{B.13})$$

$$\nabla_{\mathbf{z}} J = \mathbf{B}\{\mathbf{z} + \mathbf{H}^T \mathbf{R}^{-1}[\mathbf{H}(\mathbf{zB}) - (\mathbf{y}'^o)]\} \quad (\text{B.14})$$

Como $\mathbf{zB} = \delta \mathbf{x}$, obtemos:

$$\nabla_{\mathbf{z}} J = \mathbf{B}\{\mathbf{z} + \mathbf{H}^T \mathbf{R}^{-1}[\mathbf{H}(\delta \mathbf{x}) - (\mathbf{y}'^o)]\} \quad (\text{B.15})$$

A expressão entre chaves na Eq. B.15 é a mesma que está do lado direito na Eq. B.9, e portanto:

$$\nabla_{\mathbf{z}} J = \mathbf{B} \nabla_{\delta \mathbf{x}} J \quad (\text{B.16})$$

Este algoritmo é vantajoso porque nele não há a necessidade de se calcular explicitamente \mathbf{B}^{-1} (e.g., Eq. A.19), sendo necessária apenas uma transformação de variáveis.

APÊNDICE C - Filtro Recursivo 1D - Grau 1

Neste apêndice é apresentada a derivação de um filtro recursivo unidimensional de grau 1 baseado no trabalho de Purser et al. (2003a). Os desenvolvimentos apresentados pelo autor sugerem a construção de um filtro recursivo de 2 ou mais dimensões (aplicado em 2 ou mais dimensões) a partir de um polinômio construído a com a expansão em série de Taylor de uma função para um operador de diferenças finitas ($D_{(n)}^*$). A determinação deste operador é deveras complicada e é baseada na teoria do “cálculo de operadores” (DAHLQUIST; BJÖRCK, 1974). Portanto, a derivação apresentada aqui é fornecida com fins didáticos e dará suporte para a derivação de um filtro recursivo unidimensional de grau 2, o qual é utilizado no trabalho de Purser et al. (2003a). Em algumas partes da derivação, será necessário fazer referências à algumas equações e tabelas escritas no trabalho de Purser et al. (2003a), as quais estarão indicadas por um par de colchetes “[]”.

Partindo-se do polinômio da Eq. [3.12] de Purser et al. (2003a):

$$D_{(n)}^* = 1 + b_{1,1} \frac{\sigma^2}{2} K + \left[b_{1,2} \left(\frac{\sigma^2}{2} \right) + \frac{b_{2,2}}{2!} \left(\frac{\sigma^2}{2} \right)^2 \right] K^2 + \dots + \left[\sum_{j=1}^n \frac{b_{j,n}}{j!} \left(\frac{\sigma^2}{2} \right)^j \right] K^n \quad (\text{C.1})$$

Onde:

- $D_{(n)}^*$ é um operador de diferenças finitas composto pela n -ésima expansão de K ;
- K é um operador de diferenças finitas;
- $b_{j,n}$ são coeficientes para o filtro quasi-Gaussiano (segundo a Tabela [1] de Purser et al., 2003a);
- σ é a razão entre a e δx , de forma que $\sigma = \frac{a}{\delta x}$;
- a é o comprimento da grade unidimensional;
- δx é o espaçamento da grade.

Utiliza-se a forma da Eq. C.2 a seguir, para representar a Eq. C.1:

$$D_{(n)}^* = \prod_{p=1}^n \left(1 - \frac{K}{\kappa_p}\right) \quad (\text{C.2})$$

Expandindo-se a Eq. C.2 para $n = 1$ (expansão de primeira ordem), obtemos:

$$D_{(1)}^* = 1 - \frac{K}{\kappa_1} \quad (\text{C.3})$$

Comparando-se a Eq. C.3 com a Eq. C.1 (quando $n = 1$), obtemos:

$$-\frac{1}{\kappa_1} = b_{1,1} \frac{\sigma^2}{2} \quad (\text{C.4})$$

Na Tabela [1] de Purser et al. (2003a), temos que $b_{1,1} = 1$.

Logo,

$$-\frac{1}{\kappa_1} = \frac{\sigma^2}{2} \quad (\text{C.5})$$

$$-\kappa_1 = \frac{2}{\sigma^2} \quad (\text{C.6})$$

$$\kappa_1 = -\frac{2}{\sigma^2}, \text{ com } \sigma \neq 0 \quad (\text{C.7})$$

Onde:

- κ_1 é a raiz de um polinômio K (que é o operador de diferenças finitas).

Entre as Eqs. [A.3] e [A.4] de Purser et al. (2003a), está estabelecida a equação:

$$\omega_p = 1 - \frac{\kappa_p}{2} \quad (\text{C.8})$$

Para $p = 1$:

$$\omega_1 = 1 - \frac{\kappa_1}{2} \quad (\text{C.9})$$

E portanto,

$$\omega_1 = 1 + \frac{1}{\sigma^2}, \text{ com } \sigma \neq 0 \quad (\text{C.10})$$

Onde:

- ω_1 é um coeficiente de uma equação quadrática em z , cujas raízes são os pares conjugados de ζ_p .

Partindo da Eq. [A.4] de Purser et al. (2003a):

$$\zeta_p = [\omega_p + i(\omega_p^2 - 1)^{\frac{1}{2}}]^{\pm 1} \quad (\text{C.11})$$

Para $p = 1$:

$$\zeta_1 = [\omega_1 + i(\omega_1^2 - 1)^{\frac{1}{2}}]^{\pm 1} \quad (\text{C.12})$$

Onde:

- ζ_1 é uma raiz complexa de z , e ela é representada por ζ_1' e ζ_1'' .

Nesta etapa, é importante definir as formas complexas de ζ_1' e ζ_2' . Para isto, utilizamos as formas representadas nas Eqs. C.10 e C.12:

$$\zeta_1' = [\omega_1 + i(\omega_1^2 - 1)^{\frac{1}{2}}]^{+1} \quad (\text{C.13})$$

$$\zeta_1' = \left(1 + \frac{1}{\sigma^2}\right) + i \left[\left(1 + \frac{1}{\sigma^2}\right)^2 - 1\right]^{\frac{1}{2}} \quad (\text{C.14})$$

$$\zeta_1' = \left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left[\left(\frac{\sigma^2 + 1}{\sigma^2}\right)^2 - 1\right]^{\frac{1}{2}} \quad (\text{C.15})$$

$$\zeta_1' = \left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left[\frac{(\sigma^2 + 1)^2 - \sigma^4}{\sigma^4}\right]^{\frac{1}{2}} \quad (\text{C.16})$$

$$\zeta_1' = \left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left[\frac{\sigma^4 + 2\sigma^2 + 1 - \sigma^4}{\sigma^4}\right]^{\frac{1}{2}} \quad (\text{C.17})$$

$$\zeta_1' = \left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left[\frac{2\sigma^2 + 1}{\sigma^4}\right]^{\frac{1}{2}} \quad (\text{C.18})$$

$$\zeta_1' = \left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left(\frac{2\sigma^2 + 1}{\sigma^4}\right)^{\frac{1}{2}} \quad (\text{C.19})$$

O par conjugado de ζ_1' é ζ_1'' , e portanto obtemos:

$$\zeta_1'' = \left[\left(\frac{\sigma^2 + 1}{\sigma^2}\right) + i \left(\frac{2\sigma^2 + 1}{\sigma^4}\right)^{\frac{1}{2}}\right]^{-1}, \text{ com } \sigma \neq 0 \quad (\text{C.20})$$

Segundo Purser et al. (2003a), a menor raiz entre ζ_1' e ζ_1'' , permite a fatoração da Eq. [A.5].

A partir da Eq. [A.5], obtemos:

$$1 - \frac{K}{\kappa_p} = \left(\frac{1 - \zeta_p z^{-1}}{1 - \zeta_p}\right) \left(\frac{1 - \zeta_p z}{1 - \zeta_p}\right) \quad (\text{C.21})$$

Para $p = 1$:

$$1 - \frac{K}{\kappa_1} = \left(\frac{1 - \zeta_1 z^{-1}}{1 - \zeta_1}\right) \left(\frac{1 - \zeta_1 z}{1 - \zeta_1}\right), \text{ com } |\zeta_1| \neq 1 \quad (\text{C.22})$$

A partir das Eqs. [A.6] e [A.7], obtemos:

$$\mathbf{A} = \prod_{p=1}^n \left(\frac{1 - \zeta_p z^{-1}}{1 - \zeta_p} \right) \quad (\text{C.23})$$

$$\mathbf{B} = \prod_{p=1}^n \left(\frac{1 - \zeta_p z}{1 - \zeta_p} \right) \quad (\text{C.24})$$

Para $n = 1$:

$$\mathbf{A} = \left(\frac{1 - \zeta_1 z^{-1}}{1 - \zeta_1} \right) \quad (\text{C.25})$$

$$\mathbf{B} = \left(\frac{1 - \zeta_1 z}{1 - \zeta_1} \right), \text{ com } |\zeta_1| \neq 1 \quad (\text{C.26})$$

Onde:

- z é um operador de deslocamento para a direita;
- z^{-1} é um operador de deslocamento para a esquerda.

Com as Eqs. C.25 e C.26, finalmente obtemos a representação de um filtro recursivo unidimensional em termos de z , z^{-1} e ζ_1 :

$$D_{(1)}^* = \left(\frac{1 - \zeta_1 z^{-1}}{1 - \zeta_1} \right) \left(\frac{1 - \zeta_1 z}{1 - \zeta_1} \right) \quad (\text{C.27})$$

APÊNDICE D - Filtro Recursivo 1D - Grau 2

A derivação de um filtro recursivo unidimensional de grau 2 é apresentado neste apêndice, como uma extensão do filtro derivado no Apêndice C. A derivação segue a mesma ideia apresentada anteriormente, mas parte da expansão em série de Taylor em grau 2 do operador de diferenças finitas ($D_{(n)}^*$), seguindo as ideias apresentadas em Purser et al. (2003a). Neste apêndice, também serão utilizados “[]” para fazer referências à determinadas equações do trabalho de Purser et al. (2003a).

Partindo-se da forma apresentada na Eq. C.2 para $n = 2$, obtemos:

$$D_{(2)}^* = \left(1 - \frac{K}{\kappa_1}\right) \left(1 - \frac{K}{\kappa_2}\right) \quad (\text{D.1})$$

$$D_{(2)}^* = 1 - \frac{K}{\kappa_2} - \frac{K}{\kappa_1} + \frac{K^2}{\kappa_1\kappa_2} \quad (\text{D.2})$$

$$D_{(2)}^* = 1 - \left(\frac{1}{\kappa_1} - \frac{1}{\kappa_2}\right) K + \left(\frac{1}{\kappa_1\kappa_2}\right) K^2 \quad (\text{D.3})$$

Comparando a Eq. D.3 com a Eq. C.2, para $n = 2$:

$$-\left(\frac{1}{\kappa_1} - \frac{1}{\kappa_2}\right) = b_{1,1} \frac{\sigma^2}{2} \quad (\text{D.4})$$

Com $b_{1,1} = 1$ (segundo a Tabela [1] de Purser et al., 2003a):

$$-\left(\frac{1}{\kappa_1} - \frac{1}{\kappa_2}\right) = \frac{\sigma^2}{2} \quad (\text{D.5})$$

$$\frac{\kappa_1 + \kappa_2}{\kappa_1\kappa_2} = -\frac{\sigma^2}{2} \quad (\text{D.6})$$

Além disso, temos que:

$$\frac{1}{\kappa_1\kappa_2} = b_{1,2} \frac{\sigma^2}{2} + \frac{b_{2,2}}{2!} \left(\frac{\sigma_2}{2}\right)^2 \quad (\text{D.7})$$

Com $b_{1,2} = \frac{1}{2}$ e $b_{2,2} = 1$ (segundo a Tabela [1] de Purser et al., 2003a):

$$\frac{1}{\kappa_1 \kappa_2} = \frac{\sigma^2}{24} + \frac{\sigma^4}{8} \quad (\text{D.8})$$

$$\frac{1}{\kappa_1 \kappa_2} = \frac{\sigma^2 + 3\sigma^4}{24} \quad (\text{D.9})$$

$$\kappa_1 \kappa_2 = \frac{24}{\sigma^2 + 3\sigma^4}, \text{ com } \sigma \neq 0 \quad (\text{D.10})$$

Substituindo a Eq. D.10 na Eq. D.6:

$$\frac{\frac{\kappa_1 + \kappa_2}{24}}{\sigma^2 + 3\sigma^4} = -\frac{\sigma^2}{2} \quad (\text{D.11})$$

$$(\kappa_1 + \kappa_2) \left(\frac{24}{\sigma^2 + 3\sigma^4} \right) = -\frac{\sigma^2}{2} \quad (\text{D.12})$$

$$\kappa_1 + \kappa_2 = -\frac{12\sigma^2}{\sigma^2 + 3\sigma^4}, \text{ com } \sigma \neq 0 \quad (\text{D.13})$$

Assim como no caso do filtro recursivo de grau 1, precisamos determinar κ_1 e κ_2 , que são as raízes do polinômio K , dado na Eq. C.2.

Podemos utilizar a fórmula de Bhaskara para determinar os coeficientes a , b e c abaixo, para κ_1 e κ_2 :

$$\kappa_1 = -\frac{b + (b^2 - 4ac)^{\frac{1}{2}}}{2a} \quad (\text{D.14})$$

$$\kappa_2 = -\frac{b - (b^2 - 4ac)^{\frac{1}{2}}}{2a} \quad (\text{D.15})$$

Com isso, obtém-se:

$$b = -\left(\frac{1}{\kappa_1} - \frac{1}{\kappa_2} \right) = -\left(\frac{\kappa_1 + \kappa_2}{\kappa_1 \kappa_2} \right) = \frac{\sigma^2}{2} \quad (\text{D.16})$$

Que é exatamente a Eq. D.6.

Além disso:

$$a = \left(\frac{1}{\kappa_1 \kappa_2} \right) = \frac{\sigma^2 + 3\sigma^4}{24} \quad (\text{D.17})$$

Que é exatamente a Eq. D.10. Consequentemente, $c = 1$.

Substituindo a , b , e c nas Eqs. D.14 e D.15, obtemos:

$$\kappa_1 = \frac{\frac{\sigma^2}{2} + \left[\left(-\frac{\sigma^2}{2} \right)^2 - 4 \left(\frac{\sigma^2 + 3\sigma^4}{24} \right) \right]^{\frac{1}{2}}}{2 \left(\frac{\sigma^2 + 3\sigma^4}{24} \right)} \quad (\text{D.18})$$

$$\kappa_1 = \frac{\frac{\sigma^2}{2} + \left[\frac{\sigma^4}{4} - \left(\frac{\sigma^2 + 3\sigma^4}{6} \right) \right]^{\frac{1}{2}}}{\left(\frac{\sigma^2 + 3\sigma^4}{12} \right)} \quad (\text{D.19})$$

$$\kappa_1 = \frac{\frac{\sigma^2}{2} + \left[\frac{3\sigma^4 - 2\sigma^2 - 6\sigma^4}{12} \right]^{\frac{1}{2}}}{\left(\frac{\sigma^2 + 3\sigma^4}{12} \right)} \quad (\text{D.20})$$

$$\kappa_1 = \frac{\frac{\sigma^2}{2} + \left[-\left(\frac{3\sigma^4 + 2\sigma^2}{12} \right) \right]^{\frac{1}{2}}}{\left(\frac{\sigma^2 + 3\sigma^4}{12} \right)} \quad (\text{D.21})$$

$$\kappa_1 = \left(\frac{12}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} + i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right] \quad (\text{D.22})$$

Consequentemente,

$$\kappa_2 = \left(\frac{12}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} - i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right], \text{ com } \sigma \neq 0 \quad (\text{D.23})$$

As raízes κ_1 e κ_2 do polinômio K são complexas. Além disso, o fator $\frac{12}{\sigma^2 + 3\sigma^4}$ que multiplica os números complexos são exatamente as somas de κ_1 e κ_2 .

Da mesma forma como foi feito para o filtro recursivo 1D de grau 1, precisamos determinar ω_1 e ω_2 que dependem dos valores definidos de κ_1 e κ_2 , respectivamente. ω_1 e ω_2 serão utilizados para determinar ζ_1 e ζ_2 , que por sua vez definem as matrizes de transição para frente e para trás do filtro recursivo.

Considerando κ_1 e κ_2 a partir da Eq. C.1:

$$\omega_1 = 1 - \frac{\kappa_1}{2} \quad (\text{D.24})$$

$$\omega_1 = 1 - \frac{1}{2} \left\{ \left(\frac{12}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} + i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right] \right\} \quad (\text{D.25})$$

$$\omega_1 = 1 - \left\{ \left(\frac{6}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} + i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right] \right\} \quad (\text{D.26})$$

$$\omega_1 = 1 - \frac{3\sigma^2}{\sigma^2 + 3\sigma^4} - i \left(\frac{6}{\sigma^2 + 3\sigma^4} \right) \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \quad (\text{D.27})$$

O mesmo pode ser feito para ω_2 :

$$\omega_2 = 1 - \frac{\kappa_2}{2} \quad (\text{D.28})$$

$$\omega_2 = 1 - \frac{1}{2} \left\{ \left(\frac{12}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} - i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right] \right\} \quad (\text{D.29})$$

$$\omega_2 = 1 - \left\{ \left(\frac{6}{\sigma^2 + 3\sigma^4} \right) \left[\frac{\sigma^2}{2} - i \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \right] \right\} \quad (\text{D.30})$$

$$\omega_2 = 1 - \frac{3\sigma^2}{\sigma^2 + 3\sigma^4} + i \left(\frac{6}{\sigma^2 + 3\sigma^4} \right) \left(\frac{3\sigma^4 + 2\sigma^2}{12} \right)^{\frac{1}{2}} \quad (\text{D.31})$$

Utilizando as Eqs. C.11 e C.12, podemos escrever as expressões para ζ_1 e ζ_2 :

$$\zeta_1' = [\omega_1 + i(\omega_1^2 - 1)^{\frac{1}{2}}]^{+1} \quad (\text{D.32})$$

$$\zeta_1'' = [\omega_1 + i(\omega_1^2 - 1)^{\frac{1}{2}}]^{-1} \quad (\text{D.33})$$

E também:

$$\zeta_2' = [\omega_2 + i(\omega_2^2 - 1)^{\frac{1}{2}}]^{+1} \quad (\text{D.34})$$

$$\zeta_2'' = [\omega_2 + i(\omega_2^2 - 1)^{\frac{1}{2}}]^{-1} \quad (\text{D.35})$$

Considerando ζ_1 a menor raiz entre ζ_1' e ζ_1'' e ζ_2 a menor raiz entre ζ_2' e ζ_2'' , podemos finalmente escrever as expressões para as matrizes \mathbf{A} e \mathbf{B} (utilizando as Eqs. [A.6] e [A.7] de Purser et al., 2003a), de forma a obtermos a expressão final do polinômio $D_{(2)}^*$:

$$\mathbf{A} = \begin{pmatrix} \frac{1 - \zeta_1 z^{-1}}{1 - \zeta_1} \end{pmatrix} \begin{pmatrix} \frac{1 - \zeta_2 z^{-1}}{1 - \zeta_2} \end{pmatrix} \quad (\text{D.36})$$

$$\mathbf{B} = \begin{pmatrix} \frac{1 - \zeta_1 z}{1 - \zeta_1} \end{pmatrix} \begin{pmatrix} \frac{1 - \zeta_2 z}{1 - \zeta_2} \end{pmatrix} \quad (\text{D.37})$$

Como $D_{(n)}^* = \mathbf{A}\mathbf{B}$, logo:

$$D_{(2)}^* = \begin{pmatrix} \frac{1 - \zeta_1 z^{-1}}{1 - \zeta_1} \end{pmatrix} \begin{pmatrix} \frac{1 - \zeta_2 z^{-1}}{1 - \zeta_2} \end{pmatrix} \begin{pmatrix} \frac{1 - \zeta_1 z}{1 - \zeta_1} \end{pmatrix} \begin{pmatrix} \frac{1 - \zeta_2 z}{1 - \zeta_2} \end{pmatrix} \quad (\text{D.38})$$

ANEXO A - Artigo Científico I

Submission Confirmation

[Print](#)

Thank you for your revision

Submitted to Revista Brasileira de Meteorologia

Manuscript ID RBMET-2016-0146.R1

Title Matriz de Covariâncias dos Erros de Previsão aplicada ao Sistema de Assimilação de Dados Global do CPTEC: Experimentos com Observação Única

Authors Bastarz, Carlos
Herdies, Dirceu
Sapucci, Luiz Fernando

Date Submitted 06-Mar-2017

[Author Dashboard](#)

Revista Brasileira de Meteorologia

Decision Letter (RBMET-2016-0146.R1)**From:** nelsonluisdias@gmail.com**To:** carlos.frederico@cptec.inpe.br**CC:****Subject:** Revista Brasileira de Meteorologia - Decision on Manuscript ID RBMET-2016-0146.R1**Body:** 03-May-2017

Dear Mr. Bastarz:

It is a pleasure to accept your manuscript entitled "Matriz de Covariâncias dos Erros de Previsão aplicada ao Sistema de Assimilação de Dados Global do CPTEC: Experimentos com Observação Única" in its current form for publication in the Revista Brasileira de Meteorologia. The comments of the reviewer(s) who reviewed your manuscript are included at the foot of this letter.

Thank you for your fine contribution. On behalf of the Editors of the Revista Brasileira de Meteorologia, we look forward to your continued contributions to the Journal.

Sincerely,
Dr. Nelson Dias
Associate Editor, Revista Brasileira de Meteorologia
nelsonluisdias@gmail.com, nldias@ufpr.br

Entire Scoresheet:
Reviewer: 1

Recommendation: Accept

Comments:
(There are no comments.)

Additional Questions:
Rating:

Interest: 1. Excellent

Quality: 2. Good

Originality: 2. Good

Overall: 2. Good

Date Sent: 03-May-2017 Close Window

© Thomson Reuters | © ScholarOne, Inc., 2017. All Rights Reserved.

Resumo: A matriz de covariâncias dos erros de previsão representa uma importante componente de um sistema de assimilação de dados. Pode-se mostrar matematicamente que os incrementos de análise são diretamente proporcionais à matriz de covariâncias. Considerando-se este resultado, é correto afirmar que a habilidade de um sistema de assimilação de dados está diretamente relacionada às características da matriz de covariâncias, sejam elas representadas por comprimentos de escala horizontais e verticais e amplitudes (desvios-padrão e variâncias). Se o sistema de assimilação de dados utiliza as observações para corrigir as previsões do modelo, ponderando-se os erros das observações e previsões, então o uso de uma matriz de covariâncias não ajustada pode impactar de forma significativa o processo de assimilação de dados. No CPTEC, tem-se investido esforços para o ajuste da matriz de covariâncias utilizada na assimilação de dados. Neste trabalho é feita, portanto, uma discussão conceitual da matriz de covariâncias, expressando a sua importância e a forma como é aplicada em um sistema operacional. Além disso, é apresentada uma caracterização quantitativa e qualitativa dessa matriz para o modelo de circulação geral da atmosfera do CPTEC, e quais são as características resultantes dos incrementos de análise produzidos no sistema GSI utilizado pelo centro.

Palavras-Chave: Matriz de Covariâncias dos Erros de Previsão, Assimilação de Dados, Previsão Numérica do Tempo.

Abstract: The background error covariance matrix represents a key component of a data assimilation system. It can be shown mathematically that the analysis increments are directly proportional to the covariance matrix. As a result it is correct to state that the performance of a data assimilation system is related to the characteristics of the covariance matrix, in terms of: horizontal and vertical length scales, standard deviations and variances. Considering the information that the data assimilation system uses the observations to correct the model forecasts weighting the model and observations errors, thus the use of an unadjusted covariance matrix can impact the resulting analysis at a great level. At CPTEC efforts has been made in order to adjust the covariance matrix for its application at the operations. In this work it is presented a conceptual overview on the subject, enlightening the importance of the background error covariance matrix and how it is treated in an operational data assimilation system. Furthermore, it is also made a quantitative and qualitative characterization of the background error covariance matrix calculated using the CPTEC global forecast model and the differences in the resulting analysis increments.

Keywords: Background Error Covariance Matrix, Data Assimilation, Numerical Weather Prediction.

1 Introdução

Assimilação de Dados é uma técnica fundamentada em métodos de estimação e otimização. Estes métodos são utilizados para realizar a tarefa de controlar um sistema dinâmico (como um modelo numérico da atmosfera terrestre) na presença de incertezas (Stengel, 1994). Muitas vezes tais incertezas não podem ser diretamente inferidas e é comum assumir uma determinada distribuição estatística para os erros (eg., Gaussiana) para a sua aplicação nos mais variados problemas que podem ser tratados utilizando esses métodos. Nesse processo, estão presentes duas principais fontes de incertezas: a confiança atribuída para as observações na correção das previsões de um modelo; e a parcela das informações fornecidas por esse modelo e que devem ser preservadas na solução final. A primeira fonte de incertezas é tratada no que convencionalmente chama-se de matriz de covariância dos erros das observações (geralmente referenciada como matriz \mathbf{R} , eg., Ide et al., 1997). A incerteza do modelo é tratada, portanto, na matriz de covariância dos erros de previsão ou modelagem (geralmente referenciada como matriz \mathbf{B}) e é uma das principais - se não dizer, a mais importante componente de um sistema de assimilação de dados. Em sistemas variacionais e sequenciais, estas duas componentes são relacionadas por uma equação matemática onde são ponderadas as contribuições das observações e do modelo no cálculo das inovações trazidas pelas observações e dos incrementos de análise calculados.

Sistemas de assimilação de dados utilizam observações que são distribuídas de forma irregular no tempo e no espaço para corrigir/atualizar um estado da atmosfera previamente simulado por um modelo de Previsão Numérica de Tempo (PNT). Este estado atmosférico é também chamado de previsão de curto prazo e além de ser utilizado nesse processo de atualização com as observações, ele serve também para complementar parte das informações que são trazidas pelas observações. Então, para que seja possível realizar algum tipo de operação matemática entre a previsão do modelo e as observações, ambas as informações devem estar sob o mesmo domínio, de forma que a inovação (ie., a contribuição trazida pelas observações ao campo representado pela previsão de curto prazo) possa ser calculada. Desta forma, a previsão de curto prazo é uma importante fonte de informações quando observações indiretas da atmosfera são utilizadas, pois este tipo de informação requer a construção de operadores não lineares os quais dependem da informação representada pela previsão de curto prazo (Daley, 1993; Bannister, 2008).

Técnicas variacionais de assimilação de dados são bastante populares nos centros operacionais de PNT. Embora estes centros estejam migrando suas suítes operacionais para sistemas que utilizam algum tipo de técnica por conjunto (eg., um conjunto de realizações dos sistemas 3DVar ou 4DVar; um conjunto de realizações do sistema Filtro de Kalman ou uma combinação de ambos), os sistemas variacionais apresentam a vantagem do fácil tratamento das observações e dos seus operadores, além de permitir o controle de solu-

ções numéricas artificiais como a umidade negativa (proveniente do modelo de previsão, eg., Rasch e Williamson, 1991) e a umidade supersaturada (proveniente do sistema de assimilação de dados, eg., Dee e Silva, 2003). Além disso, para que seja possível o adequado tratamento do uso das informações da previsão de curto prazo e das observações, é necessário o uso de matrizes de covariâncias de erros associadas à representação dessas quantidades. Este artigo trata do problema da representação dos erros associados à modelagem, através da matriz de covariâncias dos erros de previsão (matriz **B**).

Covariâncias dos erros de previsão representados na matriz **B**, tem a importante propriedade de contabilizar a contribuição das quantidades observadas correspondentes ao modelo interpoladas no ponto das observações, e como elas serão transformadas em incrementos de análise e espalhadas nas direções horizontal e vertical. Estas propriedades são atingidas através de estruturas específicas que modulam a forma como o incremento de análise é aplicado. Nesse contexto, montar uma matriz e covariâncias completa considerando todas as variáveis de estado e todos os graus de liberdade associados (eg., número de pontos de grade, níveis verticais etc), resultaria em uma matriz impossível de ser armazenada. Consequentemente, há a necessidade de se modelar ou parametrizar as quantidades de interesse de forma que a representação e a aplicação das covariâncias da matriz **B** possam ser feitas. Uma forma bastante comum de se parametrizar as covariâncias é exprimi-las por meio de variâncias e correlações. No caso escalar, a correlação representa a razão entre a covariância de duas quantidades escalares e as variâncias de cada uma dessas quantidades. Com efeito, sendo a matriz **B** uma matriz de covariâncias, ela pode ser representada em termos destas quantidades. Esta decomposição é um tipo de parametrização e ajuda a amenizar o problema da representação de uma matriz muito grande, permitindo também que termos extras sejam inclusos com o objetivo de que relações físicas desejáveis como o balanço geostrófico sejam respeitadas e as variações diárias do fluxo atmosférico ou os “erros do dia” Kalnay et al. (1997), sejam - em algum grau, representadas.

No CPTEC, desde o início das operações em assimilação de dados com o sistema *Grid-point Statistical Interpolation* (GSI, Wu et al., 2002; Kleist et al., 2009) utilizando o 3DVar como esquema de assimilação de dados, foi utilizada uma matriz de covariâncias calculada com base no modelo *Global Forecast System* (GFS), e um esforço tem sido feito para a obtenção de uma nova matriz de covariâncias específica para o Brazilian Atmospheric Model CPTEC/INPE (BAM-CPTEC/INPE, Figueroa et al., 2016). Com o objetivo de verificar quais as principais diferenças entre as duas matrizes (a matriz original calculada com base no modelo GFS e aquela calculada com base no modelo BAM-CPTEC/INPE), este artigo apresenta a metodologia empregada no cálculo, as estruturas da matriz calculada, as diferenças principais e o incremento de análise resultante através da assimilação de uma única observação.

O artigo está organizado da seguinte forma: na seção 2 são apresentados os materiais e métodos utilizados na pesquisa. Na seção 3, os resultados e a discussão e na seção 4, as conclusões.

2 Material e Métodos

2.1 Sistema G3DVAR

G3DVAR é o nome dado ao acoplamento entre a versão tridimensional do sistema de assimilação de dados GSI (Wu et al., 2002; Kleist et al., 2009) e o modelo atmosférico BAM-CPTEC/INPE (Figueroa et al., 2016). Este sistema esteve em operação no CPTEC entre Janeiro de 2013 e Dezembro de 2015. As análises do sistema eram geradas em uma grade tridimensional de 900 x 450 x 64 pontos e depois eram pré-processadas onde as quantidades analisadas eram decompostas em coeficientes espectrais com truncamento na onda zonal de número 299 e 64 níveis sigma na vertical (TQ0299L064, com ~45 km de resolução horizontal próximo à linha do Equador). Apesar das análises serem escritas em uma grade correspondente à grade da previsão pós-processada do modelo, a matriz de covariâncias utilizada proveniente do modelo GFS do NCEP, era discretizada em uma grade de 768 x 386 pontos e 64 níveis na vertical. Esta foi uma limitação imposta ao sistema e que impedia alteração da resolução horizontal e limitava as análises em 64 níveis verticais. Um dos requerimentos do GSI é que a análise seja calculada com as mesmas dimensões de grade que a matriz de covariâncias. Isso se faz necessário porque o cálculo do incremento de análise depende da matriz de covariâncias, como será mostrado mais adiante. A solução para esta limitação é o cálculo de uma matriz de covariâncias utilizando os pares de previsões do modelo de interesse e com as dimensões corretas para a grade na qual as análises serão calculadas.

Nas seções a seguir, são apresentados os detalhes das componentes do sistema G3DVAR.

2.2 Modelo Atmosférico Brasileiro do CPTEC/INPE

O Modelo Atmosférico Brasileiro do CPTEC/INPE (o *Brazilian Atmospheric Model - BAM-CPTEC/INPE*) utilizado neste trabalho é a inicial de lançamento em Janeiro de 2016 (versão 0.0). Embora esta não seja uma versão mais atualizada do modelo, alguns detalhes sobre as características principais podem ser obtidos em Kubota (2012) e mais recentemente em Figueroa et al. (2016), onde estão descritos com mais detalhes as opções de dinâmica e física do modelo. Esta versão, embora não tenha sido parte integrante da versão 1.1.3 do sistema G3DVAR, assemelha-se a versão operacional utilizada no G3DVAR pelas opções escolhidas em sua configuração. As principais opções dos pacotes de dinâmica e física do modelo numérico utilizado, estão sumarizadas na Tabela 2.1.

Tabela 2.1 - Principais opções e configurações do BAM-CPTEC/INPE utilizado nos experimentos propostos.

Dinâmica		Física	
Resolução	TQ0299L064	Radiação de Onda Longa	Harshvardhan
Passo de Integração	200 segundos	Radiação de Onda Curta	Clirad
Transporte (vento)	Euleriano	Convec. Prof./Fech.	Grell
Transporte (massa)	Semi-Lagrangeano	Convecção Rasa	Tiedke
Inicialização	Diabática	Topo da Camada Limite	Hostlag e Boville
Restart	Superfície	Base da Camada Limite	Mellor e Yamada
Conservação de Massa	$\ln(p)$	Modelo de Superfície	SSiB

2.3 Gridpoint Statistical Interpolation - GSI

Na assimilação tridimensional variacional (3DVar), assume-se que a distribuição dos erros de modelagem é Gaussiana, ou seja, assume-se que média da distribuição dos erros das variáveis de estado seja zero e que o desvio padrão é característico do erro da observação analisada. Esta é a suposição mais largamente utilizada a sua especificação é fundamental para o sucesso do processo de assimilação de dados variacional (Lorenc, 1986).

No sistema 3DVar do GSI, a função custo representada (de forma simplificada) pela Eq. 2.1:

$$J(x) = \underbrace{\frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b)}_{J_b} + \underbrace{\frac{1}{2}(y - \mathbf{H}x)^T \mathbf{R}^{-1}(y - \mathbf{H}x)}_{J_o} \quad (2.1)$$

onde:

- x é o vetor de estado a ser analisado (dimensão: $n \times 1$);
- x_b é o vetor de estado previsto ($n \times 1$);
- $x - x_b$ é o vetor incremento de análise ($n \times 1$);
- \mathbf{B} é a matriz de covariâncias dos erros de previsão ($n \times n$);
- y é o vetor de estado observado ($p \times 1$);
- $y - \mathbf{H}x$ é o vetor inovação ou incremento de observação ($p \times n$);
- \mathbf{H} é o operador observação (responsável por interpolar as variáveis de estado para os respectivos pontos de observação);

- \mathbf{R} é a matriz de covariâncias dos erros de observação ($p \times p$);
- J_b representa o termo da função custo referente à previsão de curto prazo;
- J_o representa o termo da função custo referente às observações.

A função custo representada pela Eq. 2.1 é quadrática e o resultado da sua minimização iterativa pode ser interpretada como uma medida da redução da distância geométrica entre a posição da variável de controle (ie., a variável em torno da qual é realizada a minimização da função custo) e o ponto de mínimo global da função.

A análise 3DVar calculada pelo GSI está representada na Eq. 2.2. A análise é calculada a partir do mínimo da função custo representada pela Eq. 2.1, quando $\nabla J = 0$ implicando em $x = x_a$.

$$x_a = x_b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} [(\mathbf{H}^T \mathbf{R}^{-1})(y - \mathbf{H}x_b)] \quad (2.2)$$

Na Eq. 2.2 tem-se que o estado da análise x_a é igual ao estado da previsão de curto prazo (doravante denominado *background* para indicar que esta previsão não é apenas um “chute inicial”, mas também a informação do fluxo atmosférico utilizada na correção cíclica com as observações) x_b mais a contribuição da inovação trazida pelas observações $(y - \mathbf{H}x_b)$, ponderado pela razão entre as matrizes de covariâncias dos erros de observação (\mathbf{R}) e modelagem (\mathbf{B}). A presença da matriz \mathbf{H}^T aplicada em \mathbf{R} , indica que as operações com as matrizes são realizadas em ponto de grade, ie., \mathbf{R}^{-1} é interpolada para o espaço do modelo.

Tipicamente, em aplicações globais, o GSI analisa as seguintes variáveis: a função de corrente, a parte desbalanceada da velocidade potencial, a parte desbalanceada da temperatura virtual, a parte desbalanceada da pressão em superfície e a umidade relativa normalizada.

2.4 Matriz de Covariâncias dos Erros de Previsão (\mathbf{B})

A Matriz de Covariâncias dos erros de Previsão (\mathbf{B}) tem o papel de filtrar e propagar espacialmente a informação das observações (Berre et al., 2013), contribuindo para a transformação de parte das inovações $(y - \mathbf{H}x)$ em incrementos de análise $(x - x_b)$. Segundo Fisher (2003), para apenas uma observação convencional em ponto de grade, o operador observação linear da Eq. 2.1, pode ser escrito como:

$$\mathbf{H} = [0, \dots, 0, 1, 0, \dots, 0] \quad (2.3)$$

O gradiente da função custo, é então calculado aplicando-se o operador gradiente (∇J) diretamente na Eq. 2.1 e igualando-se a expressão resultante a zero para se determinar o mínimo da função (Eq. 2.4). Observe que neste caso, não se está considerando a hipótese de linearização do operador observação (ie., $y - \mathbf{H}x + \mathbf{H}(x - x_b)$), o que acarretaria no peso obtido na expressão do lado direito da Eq. 2.2. Resolvendo a equação resultante para o incremento de análise $x - x_b$ (Eq. 2.6), obtemos:

$$\nabla J(x) = \mathbf{B}^{-1}(x - x_b) + \mathbf{H}^T \mathbf{R}^{-1}(y - \mathbf{H}x) = 0 \quad (2.4)$$

$$x - x_b = \mathbf{B}\mathbf{H}^T \mathbf{R}^{-1}(y - \mathbf{H}x) \quad (2.5)$$

Como temos apenas uma única observação a ser analisada, logo o termo $\mathbf{R}^{-1}(y - \mathbf{H}x)$ é apenas um escalar, pois \mathbf{R} será um escalar e y e x também. Além disso, \mathbf{H} irá apenas interpolar x para o ponto de grade onde y se situa de forma que seja possível realizar a diferença $y - \mathbf{H}x$. Portanto:

$$\delta x \propto \mathbf{B}\mathbf{H}^T \quad (2.6)$$

Ou seja, o incremento de análise δx é proporcional às colunas da matriz de covariâncias \mathbf{B} .

Na prática, os sistemas operacionais de assimilação de dados são implementados com diferentes técnicas que minimizam o custo computacional envolvido e facilitam o tratamento das matrizes. No caso da matriz de covariâncias dos erros de previsão, a minimização da função custo envolve a inversão das matrizes \mathbf{B} e \mathbf{R} . Devido às dimensões de \mathbf{B} , esta operação pode tornar-se muito custosa ou praticamente impossível, dependendo do número de graus de liberdade do sistema. Por esta razão, algumas implementações permitem o pré-condicionamento da função custo de forma que a sua minimização seja feita com base em uma mudança de variáveis (por exemplo, $z = \mathbf{B}^{-1}\delta x$, onde δx é o incremento de análise $x - x_b$ (Derber e Rosati, 1989)). Este procedimento permite, por exemplo, que a minimização da função custo seja atingida mais rapidamente e sem a necessidade de inversão da matriz \mathbf{B} , mas com o custo extra de se calcular $\nabla_z J$ além de $\nabla_{\delta x} J$. Além disso, as informações armazenadas na matriz de covariâncias não represen-

tam simplesmente apenas as informações de variância e covariância de cada variável. Ao invés disso, são armazenadas informações que são utilizadas durante o procedimento de minimização da função custo, tais como, comprimentos de escala e relações de balanço entre as variáveis de controle.

O cálculo da matriz de covariâncias para o sistema G3DVAR, incluiu a utilização de 1460 pares de previsões de 48 e 24 horas válidos para 1 ano, no formato espectral do BAM-CPTEC/INPE na resolução TQ0299L064. Os pares de previsões foram organizados na forma a seguir: considerando-se o dia 2014010100, o primeiro par de previsões válido foi gerado com as análises dos dias 2013123100 (válido para uma previsão de 24 horas) e 2013123000 (válido para uma previsão de 48 horas). A metodologia para se calcular as amplitudes (representando as variâncias e as covariâncias) utilizando-se pares de previsão é baseada no método denominado *National Modeling Center* (NMC) de Parrish e Derber (1992). Este método preconiza que a correlação espacial dos erros do modelo de previsão são semelhantes à correlação espacial das diferenças entre as previsões de 48 e 24 horas.

Este método foi empregado para o cálculo da matriz de covariâncias pela conveniência de já se ter os pares de previsões prontos (obtidos a partir das realizações operacionais do G3DVAR). Apesar disto, pode-se discutir sobre a hipótese de que os pares de previsões de 48 e 24 horas são mais adequados ou não. Em tese, se o objetivo de se obter uma medida da variância do erro associado à modelagem - e principalmente, como a dinâmica e a física podem ser representadas neste contexto, é plausível que ao invés de se utilizar dois pares de previsão, se utilize uma análise e uma previsão. Outro aspecto que também pode ser discutido, é o tempo necessário para o cálculo das covariâncias. A princípio, o tempo mínimo para que estas covariâncias representem qualitativamente as estruturas calculadas a partir da utilização de um ano completo de pares de previsões pode depender de cada variável.

O algoritmo para o cálculo da matriz de covariâncias, possui as seguintes etapas:

- Leitura e organização dos pares de previsões de 48 e 24 horas;
- Remoção de viés (em toda a coluna vertical);
- Cálculo das matrizes de balanço que permitirão as transformações entre função de corrente (ψ) e as componentes desbalanceadas de velocidade potencial (χ), pressão em superfície (p) e temperatura (T);
- Cálculo das variâncias dos erros de cada uma das variáveis de controle (ψ , χ , q , oz , cw , p);
- Cálculo dos comprimentos de correlação verticais (em unidades inversas em

ponto de grade);

- Cálculo dos comprimentos de correlação horizontais (em km).

Se todas as amplitudes das variáveis de estado pudessem ser estimadas, então a matriz de covariâncias dos erros de previsão poderia ser caracterizada de forma ideal, como sendo uma matriz do tipo bloco onde estão organizadas as estruturas de variâncias na diagonal principal e as estruturas de covariâncias fora da diagonal principal (Figura 2.1). A matriz de covariâncias descreve, portanto, as relações univariadas e multivariadas (no tempo e no espaço) entre as diferentes variáveis do sistema (Petrie, 2012). Ao contrário da matriz \mathbf{R} que é diagonal (ie., podem apenas ser definidos as variâncias dos erros das observações), em \mathbf{B} são os elementos fora da diagonal principal quem tem o papel de espalhar as informações da variância para fora da diagonal principal e entre as diferentes variáveis envolvidas na assimilação de dados.

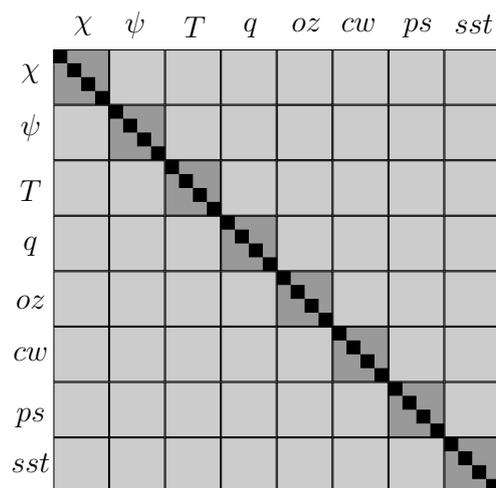


Figura 2.1 - Representação idealizada da matriz \mathbf{B} : na diagonal principal (quadrados pretos), estão os elementos de variância de cada variável de estado; em cinza escuro, estão os elementos de autocovariância (ie., variância da mesma variável em pontos de grade diferentes) e em cinza claro, os elementos de covariância multivariados (Adaptado de Petrie, 2012).

Na prática, a matriz de covariâncias do GSI possui as seguintes estruturas: amplitudes (representadas pelos desvios-padrão, variâncias e covariâncias), distâncias (representadas pelos comprimentos de escalas horizontais e verticais) e matrizes de projeção de balanço (representadas por coeficientes de regressão). Nesta versão da matriz de covariâncias utilizada pelo G3DVAR as amplitudes são representadas por médias zonais (as quais podem ser ponderadas durante o processo de minimização da função custo), variando apenas

nas latitudes e na vertical. As partes desbalanceadas das variáveis velocidade potencial, temperatura virtual e pressão em superfície, possuem matrizes de projeção (\mathbf{G} e \mathbf{W} nas Eqs. 2.7 e 2.8, respectivamente) que são responsáveis por projetar o incremento de análise da função de corrente no perfil vertical da parte balanceada do incremento de cada uma destas variáveis. Para a velocidade potencial, uma matriz de correlação (\mathbf{C} na Eq. 2.9) é utilizada para contabilizar a correlação positiva entre divergência e vorticidade.

$$T_b = \mathbf{G}\phi \quad (2.7)$$

$$P_b = \mathbf{W}\phi \quad (2.8)$$

$$\chi_b = \mathbf{C}\phi \quad (2.9)$$

Esta é a forma utilizada pelo GSI para aplicar os incrementos de análise de forma filtrada, ie., decompõe-se as variáveis de controle em uma parte balanceada (referente à parte lenta do fluxo atmosférico) e em uma parte não balanceada (referente à parte rápida e ruidosa do fluxo atmosférico) e então os incrementos de análise são realizados apenas na parte não balanceada (a qual deve ser ajustada).

No GSI, a matriz de covariância não é explicitamente construída (como mostrado na Figura 2.1) e, ao invés disso, filtros recursivos (Purser et al., 2003a; Purser et al., 2003b) são utilizados para realizar a aplicação das covariâncias. Segundo Wu et al. (2002), a matriz \mathbf{B} é então decomposta da seguinte forma:

$$\mathbf{B} = \mathbf{B}_z(V^1 \mathbf{B}_x^1 \mathbf{B}_y^1 \mathbf{B}_x^1 \mathbf{B}_y^1 V^1 + V^2 \mathbf{B}_x^2 \mathbf{B}_y^2 \mathbf{B}_x^2 \mathbf{B}_y^2 V^2) \mathbf{B}_z \quad (2.10)$$

onde:

- V^1 e V^2 são os desvios-padrão de cada variável de controle (obtidos a partir da própria matriz de covariâncias calculada);
- \mathbf{B}_x , \mathbf{B}_y e \mathbf{B}_z representam a aplicação dos filtros recursivos nas direções x (direção Oeste-Leste), y (direção Sul-Norte) e z (na vertical);
- \mathbf{B}^1 e \mathbf{B}^2 representam a aplicação do filtro recursivo nas escalas horizontais indicadas (direções x e y).

O filtro recursivo utilizado para a aplicação das covariâncias no GSI tem também a importante propriedade de ajustar as amplitudes calculadas de forma que seu aspecto possua algum grau de anisotropia (ie., variação do aspecto representado em todas as direções) permitindo que as covariâncias se adaptem ao fluxo representado pelo *background*. Os parâmetros de escala horizontais utilizados pelo filtro recursivo neste processo, são tabelados e uma discussão sobre a sua determinação pode ser encontrada em Wu et al. (2002).

Um outro aspecto tratado durante o cálculo da matriz de covariância é o viés do modelo numérico. O viés (β^t) é uma medida do erro sistemático do modelo de PNT e pode ser representada pelo modelo (ω^f) subtraído do erro real (ω^t) e de um erro aleatório, cuja esperança é zero, ie., $\langle \varepsilon^t \rangle = 0$ (Chepurin et al., 2005). A remoção de viés (Eq. 2.11) é realizada com a finalidade de se eliminar os erros mais grosseiros, e é aplicada em toda a coluna vertical do modelo. Este procedimento é realizado para as seguintes variáveis: parte balanceada da função de corrente, parte desbalanceada da velocidade potencial, temperatura virtual, umidade relativa normalizada, razão de mistura de ozônio, razão de mistura de condensação de água em nuvens e pressão em superfície.

$$\beta^t = \omega^f - \omega^t - \varepsilon^t \quad (2.11)$$

Os comprimentos de escala horizontais são importantes estruturas da matriz \mathbf{B} e são dependentes dos valores de variâncias calculados - Eq. 2.12, onde V_ψ e V_ζ , representam as variâncias da função de corrente e da vorticidade, respectivamente. Os comprimentos de escala verticais podem ser estimados através da correlação vertical de cada variável de controle (Wu et al., 2002).

$$L = \left(\frac{8V_\psi}{V_\zeta} \right)^{\frac{1}{4}} \quad (2.12)$$

Os comprimentos de escala descrevem o quão distantes as covariâncias entre os erros das variáveis de controle podem ser relacionadas espacialmente. É através deles que os elementos da diagonal principal da matriz \mathbf{B} são transportados até uma determinada distância para fora da diagonal principal de \mathbf{B} . Portanto, a variação do comprimento de escala descreve o modo como a covariância varia com a distância. Além disso, é através dos comprimentos de escala que os filtros recursivos são aplicados para a modelagem das matrizes de correlação do GSI. Pannekoucke et al. (2008) apresentam uma revisão sobre este assunto comparando diversas formulações para os comprimentos de escala e quais os seus impactos em situações idealizadas.

2.5 Ciclo de Assimilação de Dados Global

O ciclo de assimilação de dados do sistema G3DVAR é realizado nos quatro horários sinóticos padrão (00, 06, 12 e 18Z). O sistema de assimilação de dados GSI utiliza a técnica *First Guess at Appropriate Time* (FGAT) em que o termo de inovação $y - \mathbf{H}x$ é calculado utilizando-se o horário das observações como referência (Lee e Barker, 2005). Dessa forma, o GSI é capaz de utilizar as observações na correção do *background* em horários mais próximos dos horários de coleta das observações, ao invés de considerar que todas as observações são válidas para o horário central na janela de assimilação de dados. Para a parte atmosférica, o GSI calcula os incrementos de análise das variáveis de controle com base no incremento da função de corrente. Para a parte da superfície, embora a temperatura da superfície continental seja atualizada, na implementação atual do ciclo de assimilação de dados, esta parte não é considerada quando o modelo atmosférico é integrado no tempo e as variáveis de estado da superfície são lidas dos arquivos de *restart* do modelo. Dentro do ciclo de assimilação de dados, os coeficientes espectrais representando as quantidades correspondentes dos campos prognósticos do modelo são recompostas para ponto de grade, de forma que a assimilação de dados seja feita no espaço das observações. Ao final deste processo, as quantidades atualizadas da análise são decompostas novamente em coeficientes espectrais.

2.6 Experimentos

Para se verificar as características principais da nova matriz calculada com base no histórico de previsões do BAM-CPTEC/INPE, foi feita uma descrição comparativa da nova matriz e esta foi comparada com a matriz do GFS/NCEP. Além disso foram realizados dois experimentos com a assimilação de uma única observação sintética (vento horizontal e temperatura do ar). Nestes experimentos, os seguintes parâmetros puderam ser controlados: a magnitude do erro da observação, a magnitude da inovação, a posição horizontal (*lat X lon*) e vertical (nível de pressão). Os experimentos com observação única permitem que se obtenha uma visão geral de como o incremento de análise é espalhado na horizontal e na vertical de acordo com os diferentes parâmetros disponíveis e como ele se adapta ao fluxo atmosférico representado pelo *background*.

3 Resultados e Discussão

Nesta seção são apresentadas as características principais da matriz de covariâncias calculada, comparando-a com a matriz de referência do GFS/NCEP. As estruturas das amplitudes das variáveis de controle bem como os valores máximos e mínimos são apresentados e discutidos. Finalmente, a aplicação da matriz calculada é testada através de um experimento com observação única com o objetivo de se avaliar o aspecto dos incrementos de análises produzidos.

3.1 Caracterização da Matriz de Covariâncias Calculada

Para se verificar o aspecto das amplitudes e distâncias calculadas a partir dos pares de previsões do BAM-CPTEC/INPE, a matriz de covariâncias calculada considera todos os horários sinóticos padrão (00, 06, 12 e 18Z) para o período de 1 ano, totalizando 1460 pares de previsões. As matrizes comparadas (“B NCEP” e “B CPTEC”) foram utilizadas nos experimentos descritos na Seção 2.6 e os valores máximos e mínimos de suas amplitudes e comprimentos de escala das variáveis função de corrente, velocidade potencial, temperatura, umidade, pressão em superfície, encontram-se nas Tabelas 3.1 e 3.2.

As dimensões horizontais das matrizes calculadas com base nos pares de previsões do BAM-CPTEC/INPE, possuem as mesmas dimensões que a matriz do NCEP, ie., 768 x 386. Isso foi feito para que fosse possível realizar a comparação com a matriz operacional utilizada. Além disso, uma limitação na representação da razão de mistura do ozônio pelo BAM-CPTEC/INPE, impedia que as amplitudes e comprimentos calculados pudessem ser efetivamente utilizados pelo GSI. Como forma de contornar este problema, foram especificados os mesmos valores de amplitudes e comprimentos para a matriz do CPTEC.

Em uma comparação quantitativa, as amplitudes (Tabela 3.1) resultantes das variáveis função de corrente e velocidade potencial, mostram que os valores máximos representados na matriz “B CPTEC” excedem em 1 a ordem de magnitude das mesmas quantidades representadas na matriz “B NCEP”. Sendo a amplitude uma representação da variância destas variáveis, a interpretação que se faz é que na matriz “B CPTEC” existe maior variação temporal da função de corrente e da velocidade potencial possivelmente devido à quantidade de pares de previsões utilizado. Para as demais variáveis, as quantidades representadas são diretamente comparáveis. Para as variáveis razão de mistura de ozônio, estas quantidades possuem ordem de magnitude de -7 (máximo e mínimo, respectivamente) para ambas as matrizes e para a razão de mistura de água líquida em nuvens, possuem ordem de magnitude de -4 e -7 (máximo e mínimo, respectivamente), para ambas as matrizes. Para a temperatura da superfície do mar, estes valores possuem ordem de magnitude de -1 (máximo e mínimo, respectivamente) para ambas as matrizes, indicando

uma variação muito pequena entre os dois extremos.

Tabela 3.1 - Valores máximos e mínimos das amplitudes e dos comprimentos de escala horizontais das matrizes de covariâncias “B CPTEC” e “B NCEP”.

	Amplitudes				Comprimentos de Escala Horizontais			
	B NCEP		B CPTEC		B NCEP		B CPTEC	
	Mín	Máx	Mín	Máx	Mín	Máx	Mín	Máx
ψ	$4,3 \times 10^5$	$7,99 \times 10^6$	$4,6 \times 10^5$	$2,8 \times 10^7$	$2,2 \times 10^5$	$1,3 \times 10^6$	$4,55 \times 10^5$	$1,75 \times 10^6$
χ	0	$3,77 \times 10^6$	0	$1,07 \times 10^7$	0	$1,1 \times 10^6$	0	$1,82 \times 10^6$
t	0	2,23	0	4,04	0	$3,1 \times 10^5$	0	$5,9 \times 10^5$
q	$-4,06 \times 10^{-7}$	0,41	$-2,15 \times 10^{-7}$	0,29	0	$7,9 \times 10^6$	0	$2,22 \times 10^7$
ps	$4,88 \times 10^{-2}$	$1,01 \times 10^{-1}$	$6,35 \times 10^{-2}$	$1,48 \times 10^{-1}$	$1,19 \times 10^5$	$1,89 \times 10^5$	$1,83 \times 10^5$	$4,18 \times 10^5$

Os valores máximos e mínimos dos comprimentos de escala horizontais (Tabela 3.1) da matriz “B CPTEC” calculada, indicam maior variação espacial da variância da umidade, em relação à matriz “B NCEP”. Esta diferença corresponde a uma ordem de magnitude em relação aos valores máximos de ambas as matrizes. Para as demais variáveis apresentadas, os valores máximos e mínimos são compatíveis. Para as variáveis razão de mistura de ozônio e razão de mistura de água líquida em nuvens, a ordem de magnitude dos valores máximos e mínimos nas matrizes é de 6 e 0, respectivamente. A escala horizontal da temperatura da superfície do mar possui ordem de magnitude máxima e mínima igual a 2.

As escalas verticais de variação das amplitudes (Tabela 3.2) é maior na função de corrente representada para a matriz “B CPTEC” calculada, sendo uma ordem de magnitude maior do que está representado no valor máximo da função de corrente da matriz “B NCEP”. Para as demais variáveis, as variações e os valores máximos e mínimos possuem a mesma ordem de magnitude.

Tabela 3.2 - Valores máximos e mínimos dos comprimentos de escala verticais das versões das matrizes de covariâncias “B CPTEC” e “B NCEP”.

	B NCEP		B CPTEC	
	Mín	Máx	Mín	Máx
ψ	$8,3 \times 10^{-2}$	$9,1 \times 10^{-1}$	$8,41 \times 10^{-2}$	1,27
χ	0	1,22	0	1,5
t	0	1,75	0	1,67
q	0	$1,3 \times 10^6$	0	$1,75 \times 10^6$

Para comparação qualitativa das matrizes consideradas, foi utilizada a matriz pré-calculada de covariâncias do NCEP, utilizada na versão operacional do sistema G3DVAR no CPTEC. Em todos os casos, a grade utilizada possui dimensões 768 x 386 (*lat X lon*) na horizontal e 64 níveis sigma na vertical.

A Figura 3.1 mostra a distribuição vertical e latitudinal das amplitudes (variâncias) da função de corrente, velocidade potencial, temperatura do ar e umidade específica provenientes das matrizes de covariâncias do NCEP (“B NCEP”), da matriz de covariâncias do CPTEC (“B CPTEC”).

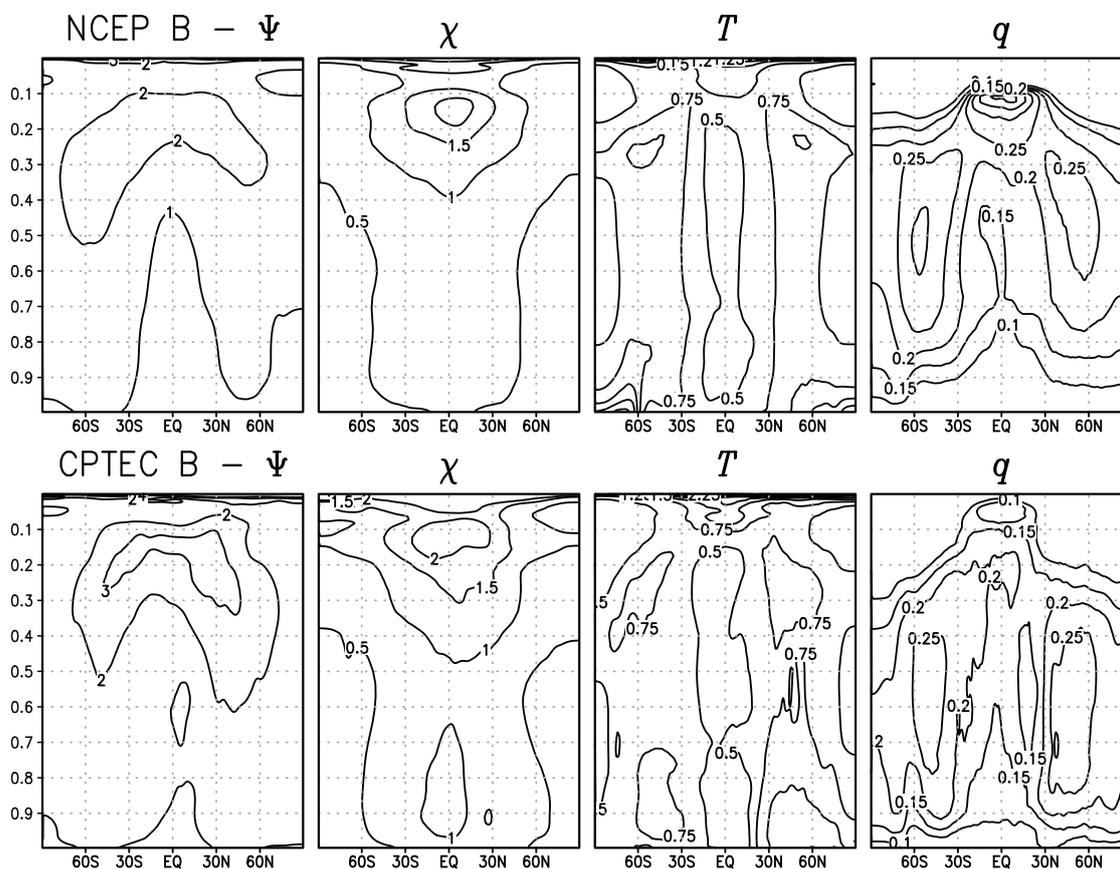


Figura 3.1 - Distribuição das amplitudes da Função de Corrente (ψ , $\times 10^6$), Velocidade Potencial (χ , $\times 10^6$), Temperatura do Ar (T) e Umidade Específica (q) ao longo das latitudes (em x) e níveis sigma (em y), calculadas utilizando-se o método NMC, para as matrizes **B** do NCEP (primeira linha) e do CPTEC (segunda linha).

Comparando-se “B NCEP” e “B CPTEC” (as duas matrizes contendo os pares de previsões de todos os horários sinóticos padrão), observa-se que na matriz “B CPTEC” as amplitudes das variâncias dos erros da parte não divergente do fluxo atmosférico (representadas através da função de corrente, primeira coluna da Figura 3.1) são mais concentradas e menos suaves do que as amplitudes observadas na matriz “B NCEP” e especialmente assimétricas. Esta característica mostra que entre os níveis sigma 1 e 0,7 (aproximadamente entre 1024 e 716,8 hPa) na matriz “B NCEP” a distribuição das amplitudes estende-se mais a partir da região tropical e sobre o Hemisfério Norte. O contrário é observado na

matriz “B CPTEC”, porém com menor extensão e intensidade, sobretudo entre os níveis sigma 1 e 0,9 (aproximadamente entre 1024 e 921,6 hPa). Estas diferenças podem estar relacionadas à forma como os pares de previsões foram gerados: para o cálculo da matriz de covariâncias “B CPTEC”, foram utilizados pares de previsões de 48 e 24 horas provenientes do ciclo de assimilação de dados do G3DVAR. Conseqüentemente, a variância do erro sobre o Hemisfério Sul, sobretudo sobre a América do Sul (entre 0 e 60S) pode ser diferente, sobretudo dependendo da forma como o controle de qualidade das observações é realizado. Um outro fator que pode contribuir para as diferenças entre a distribuição vertical das amplitudes, é a diferença entre as topografias dos modelos que foram utilizadas para o cálculo das matrizes de covariâncias. A matriz “B NCEP”, embora com uma grade horizontal de dimensões menores do que 900x450 (referente ao TQ0299L064), especula-se que as quantidades ali representadas tenham sido truncadas de uma resolução espectral maior (eg., TQ0574L64, cuja grade tridimensional possui dimensões de 1760 x 880 x 64). Este truncamento para uma representação dos coeficientes espectrais da topografia com menos ondas, pode gerar diferenças na representação da topografia em ponto de grade. Uma comparação entre as topografias das análises do NCEP e do sistema G3DVAR, mostra que as diferenças podem variar de alguns poucos metros até dois quilômetros (figura não mostrada).

A distribuição das amplitudes da parte divergente do fluxo atmosférico (representada através da velocidade potencial, segunda coluna da Figura 3.1), assim como para a função de corrente, mostram que sobre a região tropical (entre 30S e 30N), a variância do erro representada na matriz “B CPTEC” é mais suave abrangendo uma área maior do que é representado na matriz “B NCEP”, sobretudo entre os níveis mais baixos do modelo até o nível sigma 0,5 (aproximadamente 512 hPa). Esta diferença pode ser refletida na divergência do vento horizontal em níveis mais próximos à superfície, ocasionando com que os campos de vento horizontal derivados da análise do sistema G3DVAR com a matriz “B CPTEC” sejam mais intensos. Uma das implicações desta diferença, é a produção de mais chuva nas vizinhanças das regiões de maior divergência.

Para a temperatura (terceira coluna da Figura 3.1), a matriz “B CPTEC”, a maior parte da variância dos erros encontra-se concentrada sobre o Hemisfério Norte a partir de 30N. Na matriz “B NCEP”, uma característica interessante pode ser assinalada sobre a latitude de 60N, próximo ao nível sigma 1 (1024 hPa). O padrão da variância do erro da temperatura apresenta uma distribuição simétrica, com amplitude se estendendo até o nível sigma 0,9 (aproximadamente 921,6 hPa). Esta estrutura não é representada pela matriz de covariâncias do CPTEC; porém, a isolinha de valor 0,75 é observada aproximadamente na mesma região com a diferença de que na matriz “B CPTEC”, esta estrutura é fechada e situa-se entre as latitudes 60S e 30S e entre os níveis sigma 1 e 0,7 (aproximadamente entre 1024 e 716,8 hPa). Especula-se, entretanto, que tais diferenças possam estar rela-

cionadas às diferenças de representação dos fluxos de calor mais próximos à superfície, entre os modelos globais do CPTEC e do NCEP. Tal como as diferenças encontradas nas amplitudes da velocidade potencial (as quais, por sua vez, podem estar relacionadas com a precipitação), estas diferenças também podem ser devidas à forma como os modelos prognosticam a temperatura virtual e sensível.

As amplitudes da umidade (quarta coluna da Figura 3.1) representadas na matriz “B CPTEC”, não diferem muito em distribuição ao longo das latitudes, mas a amplitude da variância na região subtropical em ambos os hemisférios é mais definida e menos intensa do que aquela representada na matriz “B NCEP”. Neste caso, a variância do erro da umidade em ambos os hemisférios é simétrica, o que pode ser um indicativo da necessidade de se remover o viés no modelo do CPTEC. Outra característica que pode ser assinalada também, é a concentração das amplitudes no nível sigma 0,1 da matriz “B NCEP” sobre a região Equatorial. Apesar de se ter utilizado a quantidade umidade relativa normalizada como componente úmida no GSI (esta opção em contraste à pseudo umidade relativa, permite que a umidade varie de acordo com as variações na pressão em superfície), Wu et al. (2002) mostram que a mesma seção para a quantidade pseudo umidade relativa também apresenta concentrações na mesma região. Além disso, as isolinhas mais próximas à superfície (até o nível sigma 0,8, aproximadamente 819,2 hPa), apresentam valores com quase o dobro de amplitude, se comparados com a mesma região na matriz “B NCEP”. De outra forma, pode indicar também alguma característica da física do modelo numérico na representação dos processos úmidos, como por exemplo, a forma da representação da precipitação convectiva e de larga escala, a representação da radiação de onda curta ou mesmo a difusão horizontal e vertical.

A Figura 3.2 apresenta a distribuição latitudinal das médias zonais da variância do erro da pressão em superfície. As amplitudes da matriz “B NCEP” apresentam uma distribuição simétrica, com dois picos sobre as latitudes de 60S e 60N. Na matriz do CPTEC, a distribuição das amplitudes é assimétrica, sendo que os picos das variâncias dos erros são maiores sobre a latitude de 60N do que sobre a latitude de 60S. A mesma amplitude apresentada em Wu et al. (2002), mostra que os picos da variância dos erros situam-se em torno destas latitudes, especialmente sobre o Hemisfério Norte, onde o pico local situa-se entre as latitudes de 30N e 60N.

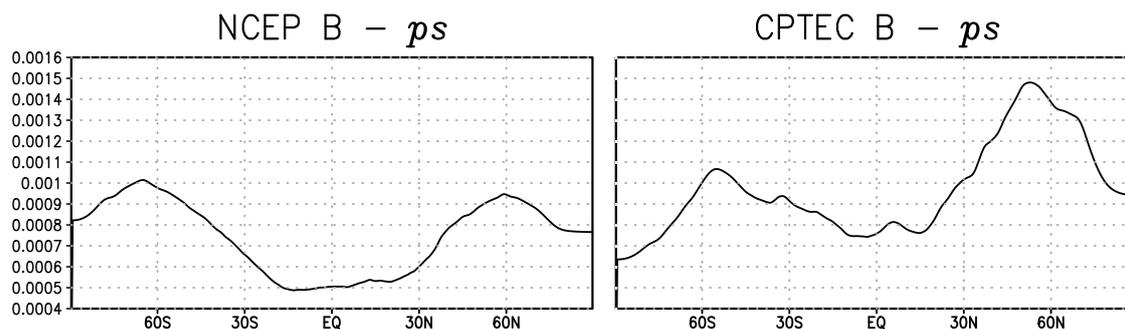


Figura 3.2 - Distribuição das amplitudes da Pressão em Superfície (ps) ao longo das latitudes (em x) e níveis sigma (em y), calculadas utilizando-se o método NMC. A comparação é feita com o NCEP.

O sistema G3DVAR experimentado não foi configurado para assimilar observações de ozônio e a representação desta quantidade com as configurações utilizadas no BAM-CPTEC/INPE, é feita por meio de uma climatologia. As amplitudes originais calculadas para esta quantidade utilizando-se a climatologia são bastante discrepantes daquelas apresentadas pela matriz “B NCEP”. A razão de mistura de ozônio, a razão de mistura de água condensada em nuvem e a umidade são três tipos de traçadores utilizados pelo modelo de transferência radiativa do GSI para o cálculo dos perfis de temperatura de brilho e dos perfis de radiancias simulados (perfis referentes ao modelo).

Outro fator que também deve ser apontado nas comparações é que, apesar da matriz de covariâncias “B NCEP” ser calculada utilizando-se pares de previsões válidos para 1 ano, não se pode dizer ao certo como estes pares de previsões foram distribuídos ao longo desse período. Wu et al. (2002) mencionam que foram utilizados 49 pares de previsões distribuídos ao longo de um período de 1 ano para o cálculo da matriz de covariâncias do NCEP, enquanto que para a matriz de covariâncias do CPTEC calculada, foram utilizados 1460 pares de previsão, o que ajuda a explicar o grau de suavização encontrado nas estruturas verticais das amplitudes da matriz e também na magnitudes dos valores máximos e mínimos das amplitudes e comprimentos de escala verificados.

3.2 Experimentos com Assimilação de Observação Única

Como forma prática de se verificar como a nova matriz de covariâncias do CPTEC aplica o incremento de análise no sistema G3DVAR, foi feita a inclusão de uma única observação sintética no sistema e avaliado qual o resultado obtido com o incremento de análise calculado. Esta observação sintética representa a simulação de uma observação real em que podem ser controlados dois importantes parâmetros: a magnitude do erro da observação e a magnitude do termo de inovação da função custo (ie., $y - \mathbf{H}x$). A fim de que

seja possível caracterizar o aspecto do incremento de análise resultante, as Figuras 3.3 e 3.4 apresentam uma comparação entre os incrementos de vento e temperatura produzidos pelas matrizes de covariâncias do CPTEC e do NCEP. Neste caso, foi utilizado o mesmo *background* do campo de vento e de temperatura nos testes com as matrizes, de forma que as diferenças apresentadas seja devidas apenas à matriz de covariâncias. Para isso, foram considerados os mesmos parâmetros de configuração que o sistema GSI dispõe para a aplicação da matriz de covariâncias. Como o objetivo deste artigo não é buscar quais são os valores ótimos destes parâmetros, mas apresentar e discutir a nova matriz de covariâncias, os testes empíricos relacionados ao ajuste do incremento de análise serão feitos em um estudo à parte. A Tabela 3.3 apresenta, portanto, os valores dos parâmetros principais utilizados para produzir os incrementos apresentados nas figuras desta seção.

Tabela 3.3 - Parâmetros de configuração utilizados no sistema GSI para a aplicação das matrizes de covariâncias do CPTEC e do NCEP.

Parâmetro	Valor	Descrição
<i>vs</i>	0,7	Fator de escala do comprimento de correlação vertical
<i>hzscl</i>	1,7; 0,8; 0,5	Fator de suavização para as 3 escalas horizontais
<i>hswgt</i>	0,45; 0,3; 0,25	Pesos aplicados a cada escala horizontal
<i>bw</i>	0	Fator no cálculo da covariância do erro
<i>norsp</i>	4	Ordem de suavização horizontal das covariâncias do erro de background
<i>bkg_flowdep</i>	True/False	Flag para usar (True = anisotrópico) ou não (False = isotrópico) a dependência das variâncias do erro do background
<i>bkg_rewgtfct</i>	1,5	Fator usado para reponderar as covariâncias do erro do background

A Figura 3.3, mostra o incremento de análise (previsão subtraída da análise, ou AMF) produzido pela assimilação de uma observação sintética da componente zonal do vento horizontal no nível de pressão 1000 hPa na posição de coordenadas 45N e 180 (*lat X lon*, respectivamente). Além disso, ajustou-se o erro da observação para 1 ms^{-1} , bem como a magnitude da inovação trazida pela observação (previsão subtraída da observação, ou OMF). Na Figura 3.3, são apresentados dois painéis: o painel da esquerda apresenta as seções *lat X lon*, *lat X lev* e *lon X lev* dos incrementos de análise produzidos por uma observação sintética do vento horizontal utilizando a matriz “B CPTEC”. O painel da direita, apresenta os resultados obtidos com a matriz “B NCEP”. Em ambos os painéis são apresentados também os resultados obtidos com a aplicação do filtro recursivo isotrópico e anisotrópico, mostrando como os incrementos de análise podem ser ajustados aos gradientes representados no campo de *background* (opção “*bkg_flowdep*”, indicado na Tabela 3.3). Esta opção é uma importante funcionalidade do sistema GSI, pois ela permite que a matriz de covariâncias seja ajustada com base na tendência temporal impressa nos campos de *background* utilizados para o cálculo da análise.

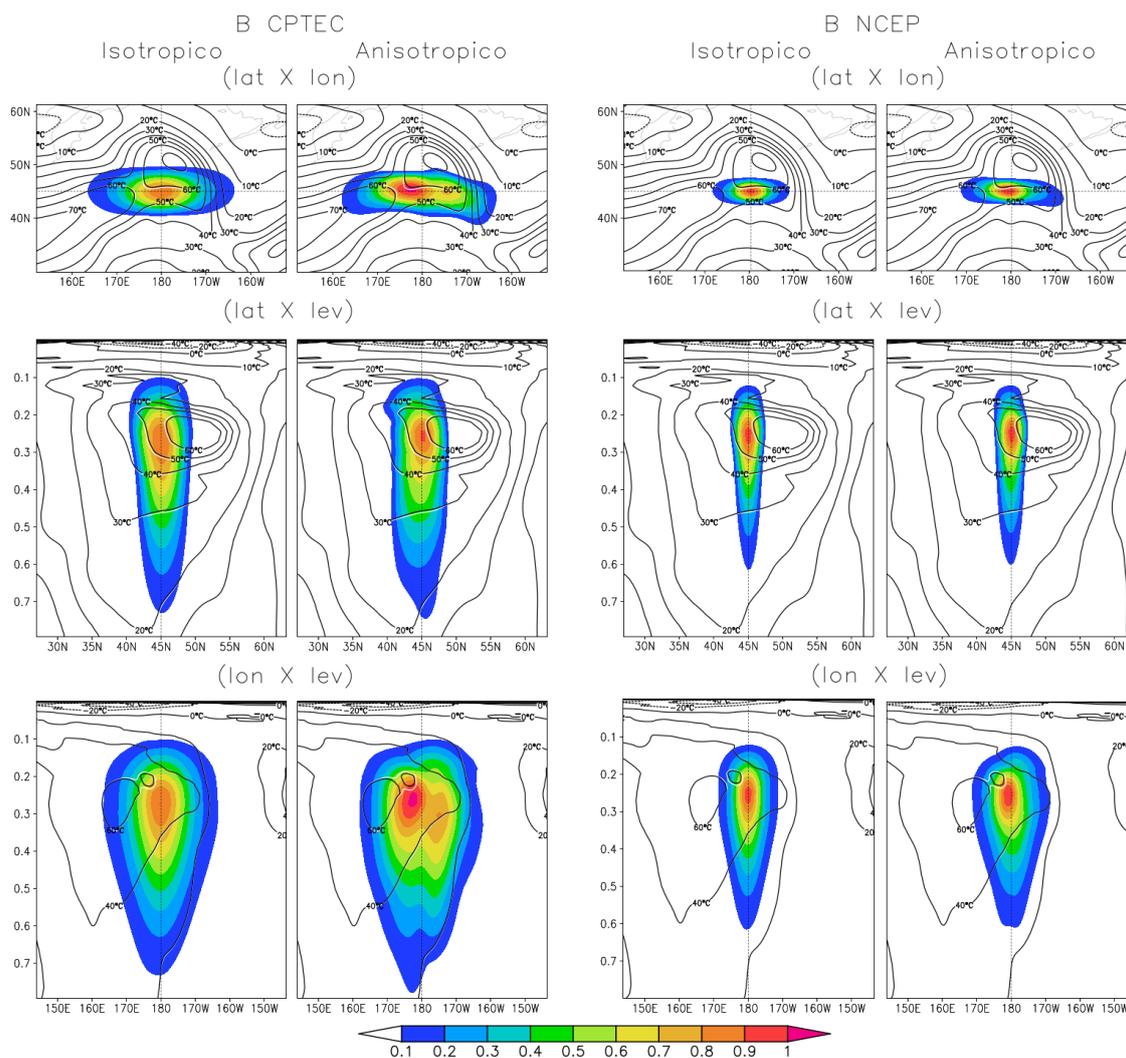


Figura 3.3 - Incrementos da análise (sombreado) do campo do vento horizontal (isolinhas) em 250 hPa, produzidos pelas matrizes de covariâncias do CPTEC e NCEP, respectivamente, onde estão destacadas seções latitudinal, longitudinal e vertical. À esquerda, incremento isotrópico; à direita incremento anisotrópico.

As linhas tracejadas na horizontal e na vertical nas figuras de ambos os painéis estão destacadas para auxiliar na identificação do ponto em que a observação sintética foi posicionada, além de indicar as seções latitudinais, longitudinais e verticais feitas para mostrar os diferentes aspectos do incremento de análise produzido. As isolinhas pretas indicam o *background* da componente zonal do vento (indicado em ms^{-1}) e na região sombreada do incremento de análise, estão destacadas isolinhas brancas representando a análise, como sendo o resultado do ajuste/correção realizada pela observação no campo de vento do *background*. Por inspeção visual, as diferenças principais entre os incrementos produzidos pelas matrizes, mostra que o incremento da matriz “B CPTEC” é mais largo na horizontal e também mais alto na vertical. A seção *lon X lev* dos painéis da esquerda e da direita, mostra como o incremento anisotrópico das matriz se ajustaram à isolinha de 20

ms^{-1} , a partir do nível sigma 0,6 (aproximadamente 600 hPa).

A Figura 3.4 apresenta o incremento de análise produzido pelas matrizes de covariâncias testadas no sistema G3DVAR, assimilando uma observação sintética de temperatura do ar, posicionada no nível de 1000 hPa e com a magnitude de erro e inovação iguais a 1 grau Celsius, respectivamente. A observação foi também posicionada no ponto de coordenadas 45N e 180 (*lat X lon*, respectivamente). De forma geral, para ambos os casos (isotrópico e anisotrópico), o incremento de análise calculado pela matriz de covariâncias do CPTEC (matriz “B CPTEC”), produz também um incremento mais largo, ie., abrangendo uma área de vários graus de latitude e longitude e também na vertical. Isso significa, que o incremento de análise calculado, é espalhado sobre uma região maior ao redor da vizinhança da observação e que a sua influência pode ser transportada para mais longe. Em uma situação física real em que, por exemplo, há a incursão de um sistema frontal e, considerando o caso anisotrópico, a assimilação de dados utilizando uma matriz de covariância com estas características pode trazer um benefício importante para a previsão gerada, porque a análise utilizada poderá possuir um balanço melhor entre observações e *background*. Isso mostra também a importância da seleção das observações, do controle de qualidade e do *thinning* (procedimento realizado no conjunto de dados observacionais com o objetivo de reduzir o custo computacional através da eliminação de dados de observações redundantes ou não confiáveis) utilizados no sistema. Além disso, a intensidade do incremento de análise produzido pela matriz de covariâncias do CPTEC é mais suave (porque se espalha sobre uma região maior), embora seja menos concentrado. Em contraste, a matriz de covariâncias do NCEP, apresenta incrementos muito mais concentrados e conseqüentemente mais localizados.

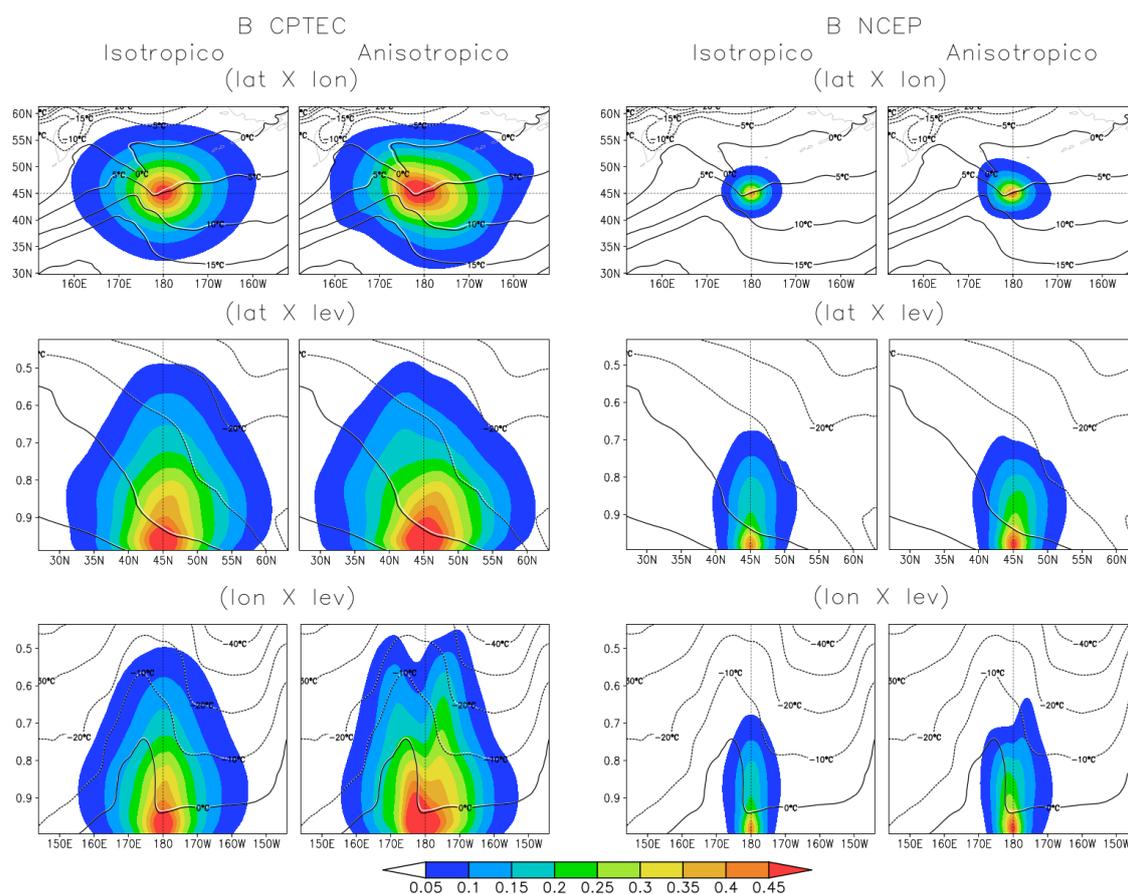


Figura 3.4 - Incrementos da análise (sombreado) do campo de temperatura do ar (isolinhas) em 1000 hPa, produzidos pelas matrizes de covariâncias do CPTEC e NCEP, respectivamente, onde estão destacadas seções latitudinal, longitudinal e vertical. À esquerda, incremento isotrópico; à direita incremento anisotrópico.

As diferenças entre os incrementos de análises produzidos pelas matrizes **B** do CPTEC e do NCEP utilizando os parâmetros indicados na Tabela 3.3, mostram dois aspectos importantes das matrizes de covariâncias: a localização espacial e a intensidade do incremento de análise. A matriz “B NCEP” apresentou incrementos mais concentrados, indicando que o incremento de análise foi restrito à área em que a inovação trazida pelas observações foi calculada, sendo esta mais restrita também à área em que as observações sintéticas (da componente zonal do vento horizontal e da temperatura) foram posicionadas.

No GSI, o campo de vento (componentes zonal e meridional) da análise é calculado a partir dos incrementos da divergência (D) e da vorticidade (ζ). As observações utilizadas para o cálculo desta análise, são dadas em termos das componentes zonal (u) e meridional (v) do vento. O BAM-CPTEC/INPE fornece como campo de *background* para o cálculo da análise do vento horizontal, também a divergência e a vorticidade. Como na matriz de covariâncias estão representadas as quantidades função de corrente (ψ) e velocidade potencial (χ), para o cálculo do incremento de análise do campo de vento, são então

calculadas - através da Eq. 2.9 (para a parte divergente), o incremento da velocidade potencial. No final, a análise é escrita novamente em termos da divergência.

As características observadas de localização e intensidade do incremento de análise podem ser, portanto, explicadas através da verificação da ordem de magnitude do valor máximo da amplitude da função de corrente representado na matriz “B NCEP”, em comparação com “B CPTEC” (vide Tabela 3.1). Sendo o incremento de análise da função de corrente projetado na parte balanceada das variáveis de controle (parte balanceada da temperatura, pressão em superfície e velocidade potencial), maiores inovações desta variável (através das observações do vento) podem acarretar em incrementos de análise mais ou menos intensos de divergência e vorticidade. Então, quando o incremento de análise é calculado, de acordo com a Eq. 2.9, e considerando a relação entre o Laplaciano da velocidade potencial e a divergência (e função de corrente e vorticidade) na Equação de Poisson ($\nabla^2 \chi = D$, $\nabla^2 \psi = \zeta$), mais intenso poderá ser o incremento de análise no campo de vento.

4 Comentários Finais e Conclusões

A matriz de covariâncias dos erros de previsão é uma das mais importantes componentes de um sistema de assimilação de dados. Os desenvolvimentos realizados nesta área mostram uma crescente preocupação na forma como estes erros são especificados e aplicados pelos sistemas de assimilação de dados. Novas técnicas tem surgido, possibilitando que uma matriz estática estacionária seja combinada com uma matriz dinâmica (ie., que possa ser atualizada no tempo), de forma que ambas as informações se complementem e que as variações diárias do fluxo atmosférico representado pelo *background* possam ser incorporadas aos sistemas de assimilação de dados ao se calcular a análise dos modelos.

Neste trabalho foi calculada e testada uma nova matriz de covariâncias dos erros de previsão, aplicada ao sistema de assimilação de dados G3DVAR do CPTEC. A matriz em uso vigente no sistema G3DVAR, foi especificada com base em um outro modelo de PNT (modelo GFS/NCEP), tendo sido aplicada na operação do sistema. Foi mostrada a importância da matriz de covariâncias dentro da estrutura variacional tridimensional, além de uma comparação quantitativa e qualitativa, e quais as principais diferenças entre essa matriz e a nova matriz calculada. Qualitativamente, a matriz do CPTEC apresenta estruturas mais suavizadas devido à quantidade de pares de previsões utilizados (1460 pares da matriz do CPTEC x 49 pares da matriz do NCEP). As diferenças apresentadas nas estruturas podem nos levar a compreender de que maneira a amplitude e as escalas espaciais dos erros de previsão podem influenciar o espalhamento do incremento de análise. Além disso, a matriz de covariâncias traz informações das anomalias do fluxo atmosférico representadas através de variâncias espaciais e temporais. Esta é uma característica que a torna estacionária, evidenciando a necessidade de sua atualização à medida em que os sistemas meteorológicos evoluem no tempo.

Nos experimentos com observação única realizados, verificou-se o aspecto do incremento de análise da nova matriz de covariâncias. Este tipo de experimento é importante e deve ser realizado sempre que algum parâmetro ou configuração é modificado dentro do sistema. Esta é uma das únicas formas de se aferir o impacto direto na realização do sistema de assimilação de dados, frente à quantidade de observações rotineiramente assimiladas. Além disso, foi feita também uma caracterização das estruturas de comprimentos de escala, variâncias da nova matriz, e foi mostrado em comparação com o uso da matriz atual, que as estruturas associadas à variação do erro da previsão estão distribuídas de formas distintas nas matrizes comparadas (“B NCEP” e “B CPTEC”), principalmente com relação à variável de controle função de corrente. Tal fato pode conduzir ao melhor entendimento de como a assimilação de dados pode ser ajustada para as diferentes regiões do globo, principalmente sobre o Hemisfério Sul, onde é notável a menor quantidade de observações de superfície e a quantidade de observações de radiancias sobre os oceanos é

determinante. Também, como forma de verificação, este diagnóstico será útil para trabalhos futuros, em que um ajuste fino deve ser realizado buscando-se a calibração do aspecto do incremento de análise e o melhor aproveitamento das amplitudes das covariâncias. A verificação dos impactos da nova matriz de covariâncias dos erros de previsão, bem como a sua comparação com diferentes versões do modelo de circulação geral da atmosfera do CPTEC, deverão ser realizadas em um estudo separado.

Sendo a matriz de covariâncias uma representação da variabilidade espaço-temporal das estruturas de erro associadas a um modelo de previsão, a descrição desta representação - mediante o uso de um grande número de previsões, pode ser aplicada em estudos envolvendo outras escalas temporais. Neste estudo, foi dado enfoque aos efeitos diretos da atualização da matriz de covariâncias no contexto da determinação da análise de um modelo de previsão numérica de tempo. Por outro lado, modelos ajustados para a previsão de tempo estendido (isto é, até 15 dias) ou mesmo do clima nas escalas sazonal e intra-sazonal, podem também ser utilizados com o objetivo de se melhorar a previsibilidade de fenômenos e oscilações (eg., Madden-Julian) dentro desta escala de tempo.

5 Agradecimentos

Os autores agradecem ao Dr. Ricardo Todling do GMAO/NASA pelo apoio e suporte no desenvolvimento deste trabalho, ao Dr. Daryl Kleist pelas informações sobre estrutura da matriz de covariâncias, ao Dr. João Gerd Zell de Mattos e Msc. Fábio Rodrigues Diniz do CPTEC/INPE pelo auxílio nos testes iniciais com a matriz de covariâncias e aos revisores do manuscrito, cujas contribuições ajudaram a melhorar a qualidade do texto. À CAPES pela concessão da bolsa do programa PDSE processo número BEX 99999.008036/2014-04.

6 Referências Bibliográficas

BANNISTER, R. N. **A Review of Forecast Error Covariance Statistics in Atmospheric Variational Data Assimilation I: Characteristics and Measurements of Forecast Error Covariances.** *Quarterly Journal of the Royal Meteorological Society*, John Wiley & Sons, Ltd., v. 134, n. 637, p. 1951–1970, 2008. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1002/qj.339>>.

BERRE, L.; MONTEIRO, M.; PIRES, C. **An Impact Study of Updating Background Error Covariances in the ALADIN-France Data Assimilation System.** *Journal of Geophysical Research: Atmospheres*, v. 118, n. 19, p. 11,075–11,086, 2013. ISSN 2169-8996. Disponível em: <<http://dx.doi.org/10.1002/jgrd.50847>>.

CHEPURIN, G. A.; CARTON, J. A.; DEE, D. **Forecast Model Bias Correction in Ocean Data Assimilation.** *Monthly Weather Review*, v. 133, n. 5, p. 1328–1342, 2005. Disponível em: <<http://dx.doi.org/10.1175/MWR2920.1>>.

DALEY, R. *Atmospheric Data Analysis*. Cambridge University Press, 1993. (Cambridge Atmospheric and Space Science Series). ISBN 9780521458252. Disponível em: <<https://books.google.com.br/books?id=RHM6pTMRTHwC>>.

DEE, D. P.; SILVA, A. M. da. **The Choice of Variable for Atmospheric Moisture Analysis.** *Monthly Weather Review*, v. 131, n. 1, p. 155–171, 2003. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(2003\)131<0155:TCOVFA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2003)131<0155:TCOVFA>2.0.CO;2)>.

DERBER, J.; ROSATI, A. **A Global Oceanic Data Assimilation System.** *Journal of Physical Oceanography*, v. 19, n. 9, p. 1333–1347, 1989. Disponível em: <[http://dx.doi.org/10.1175/1520-0485\(1989\)019<1333:AGODAS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2)>.

FIGUEROA, S. N.; BONATTI, J. P.; KUBOTA, P. Y.; GRELL, G. A.; MORRISON, H.; BARROS, S. R. M.; FERNANDEZ, J. P. R.; RAMIREZ, E.; SIQUEIRA, L.; LUZIA, G.; SILVA, J.; SILVA, J. R.; PENDHARKAR, J.; CAPISTRANO, V. B.; ALVIM, D. S.; ENORÉ, D. P.; DINIZ, F. L. R.; SATYAMURTI, P.; CAVALCANTI, I. F. A.; NOBRE, P.; BARBOSA, H. M. J.; MENDES, C. L.; PANETTA, J. **The Brazilian Global Atmospheric Model (BAM): Performance for Tropical Rainfall Forecasting and Sensitivity to Convective Scheme and Horizontal Resolution.** *Weather and Forecasting*, v. 31, n. 5, p. 1547–1572, 2016. Disponível em: <<http://dx.doi.org/10.1175/WAF-D-16-0062.1>>.

FISHER, M. **Background Error Covariance Modelling.** In: *Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*. [S.l.: s.n.], 2003. p. 45–63.

IDE, K.; COURTIER, P.; GHIL, M.; LORENC, A. C. **Unified Notation for Data Assimilation: Operational, Sequential and Variational (Special Issue Data Assimilation in Metereology and Oceanography: Theory and Practice).** *Journal of the Meteorological Society of Japan. Ser. II*, v. 75, n. 1B, p. 181–189, 1997.

KALNAY, E.; ANDERSON, D. L. T.; BENNETT, A. F.; BUSALACCHI, A. J.; COHN, S. E.; COURTIER, P.; DERBER, J.; LORENC, A. C.; PARRISH, D.; PURSER, J.; SATO, N.; SCHLATTER, T. **Data Assimilation in the Ocean and in the Atmosphere: What Should be Next? (Special Issue Data Assimilation in Meteorology and Oceanography: Theory and Practice)**. *Journal of the Meteorological Society of Japan. Ser. II*, v. 75, n. 1B, p. 489–496, 1997.

KLEIST, D. T.; PARRISH, D. F.; DERBER, J. C.; TREADON, R.; WU, W.-S.; LORD, S. **Introduction of the GSI into the NCEP Global Data Assimilation System**. *Weather and Forecasting*, American Meteorological Society, v. 24, n. 6, p. 1691–1705, 2013/11/18 2009. Disponível em: <<http://dx.doi.org/10.1175/2009WAF2222201.1>>.

KUBOTA, P. Y. **Variabilidade da Energia Armazenada na Superfície e o seu Impacto na Definição do Padrão de Precipitação na América do Sul**. 309 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012-08-06 2012. Disponível em: <<http://urlib.net/sid.inpe.br/mtc-m19/2012/08.02.02.42>>. Acesso em: 28 nov. 2013.

LEE, M.-S.; BARKER, D. **Preliminary Tests of First Guess at Appropriate Time (FGAT) with WRF 3DVar and WRF Model**. *Journal of the Korean Meteorological Society*, n. 41, p. 495–505, 2005.

LORENC, A. C. **Analysis Methods for Numerical Weather Prediction**. *Quarterly Journal of the Royal Meteorological Society*, John Wiley Sons, Ltd, v. 112, n. 474, p. 1177–1194, 1986. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1002/qj.49711247414>>.

PANNEKOUCKE, O.; BERRE, L.; DESROZIERS, G. **Background-Error Correlation Length-Scale Estimates and their Sampling Statistics**. *Quarterly Journal of the Royal Meteorological Society*, John Wiley & Sons, Ltd., v. 134, n. 631, p. 497–508, 2008. ISSN 1477-870X. Disponível em: <<http://dx.doi.org/10.1002/qj.212>>.

PARRISH, D. F.; DERBER, J. C. **The National Meteorological Center's Spectral Statistical-Interpolation Analysis System**. *Monthly Weather Review*, American Meteorological Society, v. 120, n. 8, p. 1747–1763, 2013/11/18 1992. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2)>.

PETRIE, R. E. **Background Error Covariance Modelling for Convective-Scale Variational Data Assimilation**. Tese (Doutorado) — The University of Reading (United Kingdom, 2012).

PURSER, R. J.; WU, W.-S.; PARRISH, D. F.; ROBERTS, N. M. **Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances**. *Monthly Weather Review*, American Meteorological Society, v. 131, n. 8, p. 1536–1548, 2013/11/18 2003. Disponível em: <<http://dx.doi.org/10.1175//2543.1>>.

PURSER, R. J.; WU, W.-S.; PARRISH, D. F.; ROBERTS, N. M. **Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomogeneous and Anisotropic General Covariances**. *Monthly Weather*

Review, American Meteorological Society, v. 131, n. 8, p. 1536–1548, 2013/11/18 2003. Disponível em: <<http://dx.doi.org/10.1175//2543.1>>.

RASCH, P. J.; WILLIAMSON, D. L. **The Sensitivity of a General Circulation Model Climate to the Moisture Transport Formulation.** *Journal of Geophysical Research: Atmospheres*, v. 96, n. D7, p. 13123–13137, 1991. ISSN 2156-2202. Disponível em: <<http://dx.doi.org/10.1029/91JD01179>>.

STENGEL, R. *Optimal Control and Estimation.* Dover Publications, 1994. (Dover books on advanced mathematics). ISBN 9780486682006. Disponível em: <<https://books.google.com.br/books?id=jDjPxqm7Lw0C>>.

WU, W.-S.; PURSER, R. J.; PARRISH, D. F. **Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances.** *Monthly Weather Review*, v. 130, n. 12, p. 2905–2916, 2002. Disponível em: <[http://dx.doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2)>.

ANEXO B - Artigo Científico II

Journal of Applied Meteorology and Climatology
**GSI-BASED HYBRID 3DVAR DATA ASSIMILATION FOR THE BAM-CPTEC/INPE:
SINGLE LOW RESOLUTION EXPERIMENTS**
--Manuscript Draft--

Manuscript Number:	JAMC-D-17-0170
Full Title:	GSI-BASED HYBRID 3DVAR DATA ASSIMILATION FOR THE BAM-CPTEC/INPE: SINGLE LOW RESOLUTION EXPERIMENTS
Article Type:	Article
Corresponding Author:	Carlos Frederico Bastarz, Msc. INPE Cachoeira Paulista, Sao Paulo BRAZIL
Corresponding Author's Institution:	INPE
First Author:	Carlos Frederico Bastarz, Msc.
Order of Authors:	Carlos Frederico Bastarz, Msc. Dirceu Herdies
Manuscript Classifications:	8.052: Data assimilation; 8.148: Numerical weather prediction/forecasting
Abstract:	This paper aims to present the first results obtained with a GSI-based Hybrid Ensemble-Variational system applied to the Brazilian Atmospheric Model from Center for Weather Forecasts and Climate Studies at the National Institute for Space Research (BAM-CPTEC/INPE). The hybrid data assimilation system implemented is a hybrid 3DVar, using the Ensemble Kalman Filter (EnKF) and the Gridpoint Statistical Interpolation (GSI) 3DVar analysis system. This is an upgrade to the previous GSI 3DVar system being used at CPTEC operations. The implementation was made using a previous established model and assimilation framework. A new background error covariance matrix was calculated, in order to properly account the ensemble covariance contribution to the static climatological part. An analysis and an evaluation of the updated system is presented and the first results for analyses and 5-day forecasts at TQ0062L028 model resolution are presented as well. The tested hybrid 3DVar system at CPTEC shows an improvement for the BAM-CPTEC/INPE skill mainly over South America and Tropical region and endorses the use of ensemble covariances as a complement to the static part of the covariances as a way to reduce the misrepresentation of the background error covariances. Precipitation has been also evaluated in a comparison with observed data. Exercising with the hybrid system have shown an improvement over the pure 3DVar system in both the intensity and the distribution of the precipitation over the time mean. The results obtained with the implementation are encouraging and further improvements must be made in order to fully access the analysis and forecast skill, specially under higher resolutions.
Suggested Reviewers:	Jeffrey Whitaker jeffrey.s.whitaker@noaa.gov Jeffrey Whitaker work refers to ensemble Kalman filters and hybrid methods in data assimilation. Daryl Kleist daryl.kleist@noaa.gov Daryl Kleist work refers to the Gridpoint Statistical Interpolation system and hybrid methods in data assimilation.

Carlos Frederico Bastarz
National Institute for Space Research
Avenida dos Astronautas, 1778
Jardim da Granja
CEP 12227-010
São José dos Campos/SP Brazil

June 17, 2017

Dr. David Kristovich

Chief Editor of the Journal of Applied Meteorology and Climatology

We wish to submit a new manuscript entitled "**GSI-BASED HYBRID 3DVAR DATA ASSIMILATION FOR THE BAM-CPTEC/INPE: SINGLE LOW RESOLUTION EXPERIMENTS**" for consideration by the Journal of Applied Meteorology and Climatology.

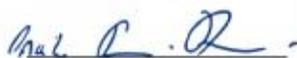
This research is related to the application of a GSI-based hybrid 3DVar data assimilation coupled to the CPTEC/INPE general circulation model. The objective of this research is to verify the impacts of the use of ensemble Kalman filter covariances combined with static covariances. This static covariance matrix was calculated using 730 pairs of 48/24 hours forecasts. The hybrid 3DVar system, is exercised with 40 members and a real non linear model. The results are encouraging and shows that when using the hybrid covariances with a greater contribution from the ensemble part, the forecasts skill are improved for various model state variables. We also show that hybrid covariances can also help the model to improve the distribution of the 24 hour precipitation, in both space and time. As we have evaluated the system performance for different regions of the globe, we found that the ensemble covariances helped the model to improve its skill in regions that are known to have a lower conventional observation coverage (eg. South Hemisphere/South America). We believe that this results whould be of the interest of the journal readers because it endorses the use of hybrid covariances in data assimilation applications, specially over the southern hemisphere.

We confirm that this work is original and has not been published elsewhere nor is it currently under consideration for publication elsewhere. All authors have approved the manuscript and agree with its submission to the Journal of Applied Meteorology and Climatology.

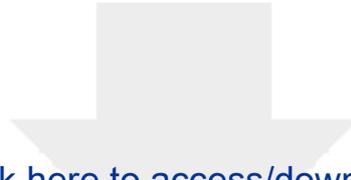
Please address all correspondence concerning this manuscript to me at carlos.frederico@cptec.inpe.br.

Thank you for your consideration of this manuscript.

Sincerely,



Carlos Frederico Bastarz on behalf of the authors



[Click here to access/download](#)

Cost Estimation and Agreement Worksheet
Cost Estimation and Agreement Worksheet_FILL.pdf



1 **GSI-BASED HYBRID 3DVAR DATA ASSIMILATION FOR THE**
2 **BAM-CPTEC/INPE: SINGLE LOW RESOLUTION EXPERIMENTS**

3 Carlos Frederico Bastarz* and Dirceu Luis Herdies †

4 *Center for Weather Forecast and Climate Studies, CPTEC/INPE, Cachoeira Paulista, SP, Brazil*

5 **Corresponding author address:* Carlos Frederico Bastarz, CPTEC/INPE, Rod. Presidente Dutra,
6 km 40, Cachoeira Paulista, SP, Brazil.

7 E-mail: carlos.frederico@cptec.inpe.br

8 †Center for Weather Forecast and Climate Studies, CPTEC/INPE, Cachoeira Paulista, SP, Brazil

ABSTRACT

9 This paper aims to present the first results obtained with a GSI-based Hybrid
10 Ensemble-Variational system applied to the Brazilian Atmospheric Model
11 from Center for Weather Forecasts and Climate Studies at the National In-
12 stitute for Space Research (BAM-CPTEC/INPE). The hybrid data assimila-
13 tion system implemented is a hybrid 3DVar, using the Ensemble Kalman Fil-
14 ter (EnKF) and the Gridpoint Statistical Interpolation (GSI) 3DVar analysis
15 system. This is an upgrade to the previous GSI 3DVar system being used
16 at CPTEC operations. The implementation was made using a previous es-
17 tablished model and assimilation framework. A new background error co-
18 variance matrix was calculated, in order to properly account the ensemble
19 covariance contribution to the static climatological part. An analysis and an
20 evaluation of the updated system is presented and the first results for anal-
21 yses and 5-day forecasts at TQ0062L028 model resolution are presented as
22 well. The tested hybrid 3DVar system at CPTEC shows an improvement for
23 the BAM-CPTEC/INPE skill mainly over South America and Tropical re-
24 gion and endorses the use of ensemble covariances as a complement to the
25 static part of the covariances as a way to reduce the misrepresentation of the
26 background error covariances. Precipitation has been also evaluated in a com-
27 parison with observed data. Exercising with the hybrid system have shown an
28 improvement over the pure 3DVar system in both the intensity and the dis-
29 tribution of the precipitation over the time mean. The results obtained with
30 the implementation are encouraging and further improvements must be made
31 in order to fully access the analysis and forecast skill, specially under higher
32 resolutions.

33 **1. Introduction**

34 Hybrid data assimilation systems using an ensemble Kalman Filter and a variational technique
35 have been developed (e.g., Hamill and Snyder 2000; Lorenc 2003; Zupanski 2005) and applied
36 focusing the sampling and the representation of the spatial-temporal variations of background
37 error covariances, introducing the day-to-day statistics variations of the background flow (the so
38 called “errors of the day” - Corazza et al. 2003) to the static part of the variational covariances.
39 The representation of background error covariances is one of the main issues of the operational
40 data assimilation and its specification for deterministic systems (e.g., variational systems) is of a
41 primordial importance.

42 Variational systems such as 3DVar uses a static background error covariance matrix, which
43 means that its covariances are stationary in time and does not vary, thus the same error statistics
44 are used for the whole assimilation process. A similar situation happens to a 4DVar system, which
45 - in general, uses a pre-computed covariance background error matrix at the beginning of each
46 assimilation window, which is at least propagated in time during the assimilation of observations
47 using the tangent linear version of the forecast model. Differences in the way that background
48 error statistics varies with time due to the variations in the background flow, leads to what we call
49 flow-dependence.

50 Flow-dependent covariances are a key feature of modern data assimilation systems, and its spec-
51 ification either at the beginning or during the assimilation window, allows the analysis to correct
52 account for the time and spatial dependency of the observation innovations being used into the
53 analysis increments, in both horizontal and vertical directions. Furthermore, desirable features of
54 the background error covariances also include anisotropy and inhomogeneity (non-uniform statis-
55 tics - Rabier 2005).

56 Several efforts have been made in order to add some degree of flow-dependence and anisotropy
57 to static covariances. The development of the variational techniques for operational use during the
58 1990's, include some sort of implicit modifications to the data assimilation framework. Desroziers
59 (1997) introduced a coordinate transformation to allow the data assimilation system to better ac-
60 count for structures of frontal systems. This led to an improved analysis with flow dependent
61 covariances and anisotropic correlations.

62 The gradients represented in the background flow, also include important information that could
63 be used to allow anisotropy. Riishøjgaard (1998) showed how to use the humidity field to model
64 correlation functions that can be stretched to accommodate the covariances according to the gradi-
65 ents represented in the background flow. Other approaches to model correlation functions that lead
66 to flow-dependent background error statistics, involve the use of wavelets with spectral methods
67 (e.g., Fisher 2003), but difficulties arise with the specification of these wavelets on the sphere. Re-
68 cursive filters are also another know method to derive quasi-Gaussian shapes to use in background
69 error covariance modeling. This approach have been used successfully in several applications;
70 e.g., Hayden and Purser (1995) applied isotropic recursive filters to the processing of the Na-
71 tional Environmental Satellite and Data Information Service (NESDIS); more recently, recursive
72 filters have been applied to the Gridpoint Statistical Interpolation (GSI) to model the application
73 of background error covariances (Wu et al. 2002).

74 Within these efforts, other approaches have been developed and are more related to the represen-
75 tation of nonlinearities of the background flow into the background error statistics. The 4DVar (Th-
76 paut and Courtier 1991) was developed as an extension of 3DVar (Lorenc 1986) and can implicit
77 evolve covariances inside the assimilation window. Corazza et al. (2003) added flow-dependence
78 to covariances by accounting for the “errors of the day”, using bred-vectors. Hybrid methods (e.g.,
79 Hamill and Snyder 2000) were introduced taking advantage of the ensemble Kalman Filter meth-

ods. This new approach has several advantages including the fact that operational centers have already experienced the variational technique and have some experience with ensemble Kalman Filters. Each of these methods can be taken as a compliment to ameliorate the deficiencies of the other (Wang et al. 2007, 2009). Recently, hybrid methods with real applications have been deployed (e.g., Wang et al. 2013; Clayton et al. 2013) where covariances estimated from an ensemble Kalman Filter, are linearly combined to a 3DVar static covariances matrix.

Most recent developments at CPTEC have focused on its independence on observations pre-processing, quality control and analysis. In the past few years, some analysis systems have been tested at CPTEC, including the GSI. GSI have been used in several operational NWP centers and is the data assimilation chosen for regional and global applications at the center.

With the objective of provide the BAM-CPTEC/INPE model (Figueroa et al. 2016) with its own analysis in a data assimilation cycle, in this paper are described the details of the implemented system and the first results obtained with the GSI-based hybrid 3DVar system.

This paper is organized as follows: Section 2 describes the strategy and the details of the implemented hybrid 3DVar system and the static part of the background error covariance matrix; Section 3 presents a description of the idealized experiments to test the hybrid system. Results from the experiments carried out are represented in Section 4 and finally, Section 5 presents the discussion regarding the results and future plans.

2. CPTEC Global Hybrid 3DVar System

In the past years, hybrid data assimilation systems have shown an increasing amount of development and several centers have been made efforts to take advantage of its benefits. Hybrids systems using the 3DVar and an ensemble Kalman Filter (EnKF) appears to be a common choice to start with. The GSI system comprises a collection of routines that are already prepared to run such

103 a hybrid 3DVar system using the EnKF to provide the ensemble covariances to compliment the
104 static covariances used in 3DVar.

105 Since 2012 CPTEC have been experimenting the GSI 3DVar analysis with its global General
106 Circulation Model (GCM) for NWP until 7 days at a TQ0299L064 spectral model resolution
107 (roughly, 45 km at the Equator, with a corresponding grid of 450 X 900 X 64 points, *lat X lon*
108 *X lev*). This version of the system was performed with a GSI's example global background error
109 covariance matrix (with a grid of 386 X 768 X 64 points, *lat X lon X lev*). Although this config-
110 uration seems to be not the ideal, it was a first attempt to run a global analysis with radiance data
111 assimilation at a higher resolution at CPTEC.

112 With the public release of the hybrid version of GSI in 2015, CPTEC started to plan a revamped
113 version of its global data assimilation system. A new version of the BAM-CPTEC/INPE model
114 was released in the summer of 2015 and was planned a test of running the BAM-CPTEC/INPE
115 with its own analysis, but using a proper background error covariance matrix and also testing with
116 the new hybrid system structure of GSI.

117 Exercise with the hybrid 3DVar assimilation technique with the BAM-CPTEC/INPE, was made
118 using the Ensemble Square Root Filter (EnSRF - Whitaker and Hammill 2002). Two experiments
119 were carried out: one experiment using 50% of ensemble contribution to the static covariance and
120 a second experiment with 75% of ensemble contribution, were designed to test the hybrid 3DVar
121 system. Another experiment with a pure 3DVar using the BAM-CPTEC/INPE in a cycled data
122 assimilation and the realization of the BAM-CPTEC/INPE with the National Centers for Environ-
123 mental Predictions (NCEP) analysis were also performed. For the three experiments using data
124 assimilation, the new static version of the background error covariance matrix were used. As we
125 have made a new background error covariance matrix for use in the experiments with the tradi-
126 tional 3DVar and the hybrid data assimilation, Subsection 2a shows the details and main aspects

127 the the new static matrix. Subsections 2b and 2c are dedicated to present the BAM-CPTEC/INPE
128 model information and the main setup for the hybrid 3DVar system, respectively. A detailed de-
129 scription of the experiments is given in Section 3.

130 *a. Static Background Error Covariance Matrix*

131 The background error covariance matrix (**B**) has the role to filter and propagate spatially the
132 observation information (Berre et al. 2013), contributing to transform part of the observations
133 innovations into analysis increments.

134 The calculation of the background error covariance matrix for the GSI system used in this work,
135 involved the use of 730 pairs of 48 and 24 hours forecasts, valid for the 0000 and 1200UTC,
136 distributed evenly over 1 year (2013). The resolution of the forecasts matches the resolution of
137 the forecast model used in the experiments, i.e., TQ0062L028 (with 192 x 96 x 28 - *lat X lon*
138 *X lev* grid points). The forecasts pairs were organized as follows: considering 2014010100, the
139 first valid forecast pair was generated with the analysis from 2013123100 (valid for a 24 hours
140 forecast) and 2013123000 (valid for a 48 hours forecast). The methodology used to calculate the
141 amplitudes (i.e., the variances and covariances) is based on the NMC method (Parrish and Derber
142 1992). This method claims that spatial correlations of the model errors are similar to the spatial
143 correlations of the differences between the 48 and 24 hours forecasts. The NMC method was cho-
144 sen for the background error covariance calculations due to the convenience of having previously
145 analyses from operations. These analyses were used to integrate the BAM-CPTEC/INPE model
146 at TQ0062L028 resolution for one year. Within the algorithm involved in the calculations, there is
147 also a procedure to remove the bias in the vertical model column, which were also used. Covari-
148 ances were calculated for the standard GSI control variables, i.e., streamfunction (ψ), unbalanced
149 velocity potential (χ), unbalanced temperature (T), unbalanced surface pressure (ps) and the nor-

150 malized relative humidity (*rh*). The sea surface temperature (*sst*) is the same as the one used
 151 on the GFS-NCEP background error covariance matrix (i.e., the Real-Time, Global, Sea Surface
 152 Temperature - RTG_SST analysis, from which the variances and length scales were interpolated to
 153 the target grid).

154 In practice, although GSI doesn't explicit construct \mathbf{B} , it applies the covariances of \mathbf{B} using
 155 recursive filters (Purser et al. 2003a,b):

$$\mathbf{B} = \mathbf{B}_z(V^1 \mathbf{B}_x^1 \mathbf{B}_y^1 \mathbf{B}_x^1 V^1 + V^2 \mathbf{B}_x^2 \mathbf{B}_y^2 \mathbf{B}_y^2 V^2) \mathbf{B}_z \quad (1)$$

156 where,

- 157 • V^1 and V^2 are the standard deviations of each control variable (from the \mathbf{B} file);
- 158 • \mathbf{B}_x , \mathbf{B}_y and \mathbf{B}_z are the recursive filter applications in the directions of x (west-east), y (south-
 159 north) and z (vertical);
- 160 • \mathbf{B}^1 and \mathbf{B}^2 represents the application of the recursive filter in the horizontal scales (x and y).

161 An important property of the recursive filters in the GSI framework, is the adjustment of the
 162 amplitudes and the length scales in order to make the aspect of the covariances anisotropic, letting
 163 them to adjust to the background flow. These parameters and scales used in the application of the
 164 recursive filters are sort of empirical and a discussion of it can be found in Wu et al. (2002). The
 165 resulting amplitudes of the calculated \mathbf{B} are presented in Section 4a.

166 *b. The Brazilian Atmospheric Model (BAM)*

167 CPTEC has been made efforts to continuously develop and improve its own GCM for operational
 168 use at the center. As a result of this efforts, the BAM-CPTEC/INPE (Figueroa et al. 2016) brings

169 the most recent advances in both model dynamics and physics. Although the version of the BAM-
170 CPTEC/INPE model used in this work is not exactly the same version used in Figueroa et al.
171 (2016), the version used in the experiments will be referred as “BAM v0.0”, which includes the
172 same dynamical core and a simplified physics package (mostly described in Cavalcanti et al. 2002).
173 This simplified physics package is based on a large scale condensation due to the chosen model
174 resolution (TQ0062L028, roughly 200 km at the Equator) and the ensemble size (40 plus 1 of
175 the hybrid analysis). Furthermore, microphysics was avoided due to the intense computational
176 requirements (Silvio N. Figueroa, personal communication).

177 The BAM-CPTEC/INPE is a spectral model representing mass and winds in terms of vorticity
178 and divergence, with a pure sigma vertical coordinate, used at the center for operational NWP
179 (current resolution is TQ0666L064, roughly 20 km at the Equator) up to 5 days, and - at a coarse
180 resolution (TQ0126L028, roughly 100 km near the Equator) for extended NWP up to 15 days,
181 using an ensemble prediction system based on Empirical Orthogonal Functions (EOF) for the
182 perturbation of the initial condition. Seasonal and climate simulations are made as well. For these
183 applications analysis are still taken from GFS-NCEP.

184 *c. Hybrid Ensemble-Variational Analysis Cycle*

185 The hybrid 3DVar is the variational system with hybrid covariances intended to be the next
186 generation of atmospheric analysis at CPTEC. This system was exercised in an experimental basis
187 at CPTEC for weather forecasts only (up to 5 days at TQ0062L028 model resolution). Plans
188 include to use the ensemble of analysis from the hybrid system to replace or to augment the
189 current size (15 members) and consequently the spread of the center’s global Ensemble Prediction
190 System (EPS, described in Mendonça and Bonatti 2009; Cunningham et al. 2015), for extended
191 atmospheric forecasts (up to 15 days). The hybrid system used in this work is coupled with the

192 BAMv0 model through an interface that reads the spectral forecasts from the BAMv0 model and
 193 rewrites the spectral coefficients for use as a background for the GSI system.

194 The current hybrid setup requires the realization of the GSI observer using the mean ensemble as
 195 a background. No analysis is made at this point and this step is only necessary to write the diagnos-
 196 tics files from the departures. The next step is the realization of the observer using the departures
 197 related to the ensemble mean, but using every ensemble member as a background. In the setup
 198 tested, an ensemble of 40 members was used. This number represents a reasonable ensemble size,
 199 which fits all the computational requirements, regarding the allowed computational time and disk
 200 space without compromise the efficiency. Once the observer ensemble of departures is computed
 201 within GSI, the EnSRF is applied to update the ensemble of backgrounds. The final determin-
 202 istic hybrid 3DVar analysis is then achieved using the covariances from the updated ensemble of
 203 backgrounds, which are blended with the static pre-computed background error covariance matrix.
 204 Figure 1 shows a schematic diagram of the analysis cycle.

205 The hybrid 3DVar system minimizes a 3D variational cost function in the same fashion as 3DVar.
 206 The main difference is regarded to the way the background error covariance matrix is defined and
 207 applied. Equation 2 (presented here in a more general way) represents the hybrid 3DVar cost
 208 function, where x' is the new analysis increment.

$$J(x') = \frac{1}{2}(x')^T (\alpha_1 \mathbf{B} + \alpha_2 \mathbf{P}^b)^{-1} (x') + \frac{1}{2} [y_o - \mathbf{H}(x')]^T \mathbf{R}^{-1} [y_o - \mathbf{H}(x')] \quad (2)$$

209 Note that the term $(\alpha_1 \mathbf{B} + \alpha_2 \mathbf{P}^b)$ represents the linear combination between the pre-computed
 210 static \mathbf{B} and the ensemble flow-dependend ensemble covariances \mathbf{P}^b . The coefficients α_1 and α_2
 211 are the weights assigned to each parcel of the linear combination, trough which we control the
 212 amount of contribution of the ensemble covariances to the static part.

213 In Equation 2, x' is defined as:

$$x' = x + \sum_{k=1}^K (a_k \circ x_k^e) \quad (3)$$

214 This analysis increment is defined as the sum between the standard analysis increment (x) due
 215 to the static background error covariance contribution and the summing between the extension of
 216 the control variable (a_k , to account for the ensemble localization) and the ensemble perturbation
 217 of each member ($x_k^e = \frac{(x_k^b - \bar{x}^b)}{\sqrt{K-1}}$). A more detailed explanation of the GSI methodology to extend
 218 the variational framework to accommodate the ensemble covariances can be found in Wang et al.
 219 (2007, 2008a,b); Wang (2010); Wang et al. (2013).

220 Despite the fact that the hybrid 3DVar analysis cycle have been established for the BAMv0
 221 model, some caveats have been found and will demand attention on future revisions of this sys-
 222 tem. The system works in a single resolution, i.e., both the control and the ensemble of back-
 223 ground/analysis are run at the same model resolution (currently TQ0062L028). As a consequence,
 224 a recentering of the ensemble mean around the hybrid analysis, is not made. Another point to
 225 mention is the fact that the bias mass and angle of the radiance channels and instruments were not
 226 being updated. This decision was taken because, as a testing system, no compromise was made
 227 to keep the radiance information error updated. Also, there is no vertical localization within the
 228 EnSRF, so a fixed value is used for the entire vertical column.

229 3. Experimental design

230 To test the hybrid 3DVar system, a set of experiments were idealized in order to verify the
 231 analysis and forecasts skill (in terms of the Root Mean Square Error - RMSE and the Anomaly
 232 Correlation - AC), by comparing them with a pure (control) 3DVar analysis. The pure 3DVar
 233 analysis, in turn, was compared to an offline analysis experiment by integrating the BAMv0 model

234 with an independent analysis. The next subsection are dedicated to explain the design of each
235 experiment.

236 *a. Control Experiment*

237 The control experiment is intended to be a 3DVar experiment run with the new background
238 error covariance matrix. This experiment is the control run and is compared to the hybrid 3DVar
239 experiments. The control was ran using the variational analysis in TQ0062L028 resolution and
240 the model configuration (both the dynamics and the physics) is the same among every experiment.
241 The analysis cycle is setup in the following manner. A set of three background forecasts of 3, 6
242 and 9 hours (where the central analysis is always at 6 hours) are used by GSI for the update with
243 the observations in a First Guess at Appropriate Time (FGAT) approach. Observations include
244 conventional data (u , v , T , q and ps) from several sources (e.g., the Global Telecommunication
245 System - GTS) and unconventional data, like radiance, retrievals and Global Positioning System -
246 Radio Occultation (GPS-RO).

247 Table 2 in Section 3c summarizes the types of observations included in the experiments. The
248 same set of observations are used for the hybrid 3DVar experiments as well.

249 The analysis cycles continues when GSI writes the analysis file for the atmosphere and the
250 surface. It is worth to note that only the atmospheric part of the analysis was used to initialize the
251 BAMv0, although the model restart files were used to initialize the surface fields, as well clouds
252 and radiation. This same approach is taken for the hybrid experiments.

253 Another idealized experiment is the offline analysis (i.e., without an analysis cycle). In this
254 experiment the BAMv0 where integrated in time with a hi-resolution NCEP spectral operational
255 analysis which was chopped at each analysis time (i.e., 00, 06, 12 and 18UTC). From this ex-
256 periment, 120 hours forecasts are generated, in the same fashion as the pure 3DVar run. This

257 offline experiment is intended to serve as a control for the pure 3DVar experiment, whereas the
258 pure 3DVar analysis experiment serves as a control for the hybrid 3DVar experiments.

259 *b. Hybrid 3DVar Experiments*

260 Exercising with the hybrid 3DVar was made cycling the system with the BAMv0 model be-
261 tween December 2012 and January 2013. A spin up time was considered and the first month of
262 the simulations was taken apart from the results. One of the reasons for this practice is the fact
263 that the initial ensemble of analysis was generated by running the BAMv0 model from a single
264 deterministic analysis and forecasts made up to 30 days at each 12 hours (i.e., twice a day). From
265 this, a time series of 60 forecasts ranging from 12 hours up to 720 hours from the analysis time
266 was assembled. The first 40 forecasts were kept and carefully modified to reflect the background
267 time for the first analysis to be generated in all hybrid 3DVar experiments.

268 To assess the behaviour of the hybrid 3DVar system using the BAMv0 model as a background
269 and its new background error covariance matrix, two cycled data assimilation experiments were
270 ran. The experiment called 3DVar represents the variational GSI/3DVar experiment with the new **B**
271 (considering $\alpha_1 = 1$, i.e., no ensemble covariance contribution). The experiment called EnSRF50
272 (EnSRF75) represent the same data assimilation cycle ran using the hybrid **B** through the use of
273 the EnSRF (considering $\alpha_2 = 0.5$ and $\alpha_2 = 0.25$, i.e., considering 50% and 75% of contribution
274 of the ensemble covariances, respectively).

275 Table 1 summarizes the experiments designed to exercise the hybrid 3DVar system, ranking the
276 ensemble covariance contribution for each experiment.

277 *c. Observational Data*

278 The observational data used for the experiments with a cycled analysis using the 3DVar GSI and
279 the EnSRF algorithm, are the same. Table 2 shows a summary of the observational data used.

280 *d. Model Setup*

281 In all experiments, regardless the analysis used, the same model configuration had been taken.
282 Table 3 summarizes the main options adjusted within the dynamical core and physics options. It
283 is worth to note that in the data assimilation cycle (regardless the hybrid or the pure variational
284 option), the atmospheric model was restarted at every cycle (00, 06, 12 and 18UTC). The BAMv0
285 restart files comprises the previous states of the land-surface, clouds and radiation. Sea surface
286 temperature, snow cover and soil moisture are read from separated files and the dynamical state
287 (i.e., momentum) are provided by the atmospheric analysis. These conditions are valid for the
288 generation of the background files (3h, 6 and 9 hours forecasts) used by the GSI.

289 *e. GSI Setup*

290 This subsection is intended to give an overview of the basic and main GSI options that was
291 chosen to complete the experimental design. GSI has many different option that can be adjusted
292 in order to allow the system run with specific procedures or to tune the system through the use
293 of knobs that must be adjusted empirically. All of these option can be found on the official GSI
294 manual, available at the Development Testbed Center (DTC) website.

295 Two of these general options worth to mention here (the choice for humidity control variable
296 and the moisture constraint), as they were kept the same among all experiments using GSI.

297 HUMIDITY CONTROL VARIABLE

298 GSI has two options to treat the humidity control variable. They are the pseudo relative-humidity
299 and the normalized relative humidity. As we were testing a new background error covariance
300 matrix, we decided to choose the normalized relative humidity as the humidity control variable in
301 all experiments. The normalized relative humidity allows the relative humidity to change in the
302 inner loops in accordance to changes in the surface pressure, temperature or specific humidity.

303 MOISTURE CONSTRAINT

304 Supersaturated and negative moisture are non-physical solutions that arises from computation
305 modes. In GSI, there is an option that can be tuned in order to control the supersaturated and
306 negative moisture. Several tests were made in order to achieve reasonable values to allow the
307 stability of the system. The values found are 5.0 for the supersaturated moisture constraint and
308 0.005 for the negative moisture constraint.

309 **4. Results**

310 This section shows the results obtained with the application of a hybrid background error co-
311 variance matrix through the use of a hybrid 3DVar system for the BAMv0 model. Along with the
312 experimental results, the main characteristics of the calculated background error covariance matrix
313 are shown. Results from the experiments with the hybrid 3DVar are divided into three categories:
314 1) the ensemble innovation statistics; 2) forecasts skill (up to 120 hours, taken in terms of AC along
315 with t-Student tests) and 3) precipitation assessment (taken in term of a comparison with the GPCP
316 data). The evaluations are made to the regions Global (GL), North Hemisphere (NH - lons:0°-
317 360°; lats:20°N-80°N), Tropics (TR - lons:0°-360°; lats:20°S-20°N), South Hemisphere (SH -
318 lons:0°-360°; lats:80°S-20°S) and South America (SA - lons:0°-360°; lats:49.875°S-11.375°N).

319 *a. Standard Deviations of the Static \mathbf{B}*

320 Experiments with a cycled data assimilation (analysis from 3DVar and hybrid 3DVar experi-
321 ments) were made using the same version of the static background error covariance matrix. In
322 Section 2a, it was showed how the static part of the hybrid background error covariance matrix
323 was calculated and how its application within GSI is made. Following the results in Wu et al.
324 (2002), the main features of the static \mathbf{B} are presented here in a similar manner.

325 In order to make an quantitative and a qualitative description of the amplitudes represented by
326 the computed covariance matrix, Figure 2 depicts the standard deviations of the variance error for
327 the ψ , χ , T , rh , oz and ps . As a reference, Table 4 presents the maxima and minima values for the
328 amplitudes shown in Figure 2 along with the values for the horizontal and vertical length-scales.

329 The distribution of the variances of stream function (ψ) and velocity potential (χ) are presented
330 in a vertical cross-section, as shown in Figure 2. These distribution shows some of the features of
331 the non-divergent and divergent part of the wind for the BAMv0 model. The ψ variances distribu-
332 tion are sort of symmetric with respect to the Equator line. They are more elongated throughout
333 the vertical column and are concentrated at the model top. Higher values of the streamfunction
334 variances are located over the poles, between sigma levels of 0.6 (~ 614.4 hPa) and 0.2 (~ 204.8
335 hPa). In the tropical regional, these values are more confined between 0.3 (~ 307.2 hPa) and 0.1
336 (~ 102.4 hPa) sigma levels. The distribution of the variances of the velocity potential also shows a
337 symmetric pattern with respect to the Equator line and are mostly concentrated with higher values
338 over the tropical region, between sigma levels of 0.3 (~ 307.2 hPa) and 0.1 (~ 102.4 hPa).

339 Temperature (T) variances are distribute from the Equator line towards the pole, in both hemi-
340 spheres. At 30°N large values extend onwards with a maximum peak between 0.9 (~ 921.6 hPa)
341 and 0.8 (~ 819.2 hPa) sigma layers. At 60°S , a symmetric pattern is represented, as can be seen

342 in Figure 4c in Wu et al. (2002). Variances of the humidity (rh) are distributed with a peak in
343 approximately 60°S between 0.7 (~ 716.8 hPa) and 0.6 (~ 614.4 hPa) sigma layers. Layers above
344 0.6 sigma, variances diminish with height until layer 0.2 (~ 204.8 hPa), whereas the opposite pat-
345 tern seems to happen around latitude 60°N . Ozone (oz) distribution of variances are concentrated
346 in upper layers, 0.5 (~ 512 hPa) upwards, with a peak in the tropical region. In the current setup,
347 GSI does not analyses ozone observations. In the BAMv0 model ozone is transported throughout
348 the grid during the forecast. The surface pressure (ps) variances are distributed with two local
349 maxima values in the midlatitudes (higher values are located below 60°S) and minimum values
350 in the tropical region. It is worth to note that the surface pressure variance distribution are fairly
351 different from what can be found in Figure 4e from Wu et al. (2002).

352 Most of the characteristics in the distribution of the error amplitudes may be due to the represen-
353 tation of the BAMv0 topography, including the choices for the model physics and the treatment of
354 the model dynamics and resolution. At this resolution (i.e., TQ0062L028), steep topography may
355 be not well represented (peaks in topography may be too smooth), what can lead to discrepancies
356 in the representation of the pressure and winds. A separated study still has to be made in order
357 to address the sensitivity of the background error covariance matrix to the BAMv0 options and
358 the resulting variances. Furthermore, one may consider the fact that such amplitudes can be tuned
359 within GSI by setting up specific parameters for the application of the covariances using recursive
360 filters.

361 Table 4 summarizes the maximum and minimum values for the amplitudes of the standard de-
362 viations and horizontal and vertical length-scales, including for the other quantities not shown in
363 Figure 2.

364 *b. Ensemble Innovation Statistics*

365 For an evaluation of the ensemble of analyses produced in experiments using the EnSRF, an
366 ensemble innovation assessment was made in terms of a scaled bias to diagnose deficiencies in the
367 ensemble spread due to inflation and localization adjustment.

368 To assess how the ensemble deals with the system spread in the presence of observational error,
369 innovations are shown as a measure of how good the ensemble spread is due to the observation
370 innovation. In Figure 3, a comparison between the innovation statistics from EnSRF is shown for
371 three different regions (NH; TR and SH) and only for the conventional observations. We left the
372 same evaluation regarding unconventional observations for a separated study. In order to make
373 it easier to understand, the y axis shows the ratio between the standard deviation of the prior
374 innovations (i.e., $y^o - \mathbf{H}x_k^b$, where k is an ensemble member) and the square root of the ensemble
375 total spread (i.e., $\sqrt{\mathbf{S} + \mathbf{R}}$). The ensemble innovation is given by,

$$EI = \frac{y^o - \mathbf{H}x_k^b}{\sqrt{\mathbf{S} + \mathbf{R}}} \quad (4)$$

376 where,

- 377 • y^o : is the observation vector;
- 378 • x_k^b : is the vector of the k -th ensemble member;
- 379 • \mathbf{S} : is the ensemble spread;
- 380 • \mathbf{R} : is the observation error from the observation error covariance matrix.

381 As we are normalizing the standard deviation of the priors innovations by the ensemble total
382 spread, the lower the values, the better the ensemble innovation is. Ideally, the standard devi-

383 ation of the priors should match the value of the ensemble total spread if the ensemble is well
384 conditioned, i.e., with a good amount of spread and proper localization and inflation.

385 In Figure 3, the upper panel shows the statistics for 00UTC and the bottom panel the same
386 statistics for 12UTC. This was made with the purpose to reduce the size of the time series and to
387 separate the signal from both times, so this would make easier to analyze the results. Furthermore,
388 synoptic times of 06 and 18UTC were kept away from the evaluation due to the fact that there is
389 less meteorological observations at these synoptic times. Red lines (dashed and solid) refers to the
390 hybrid 3DVar experiment where 50% of ensemble contribution to the static part of the covariance
391 matrix, whereas the blue lines refers to the same experiment but denoting the case where 75% of
392 ensemble contribution to the static part of the covariances were drawn.

393 Figures 3d,e,f (in both panels) with the ensemble innovation statistics for T at NH, TR and SH,
394 regardless the relative amount of ensemble contribution, shows that the differences between the
395 priors and the posteriors are more noticeable at the NH, where T is closer to the zero line.

396 The standard deviation for the horizontal wind (uv) at 00UTC (Figure 3a), was very close to the
397 ensemble total spread, indicating that the ensemble fit to the observation was good and the amount
398 of ensemble spread was reasonable. At the TR region (Figure 3b), however, the opposite was
399 found. In the SH region (Figure 3c), a similar situation to NH happens and the amplitude of the
400 ensemble innovation signal increases indicating, possibly, that the lack of ensemble inflation due
401 to reduced number of conventional observation in the region. Innovation statistics for variables q
402 follows the trends figured for T , with increasing values over TR regions all along the simulations.
403 For the surface pressure, on the other hand, innovation statistics seems to decrease with time (in
404 both 00 and 12UTC), specially over SH region. This may be an indication that the ensemble spread
405 is not good there.

406 *c. Forecast Skill*

407 The forecast skill up to 5 days, was assessed in terms of the Anomaly Correlation (AC). As
408 CPTEC developments are focused in the Tropics and the the South America regions, results are
409 shown for these region along with for the whole Globe and North and South Hemispheres. To
410 assess how different the experiments are between each other, a t-Student test was also made with
411 a 95% of confidence level. For each plot of the AC, a t-Student test is presented and the lower
412 boxes shows whether or not the curves are different at the defined confidence level. When the
413 curves crosses their respective boxes, it means that the differences between the experiment (e.g.,
414 EnSRF50) and the experiment of reference (i.e., REF, the BAMv0 model integrated with the op-
415 erational NCEP analysis) are of no difference, indicating that the test for the null hypothesis fails.
416 In this case, the null hypothesis is that the mean of the hybrid experiment verified is statistically
417 indistinguishable from the mean of the experiment of reference.

418 Figure 4 shows the 5-day forecasts AC with a t-Student test for regions NH, TR and SH. Fig-
419 ure 5 shows the 5-day forecasts AC with the same t-Student test, but for the regions GL and SA.
420 Variables evaluated are the surface pressure ($psnm$), specific humidity at 925 hPa ($q925$), air tem-
421 perature at 850 hPa ($T850$), zonal wind component at 250 hPa ($u250$) and geopotential height at
422 500 hPa ($z500$). Experiments summarized in Table 1 were run with the same configuration (i.e.,
423 using the same code and options chosen through namelists).

424 Experiment EnSRF50 indicate the best analysis performance for all regions and variables eval-
425 uated. Surface pressure ($psnm$) at NH has practically no difference using either EnSRF75 or
426 EnSRF50 for the first 24 hours of forecasts.

427 In the TR region - a region specially difficult to predict for, forecast of $psnm$, $q925$ and $T850$
428 (Figures 4b,e,h) from EnSRF75 have shown a good performance with respect to experiments REF

429 and 3DVar. The improvements with respect to REF for *psnm* (Figure 4b), are of almost 24 hours
430 (considering 80% of the AC). Figure 4e, shows improvements of the experiment EnSRF75 with re-
431 spect to experiment 3DVar, in the order of almost 4 days in advance (considering 85% of the AC).
432 A similar improvement was found for the forecast of the *T850* (Figure 4h), with respect to experi-
433 ment to experiment 3DVar were experiment EnSRF75 still holds 80% of AC for 72 hours forecast,
434 whereas experiment 3DVar limits its forecast skill for this variable up to 36 hours. The zonal com-
435 ponent of the horizontal wind at 850 hPa (*u850* - Figures 4n), showed an slightly improvement over
436 the TR region. It is worth to note in Figures 4b,h,q the spin up time that the BAMv0 model took
437 to initialize with the GFS-NCEP analysis. However, this effect was not noticed for every evalu-
438 ated variable. Figure 4g, for example, indicates that there is no practical difference between the
439 experiments whose AC scores are very similar. The t-Student test, however, reveals that the 3DVar
440 experiment did not fail the null hypothesis test for the 120 hours forecasts, while the remaining of
441 the experiments holds the test up to 36 hours (EnSRF50) and 72 hours (EnSRF75). On the other
442 hand, however, experiment EnSRF75 was statistically different from the REF up to 72 hours.

443 Figure 5 shows the same evaluation as Figure 4, but for regions GL and SA. The SA region is
444 of great interest for CPTEC because it is Brazil's target forecast region. The same response from
445 experiments EnSRF75 and EnSRF50 were found, indicating that for the SA region, the ensemble
446 covariances can also play an important role in the determination of the forecasts skill. For the
447 GL region, AC scores are similar to what we have found on NH region. Besides that, the AC
448 for *q925* at GL region (Figure 5c), shows that the forecasts drawn from the EnSRF75 experiment
449 was good up to 120 hours. Remaining variables (i.e., *psnm*, *T850*, *u250* and *z500* - Figures
450 5a,e,g,i, respectively) showed similar AC scores in comparison with the NH region. For the SA
451 region, good results were found *psnm* (Figure 5b), for up to 84 hours forecasts; *q925* (Figure 5d),
452 with AC greater than 85% for 120 hours and *T850* (Figure 5f) with AC of 80% from experiment

453 EnSRF75 for up 72 hours forecasts. Zonal wind at 250 hPa (u_{250}) and, specially, z_{500} , did not
454 drawn good AC scores. The z_{500} (Figures 5k,l) from experiment EnSRF75 did not performed well
455 as for others variables. From 84 hours forecast ahead, its was the worse and experiment EnSRF50
456 - and even experiment 3DVar, were better.

457 A comparison between the analysis schemes used in the experiments, shows that the humidity
458 analysis drawn from the EnSRF50 experiment, was improved over the pure 3DVar analysis. For
459 these hybrid 3DVar experiments, background error covariance matrix is the same static covariance
460 matrix used within the 3DVar experiment. This is an indication that the ensemble covariance plays
461 an important role on defining how the analysis increments are applied.

462 PRECIPITATION ASSESSMENT

463 Precipitation forecasts drawn from the REF experiment and from the pure 3DVar and the hy-
464 brid 3DVar analysis has been also evaluated. Figure 6a shows the observed Global Precipitation
465 Climatology Project v2.2 (GPCPv2.2) 2.5° (Adler et al. 2003) global monthly mean precipitation
466 and the forecasted precipitation - represented as a monthly mean, drawn from the REF (Figure
467 6b), 3DVar (Figure 6c), hybrid 3DVar analyses with EnSRF50 (Figure 6d) and EnSRF75 (Figure
468 6e). The comparison between the 24 hour monthly mean forecasts and the GPCP 2.5 is valid
469 for January 2013 (12UTC) and the gridded data has been interpolated to a 2.5° grid resolution to
470 match GPCPv2.2. GPCP 24 hour monthly mean precipitation for January 2013 depicts the main
471 features of the tropical precipitation (from 8 mm above) and the subtropical precipitation (until 8
472 mm), with most of the convective precipitation being distributed over the tropical region.

473 The precipitation fields being compared in Figure 6a represents the monthly mean at 12UTC
474 beginning at 1st January 2013, until 31rd January 2013. All experiments shows the spatial features
475 of the large scale precipitation and the convective precipitation as well. Figure 6a is used as a

476 reference and the evaluation is made by comparing the precipitation intensity, distribution and the
477 spatial average of the monthly mean, annotated below each picture.

478 The precipitation forecasts from the experiments, shows a reasonable spatial distribution in com-
479 parison with GPCPv2.2 monthly mean precipitation. It is worth to note that the model configura-
480 tion for large scale condensation, cumulus convection and horizontal and vertical diffusion were
481 kept the same, so the differences are supposed to be due to the initial condition only. Main dif-
482 ferences between precipitation forecasts from experiments are mostly related to the intensity of
483 the precipitation, once the spatial distribution between them are fairly similar. Figure 6b, shows
484 the resulting monthly mean 24 hour precipitation from the BAMv0 model initialized with the
485 REF analysis. In comparison with any other experiment (including the GPCPv2.2 reference), the
486 BAMv0 model tended to produce more convective precipitation in the Tropical region. The spa-
487 tial average of the time mean amounts to 2.9718 mm/month, yielding the highest value among the
488 experiments, even higher than GPCPv2.2 (2.7197 mm/month). All experiments with data assimi-
489 lation cycle, on the other hand, have produced precipitation with less concentrated distribution and
490 a more reasonable spatial average. The spatial average in experiment EnSRF50 (Figure 6d) is the
491 one which more approximates GPCPv2.2, although the rest of the data assimilation experiments
492 has concentrated the convective precipitation in a more conformable way.

493 The exaggerated amount of precipitation produced by experiment REF must be due to fact that
494 this experiment was made by just running the BAMv0 model with the GFS-NCEP analysis, al-
495 though a topography smoothing and a initialization with normal modes has been made for every
496 analysis used in this specific experiment. The choices for the cumulus parametrization and the
497 diffusion parameters may also play an important role in the precipitation production. The exper-
498 iments with the data assimilation cycle, with the exception of experiment EnSRF75 (Figure 6e) -

499 which is the experiment that yielded the lower monthly mean, seems to be well conditioned with
500 the option chosen for the forecasts.

501 In order to better understand and to identify the regions where the experiments have accumulated
502 more or less the precipitation, Figure 7 shows the time series of the spatial averages for each
503 experiment, including the GPCP. In this figure, the GPCPv1.2 1DD (1 Degree Daily - Huffman
504 et al. 2001) was used and the observed data has been interpolated to the analysis grid (with 1.875°).
505 Accompanying Figure 7, Tables 5 and 6 shows the spatial averages (μ) and the standard deviations
506 (σ) for each experiment (including the interpolated GPCP1DD) at the regions evaluated (GL, NH,
507 TR, SH and SA).

508 In contrast to what we have found by analyzing the spatial distribution of the precipitation fore-
509 casts in Figure 6, the time series for the spatial averages showed in Figure 7, shows us that - with
510 the exception of region NH (Figure 7b), all experiments have some difficulties to reproduce the ob-
511 served precipitation in 24 hour forecasts. Deficiencies can be noted with respect to the amplitude
512 (maxima and minima) and to the average. In some specific cases, forecasted precipitation nearly
513 matched observed values - e.g., experiment EnSRF75 in TR region (Figure 7c) which matched the
514 GPCPv1.2 during some periods of the month. This results is with accordance to what we found
515 when analyzing the forecast skill (Figures 4 and 5), showing that among the hybrid 3DVar exper-
516 iments, the 75% of the EnSRF ensemble contribution to the static part of the background error
517 covariance was beneficial in most cases. On the other hand, experiment REF have exaggerated the
518 precipitation over time mainly for the GL and TR regions (Figures 7a and 7c). This is an expected
519 result, since this experiment have used an independent analysis and the model was not configured
520 to give the best result from it. For the TR and SA regions (Figures 7c,e), precipitation forecasts
521 were mostly above the observed values indicated by GPCPv1.2. In these regions, the convective
522 precipitation plays an important role and models tend to exaggerate the precipitation there. If we

523 consider the SH region (Figure 7d) - which contains more oceanic parts (in opposition to the NH
524 region), models tends to produce more large-scale precipitation whereas the convective precipita-
525 tion tends to be less important in the results. Consequently, the precipitation forecasts tends to be
526 below the observed values.

527 All experiments with the hybrid background error covariance matrix showed reasonable results
528 if compared to experiments 3DVar and REF. It appears, however, that the precipitation - specially
529 the convective precipitation, to be not so sensitive to changes in the covariance matrix, although
530 the experiment with greater ensemble contributions (EnSRF75) do perform better than the rest of
531 the experiments made.

532 **5. Summary and discussion**

533 In this work we have tested an implementation of the hybrid 3D variational (hybrid 3DVar)
534 data assimilation technique with the BAMv0 model. For this implementation, we carried out the
535 calculation of a static background error covariance matrix used in the linear combination with
536 the covariances from the ensemble part. To test the implementation, experiments were made in a
537 TQ0062L028 model and analysis resolution using an ensemble of analysis of 40 members, testing
538 the EnSRF algorithm. Results shows an general improvement of the winds, temperature, moisture
539 and surface pressure, mainly over South America and the Tropical region, which is a target region
540 for CPTEC operations. The moist part of the BAMv0 model (e.g., Specific Moisture, Precipitable
541 Water and Precipitation) resulted in a slight difference from the control run (the BAMv0 model
542 initialized with the GFS-NCEP analysis) and the from the BAMv0 model simulation initialized
543 with the pure 3DVar analysis.

544 In this work it is clear that the ensemble contributions to the static part of the covariance matrix
545 plays an important role in determining the skill of the analysis, and consequently the skill of the

546 forecasts. In some cases, we found improvements of more than 3 days, which we consider a great
547 step towards a robust analysis to use in daily operations. Furthermore, with this implementation,
548 we could not only show that our implementation is in the right way, but also that we could improve
549 the pure 3DVar analysis (despite some practical issues that have to be properly addressed). Our
550 results corroborates and contribute to the understanding of the application of hybrid background
551 error covariances matrices.

552 Further improvements to this work shall include the treatment of the caveats pointed out in
553 Section 2c. The realization of the hybrid 3DVar system at CPTEC operations will also depends
554 on the CPTEC capacity to keep the system at a higher resolution without compromise the system
555 performance and to afford the computational cost involved. A study regarding the dual resolution
556 feature for the hybrid system must be considered as the resolution of the ensemble limits the
557 system performance, and the use of the re-centering and ensemble inflation.

558 *Acknowledgments.* The authors are thankful to Ricardo Todling for the guidance and multiple
559 discussions during the implementation of the hybrid 3DVar system at CPTEC. The authors are
560 also thankful to the CPTEC staff and to the anonymous reviewers for the corrections and the sug-
561 gestions that improved the manuscript. To CAPES for the financial support in the PDSE program
562 (grant number BEX 99999.008036/2014-04).

563 **References**

564 Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project
565 (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*,
566 **4 (6)**, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2, URL [http://](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2)
567 [dx.doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2), [http://dx.doi.org/10.1175/](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2)
568 [1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).

- 569 Berre, L., M. Monteiro, and C. Pires, 2013: An Impact Study of Updating Background Error
570 Covariances in the ALADIN-France Data Assimilation System. *Journal of Geophysical Re-*
571 *search: Atmospheres*, **118 (19)**, 11,075–11,086, doi:10.1002/jgrd.50847, URL [http://dx.doi.org/](http://dx.doi.org/10.1002/jgrd.50847)
572 [10.1002/jgrd.50847](http://dx.doi.org/10.1002/jgrd.50847).
- 573 Cavalcanti, I. F. d. A., and Coauthors, 2002: Global Climatological Features in a Simulation using
574 the CPTEC-COLA AGCM. *Journal of Climate*, **15 (21)**, 2965–2988, URL [http://urlib.net/sid.](http://urlib.net/sid.inpe.br/iris@1915/2005/05.05.11.19)
575 [inpe.br/iris@1915/2005/05.05.11.19](http://urlib.net/sid.inpe.br/iris@1915/2005/05.05.11.19).
- 576 Chou, M., 1999: *A Solar Radiation Parameterization for Atmospheric Studies*. NASA technical
577 memorandum, URL https://books.google.com.br/books?id=_gYbvgAACAAJ.
- 578 Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational Implementation of a Hybrid
579 Ensemble/4D–Var Global Data Assimilation System at the Met Office. *Quarterly Journal of the*
580 *Royal Meteorological Society*, **139 (675)**, 1445–1461, doi:10.1002/qj.2054, URL [http://dx.doi.](http://dx.doi.org/10.1002/qj.2054)
581 [org/10.1002/qj.2054](http://dx.doi.org/10.1002/qj.2054).
- 582 Corazza, M., and Coauthors, 2003: Use of the Breeding Technique to Estimate the Structure of
583 the Analysis “errors of the day”. *Nonlinear Processes in Geophysics*, **10 (3)**, 233–243, doi:
584 [10.5194/npg-10-233-2003](http://www.nonlin-processes-geophys.net/10/233/2003/), URL <http://www.nonlin-processes-geophys.net/10/233/2003/>.
- 585 Cunningham, C., B. J. P., and M. Ferreira, 2015: Assessing Improved CPTEC Probabilistic
586 Forecasts on Medium-Range Timescale. *Meteorological Applications*, **22 (3)**, 378–384, doi:
587 [10.1002/met.1464](http://dx.doi.org/10.1002/met.1464), URL <http://dx.doi.org/10.1002/met.1464>.
- 588 Desroziers, G., 1997: A Coordinate Change for Data Assimilation in Spherical Ge-
589 ometry of Frontal Structures. *Monthly Weather Review*, **125 (11)**, 3030–3038, doi:
590 [10.1175/1520-0493\(1997\)125<3030:ACCFDA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1997)125<3030:ACCFDA>2.0.CO;2), URL <http://dx.doi.org/10.1175/>

591 1520-0493(1997)125(3030:ACCFDA)2.0.CO;2, [http://dx.doi.org/10.1175/1520-0493\(1997\)](http://dx.doi.org/10.1175/1520-0493(1997)125(3030:ACCFDA)2.0.CO;2)
592 125(3030:ACCFDA)2.0.CO;2.

593 Figueroa, S. N., and Coauthors, 2016: The Brazilian Global Atmospheric Model (BAM):
594 Performance for Tropical Rainfall Forecasting and Sensitivity to Convective Scheme
595 and Horizontal Resolution. *Weather and Forecasting*, **31** (5), 1547–1572, doi:10.1175/
596 WAF-D-16-0062.1, URL <http://dx.doi.org/10.1175/WAF-D-16-0062.1>, [http://dx.doi.org/10.](http://dx.doi.org/10.1175/WAF-D-16-0062.1)
597 1175/WAF-D-16-0062.1.

598 Fisher, M., 2003: Background Error Covariance Modelling. *Seminar on Recent Developments in*
599 *Data Assimilation for Atmosphere and Ocean, 8-12 September 2003*, ECMWF, Shinfield Park,
600 Reading, ECMWF, 45-64.

601 Grell, G. A., 1993: Prognostic Evaluation of Assumptions Used by Cumulus Parameteriza-
602 tions. *Monthly Weather Review*, **121** (3), 764–787, doi:10.1175/1520-0493(1993)121(0764:
603 PEOAUB)2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(1993\)121\(0764:PEOAUB\)2.0.](http://dx.doi.org/10.1175/1520-0493(1993)121(0764:PEOAUB)2.0.CO;2)
604 CO;2, [http://dx.doi.org/10.1175/1520-0493\(1993\)121\(0764:PEOAUB\)2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1993)121(0764:PEOAUB)2.0.CO;2).

605 Hamill, T. M., and C. Snyder, 2000: A Hybrid Ensemble Kalman Filter–3D Variational Anal-
606 ysis Scheme. *Monthly Weather Review*, **128** (8), 2905–2919, doi:10.1175/1520-0493(2000)
607 128(2905:AHEKFV)2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(2000\)128\(2905:](http://dx.doi.org/10.1175/1520-0493(2000)128(2905:AHEKFV)2.0.CO;2)
608 AHEKFV)2.0.CO;2.

609 Harshvardhan, R. Davies, D. A. Randall, and T. G. Corsetti, 1987: A Fast Radiation Param-
610 eterization for Atmospheric Circulation Models. *Journal of Geophysical Research: Atmo-*
611 *spheres*, **92** (D1), 1009–1016, doi:10.1029/JD092iD01p01009, URL [http://dx.doi.org/10.1029/](http://dx.doi.org/10.1029/JD092iD01p01009)
612 JD092iD01p01009.

- 613 Hayden, C. M., and R. J. Purser, 1995: Recursive Filter Objective Analysis of Meteorologi-
614 cal Fields: Applications to NESDIS Operational Processing. *Journal of Applied Meteorology*,
615 **34** (1), 3–15, doi:10.1175/1520-0450-34.1.3, URL <http://dx.doi.org/10.1175/1520-0450-34.1.3>,
616 <http://dx.doi.org/10.1175/1520-0450-34.1.3>.
- 617 Holtslag, A. A. M., and B. A. Boville, 1993: Local Versus Nonlocal Boundary–Layer
618 Diffusion in a Global Climate Model. *Journal of Climate*, **6** (10), 1825–1842, doi:
619 10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0442\(1993\)006<1825:LVNBLD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2),
620 [http://dx.doi.org/10.1175/1520-0442\(1993\)](http://dx.doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2)
621 [006<1825:LVNBLD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2).
- 622 Huffman, G. J., R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. Mc-
623 Gavock, and J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution
624 from Multisatellite Observations. *Journal of Hydrometeorology*, **2** (1), 36–50, doi:
625 10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2, URL [http://dx.doi.org/10.1175/1525-7541\(2001\)](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2)
626 [1525-7541\(2001\)002<0036:GPAODD>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2), [http://dx.doi.org/10.1175/1525-7541\(2001\)](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2)
627 [002<0036:GPAODD>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2).
- 628 Lorenc, A. C., 1986: Analysis Methods for Numerical Weather Prediction. *Quarterly Journal of*
629 *the Royal Meteorological Society*, **112** (474), 1177–1194, doi:10.1002/qj.49711247414, URL
630 <http://dx.doi.org/10.1002/qj.49711247414>.
- 631 Lorenc, A. C., 2003: Modelling of Error Covariances by 4D–Var Data Assimilation. *Quarterly*
632 *Journal of the Royal Meteorological Society*, **129** (595), 3167–3182, doi:10.1256/qj.02.131,
633 URL <http://dx.doi.org/10.1256/qj.02.131>.
- 634 Mellor, G. L., and T. Yamada, 1974: A Hierarchy of Turbulence Closure Models for
635 Planetary Boundary Layers. *Journal of the Atmospheric Sciences*, **31** (7), 1791–1806,

636 doi:10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2),
637 [http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2),
638 [http://dx.doi.org/10.1175/1520-0469\(1974\)031<1791:AHOTCM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1974)031<1791:AHOTCM>2.0.CO;2).

639 Mendonça, A. M., and J. P. Bonatti, 2009: Experiments with EOF–Based Perturbation Meth-
640 ods and Their Impact on the CPTEC/INPE Ensemble Prediction System. *Monthly Weather*
641 *Review*, **137** (4), 1438–1459, doi:10.1175/2008MWR2581.1, URL <http://dx.doi.org/10.1175/2008MWR2581.1>,
642 <http://dx.doi.org/10.1175/2008MWR2581.1>.

643 Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center’s Spectral
644 Statistical-Interpolation Analysis System. *Monthly Weather Review*, **120** (8), 1747–1763,
645 doi:10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2, URL [http://dx.doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2),
646 [http://dx.doi.org/10.1175/1520-0493\(1992\)120<1747:TNMCSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2).

647 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003a: Numerical Aspects of the
648 Application of Recursive Filters to Variational Statistical Analysis. Part I: Spatially Homo-
649 geneous and Isotropic Gaussian Covariances. *Monthly Weather Review*, **131** (8), 1524–1535,
650 doi:10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2, URL [http://dx.doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2),
651 [http://dx.doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2),
652 [http://dx.doi.org/10.1175//1520-0493\(2003\)131<1524:NAOTAO>2.0.CO;2](http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2).

653 Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003b: Numerical Aspects of the
654 Application of Recursive Filters to Variational Statistical Analysis. Part II: Spatially Inhomo-
655 geneous and Anisotropic General Covariances. *Monthly Weather Review*, **131** (8), 1536–1548,
656 doi:10.1175//2543.1, URL <http://dx.doi.org/10.1175//2543.1>.

- 657 Rabier, F., 2005: Overview of Global Data Assimilation Developments in Numerical Weather-
658 Prediction Centres. *Quarterly Journal of the Royal Meteorological Society*, **131 (613)**, 3215–
659 3233, doi:10.1256/qj.05.129, URL <http://dx.doi.org/10.1256/qj.05.129>.
- 660 Riishøjgaard, L. P., 1998: A Direct Way of Specifying Flow-Dependent Background Error Corre-
661 lations for Meteorological Analysis Systems. *Tellus A*, **50 (1)**, 42–57, doi:10.1034/j.1600-0870.
662 1998.00004.x, URL <http://dx.doi.org/10.1034/j.1600-0870.1998.00004.x>.
- 663 Sellers, P. J., and Coauthors, 1996: A Revised Land Surface Parameterization
664 (SiB2) for Atmospheric GCMs. Part I: Model Formulation. *Journal of Climate*,
665 **9 (4)**, 676–705, doi:10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2, URL
666 [http://dx.doi.org/10.1175/1520-0442\(1996\)009<0676:ARLSPF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2), [http://dx.doi.org/10.1175/1520-0442\(1996\)009<0676:ARLSPF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(1996)009<0676:ARLSPF>2.0.CO;2).
- 668 Thpaut, J.-N., and P. Courtier, 1991: Four-Dimensional Variational Data Assimilation using the
669 Adjoint of a Multilevel Primitive-Equation Model. *Quarterly Journal of the Royal Meteorolo-*
670 *gical Society*, **117 (502)**, 1225–1254, doi:10.1002/qj.49711750206, URL <http://dx.doi.org/10.1002/qj.49711750206>.
- 671
- 672 Tiedtke, M., 1983: The Sensitivity of the Time-Mean Large-Scale Flow to Cumulus Convection in
673 the ECMWF Model. *Workshop on Convection in Large-scale Numerical Models, 28 November*
674 *to 1 December 1983*, ECMWF, Shinfield Park, Reading, ECMWF, 297-316.
- 675 Wang, X., 2010: Incorporating Ensemble Covariance in the Gridpoint Statistical Interpolation
676 Variational Minimization: A Mathematical Framework. *Monthly Weather Review*, **138 (7)**,
677 2990–2995, doi:10.1175/2010MWR3245.1, URL <http://dx.doi.org/10.1175/2010MWR3245.1>,
678 <http://dx.doi.org/10.1175/2010MWR3245.1>.

- 679 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008a: A Hybrid ETKF–3DVAR Data
680 Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment.
681 *Monthly Weather Review*, **136** (12), 5116–5131, doi:10.1175/2008MWR2444.1, URL [http://dx.
doi.org/10.1175/2008MWR2444.1](http://dx.
682 doi.org/10.1175/2008MWR2444.1).
- 683 Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill, 2008b: A Hybrid ETKF–3DVAR Data
684 Assimilation Scheme for the WRF Model. Part II: Real Observation Experiments. *Monthly
685 Weather Review*, **136** (12), 5132–5147, doi:10.1175/2008MWR2445.1, URL [http://dx.doi.org/
10.1175/2008MWR2445.1](http://dx.doi.org/
686 10.1175/2008MWR2445.1).
- 687 Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007: A Comparison of Hybrid Ensem-
688 ble Transform Kalman Filter–Optimum Interpolation and Ensemble Square Root Filter Anal-
689 ysis Schemes. *Monthly Weather Review*, **135** (3), 1055–1076, doi:10.1175/MWR3307.1, URL
690 <http://dx.doi.org/10.1175/MWR3307.1>, <http://dx.doi.org/10.1175/MWR3307.1>.
- 691 Wang, X., T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2009: A Comparison of the Hybrid and
692 EnSRF Analysis Schemes in the Presence of Model Errors due to Unresolved Scales. *Monthly
693 Weather Review*, **137** (10), 3219–3232, doi:10.1175/2009MWR2923.1, URL [http://dx.doi.org/
10.1175/2009MWR2923.1](http://dx.doi.org/
694 10.1175/2009MWR2923.1), <http://dx.doi.org/10.1175/2009MWR2923.1>.
- 695 Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-Based Ensemble-Variational
696 Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments.
697 *Monthly Weather Review*, **141** (11), 4098–4117, doi:10.1175/MWR-D-12-00141.1, URL [http://
dx.doi.org/10.1175/MWR-D-12-00141.1](http://
698 dx.doi.org/10.1175/MWR-D-12-00141.1), <http://dx.doi.org/10.1175/MWR-D-12-00141.1>.
- 699 Whitaker, J. S., and T. M. Hammill, 2002: Ensemble Data Assimilation Without
700 Perturbed Observations. *Monthly Weather Review*, **130** (7), 1913–1924, doi:10.1175/

701 1520-0493(2002)130<1913:EDAWPO>2.0.CO;2, [http://dx.doi.org/10.1175/1520-0493\(2002\)](http://dx.doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2)
702 130<1913:EDAWPO>2.0.CO;2.

703 Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-Dimensional Variational
704 Analysis with Spatially Inhomogeneous Covariances. *Monthly Weather Review*,
705 **130** (12), 2905–2916, doi:10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2,
706 URL [http://dx.doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2), [http:](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2)
707 [//dx.doi.org/10.1175/1520-0493\(2002\)130<2905:TDVAWS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2).

708 Zupanski, M., 2005: Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Monthly*
709 *Weather Review*, **133** (6), 1710–1726, doi:10.1175/MWR2946.1, URL [http://dx.doi.org/10.](http://dx.doi.org/10.1175/MWR2946.1)
710 [1175/MWR2946.1](http://dx.doi.org/10.1175/MWR2946.1), <http://dx.doi.org/10.1175/MWR2946.1>.

711 **LIST OF TABLES**

712 **Table 1.** Hybrid 3DVar experiments and covariance matrix setup. 36

713 **Table 2.** Observational data used in the experiments with a cycled analysis. 37

714 **Table 3.** BAMv0 model setup (main options) used within the data assimilation cycles. 38

715 **Table 4.** Maximum and minimum values of Amplitudes, Horizontal and Vertical
716 Length-Scales of the Static TQ0062L028 background error covariance matrix. 39

717 **Table 5.** Area average of 24 hour (12UTC) precipitation (in mm/month) for the Globe
718 (GL) and South America (SA). 40

719 **Table 6.** Area average of 24 hour (12UTC) precipitation (in mm/month) for the North
720 Hemisphere (NH), Tropics (TR) and South Hemisphere (SH). 41

TABLE 1. Hybrid 3DVar experiments and covariance matrix setup.

Experiment	Hybrid Setup	Description
REF	–	based on GFS-NCEP
3DVar	–	based on BAMv0-CPTEC
EnSRF50	50% Static/Ens. (40 mem.)	based on BAMv0-CPTEC
EnSRF75	75% Ens. (40 mem.)	based on BAMv0-CPTEC

TABLE 2. Observational data used in the experiments with a cycled analysis.

Mnemonic	Type	Description
airsbufr	Unconventional	AMSU-A/AIRS radiances from AQUA
amsuabufr	Unconventional	AMSU-A 1b radiances (brightness temperature) from NOAA-15, 16, 17, 18, 19 and METOP-A
hirs4bufr	Unconventional	HIRS4 1b radiances from NOAA 18, 19 and METOP-A
mhsbufr	Unconventional	MHS moisture sounders MHS from NOAA 18, 19 and METOP-A
iasibufr	Unconventional	IASI sounders from METOP-A
gpsrobufr	Conventional	GPS radio occultation refractivities
prepbufr	Conventional	<i>ps, T, q, pw, uv, spd, dw</i> and <i>sst</i> observations

TABLE 3. BAMv0 model setup (main options) used within the data assimilation cycles.

Option	Setup
Resolution	TQ0062L028
Integration timestep	1200 seconds
Wind transport	Eulerian
Mass transport	Semi-Lagrangean
Initialization	Diabatic
Restart	Surface
Mass conservation	$\ln(p)$
LW radiation	Harshvardhan et al. (1987)
SW radiation	CLiRAD (Chou 1999)
Cumulus convection	Grell (1993)
Shallow convection	Tiedtke (1983)
Top of PBL	Holtslag and Boville (1993)
Bottom of PBL	Mellor and Yamada (1974)
Surface model	SSiB (Sellers et al. 1996)

721 TABLE 4. Maximum and minimum values of Amplitudes, Horizontal and Vertical Length-Scales of the Static
722 TQ0062L028 background error covariance matrix.

	Amplitudes		Horizontal L-S		Vertical L-S	
	Min	Max	Min	Max	Min	Max
ψ	7.60×10^5	6.83×10^6	4.59×10^5	1.31×10^6	1.00×10^{-2}	1.48×10^0
χ	0	5.05×10^6	0	1.95×10^6	0	1.60×10^0
t	0	2.43×10^0	0	7.49×10^5	0	1.77×10^0
q	0	6.92×10^{-1}	0	7.29×10^{-1}	0	6.61×10^5
oz	0	1.76×10^6	0	5.90×10^5	0	5.30×10^0
cw	0	1.95×10^6	0	5.00×10^5	0	4.91×10^0
ps	6.84×10^{-2}	2.29×10^{-1}	3.56×10^5	5.26×10^5	–	–
sst	2.00×10^{-2}	4.32×10^{-1}	5.07×10^2	8.00×10^2	–	–

723 TABLE 5. Area average of 24 hour (12UTC) precipitation (in mm/month) for the Globe (GL) and South
 724 America (SA).

	GL		SA	
	μ	σ	μ	σ
GPCP	2.8635	0.1123	3.1321	0.7743
REF	3.0478	0.1036	5.2812	0.7786
3DVar	2.7601	0.1118	5.0152	0.8779
EnSRF50	2.8103	0.0935	4.6480	0.7768
EnSRF75	2.5745	0.1089	3.8128	0.6702

725 TABLE 6. Area average of 24 hour (12UTC) precipitation (in mm/month) for the North Hemisphere (NH),
 726 Tropics (TR) and South Hemisphere (SH).

	NH		TR		SH	
	μ	σ	μ	σ	μ	σ
GPCP	1.7228	0.2653	4.2820	0.2694	2.6068	0.2053
REF	1.5547	0.2842	5.4655	0.3733	2.0758	0.2685
3DVar	1.4344	0.2438	4.7657	0.3420	2.0499	0.2512
EnSRF50	1.5605	0.2779	4.8385	0.2990	1.9999	0.2546
EnSRF75	1.4566	0.2520	4.2710	0.2209	1.9756	0.2619

727 **LIST OF FIGURES**

728 **Fig. 1.** Schematic diagram depicting the hybrid 3DVar analysis cycle for the BAMv0 model. 43

729 **Fig. 2.** Standard deviations of the variance errors distribution for ψ [$10^6 m^2 s^{-1}$], χ [$10^6 m^2 s^{-1}$], T
730 [K], q , oz and ps [$ln(hPa)$] throughout latitudes and vertical sigma levels. 44

731 **Fig. 3.** Ensemble innovation statistics for the EnSRF experiments: red (blue) color represent hybrid
732 analyses with 50% (75%) of EnKF covariances contribution. Upper (bottom) panel plots are
733 valid for 00UTC (12UTC) Dashed lines represents the priors and solid lines represents the
734 posteriors. Left column represent the North Hemisphere (NH), middle the Tropics (TR) and
735 right column, represent the South Hemisphere (SH). 45

736 **Fig. 4.** 5-day forecast skill for regions NH, TR and SH. Dashed black line represents the REF
737 experiment analyses and forecasts; solid red line, the 3DVar experiment; magenta line, the
738 EnSRF50 and cyan line, the EnSRF75. All figures are presented with their respective t-
739 Student significance test, where a 95% confidence interval is hold. 46

740 **Fig. 5.** Same as Figure 4, but for the GL and SA regions. 47

741 **Fig. 6.** Monthly average precipitation 24 hour forecast valid for January 2013 at 12UTC. Compari-
742 son is made between monthly mean precipitation from GPCPv2.2 and all experiments using
743 different analyses. In a) GPCP v2.2; b) BAM with the REF analysis; c) BAM with 3DVar;
744 d) BAM with hybrid 3DVar 50% and e) BAM with hybrid 3DVar 75%. Below each pic-
745 ture, it is annotated the spatial average of the monthly mean of the precipitation forecasts, in
746 mm/month. 48

747 **Fig. 7.** Spatial averages of the 24 hour forecast precipitation (in mm/day) valid for January 2013
748 at 12Z. Panel a) precipitation for the whole globe (GL); b) same for the North Hemi-
749 sphere (NH); c) for Tropics (TR); d) for South Hemisphere (SH) and e) for South Amer-
750 ica (SA). Precipitation from experiments are compared with the observed precipitation from
751 GPCPv1.2. 49

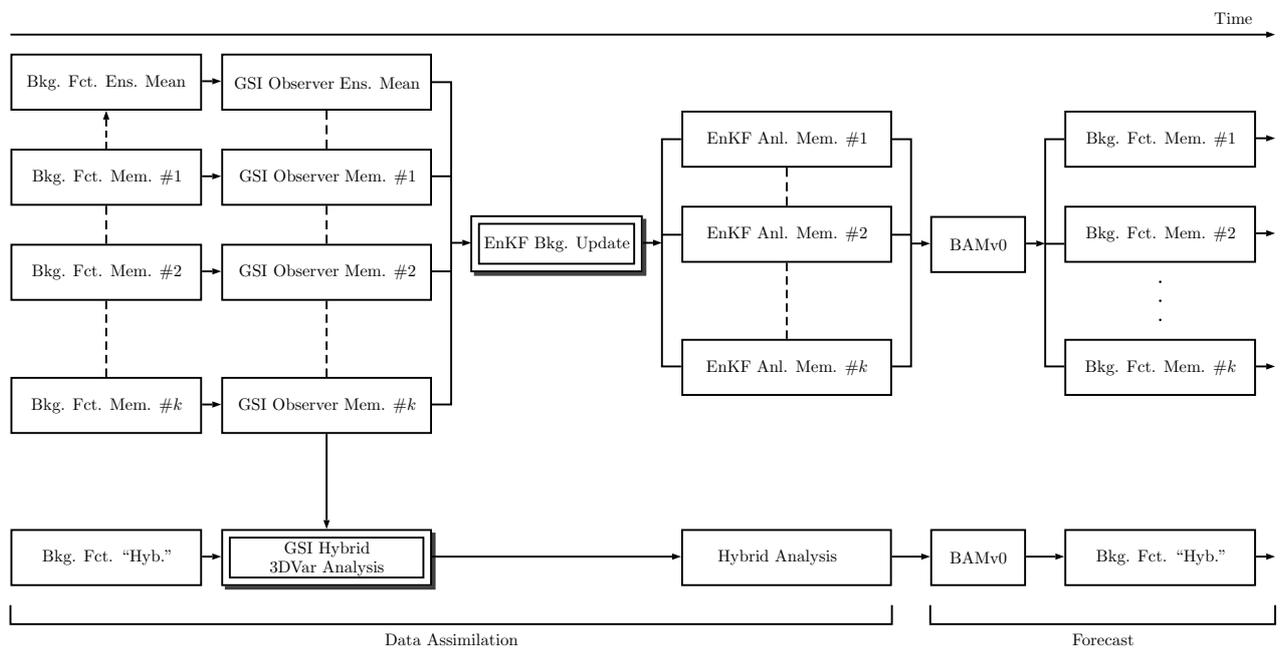
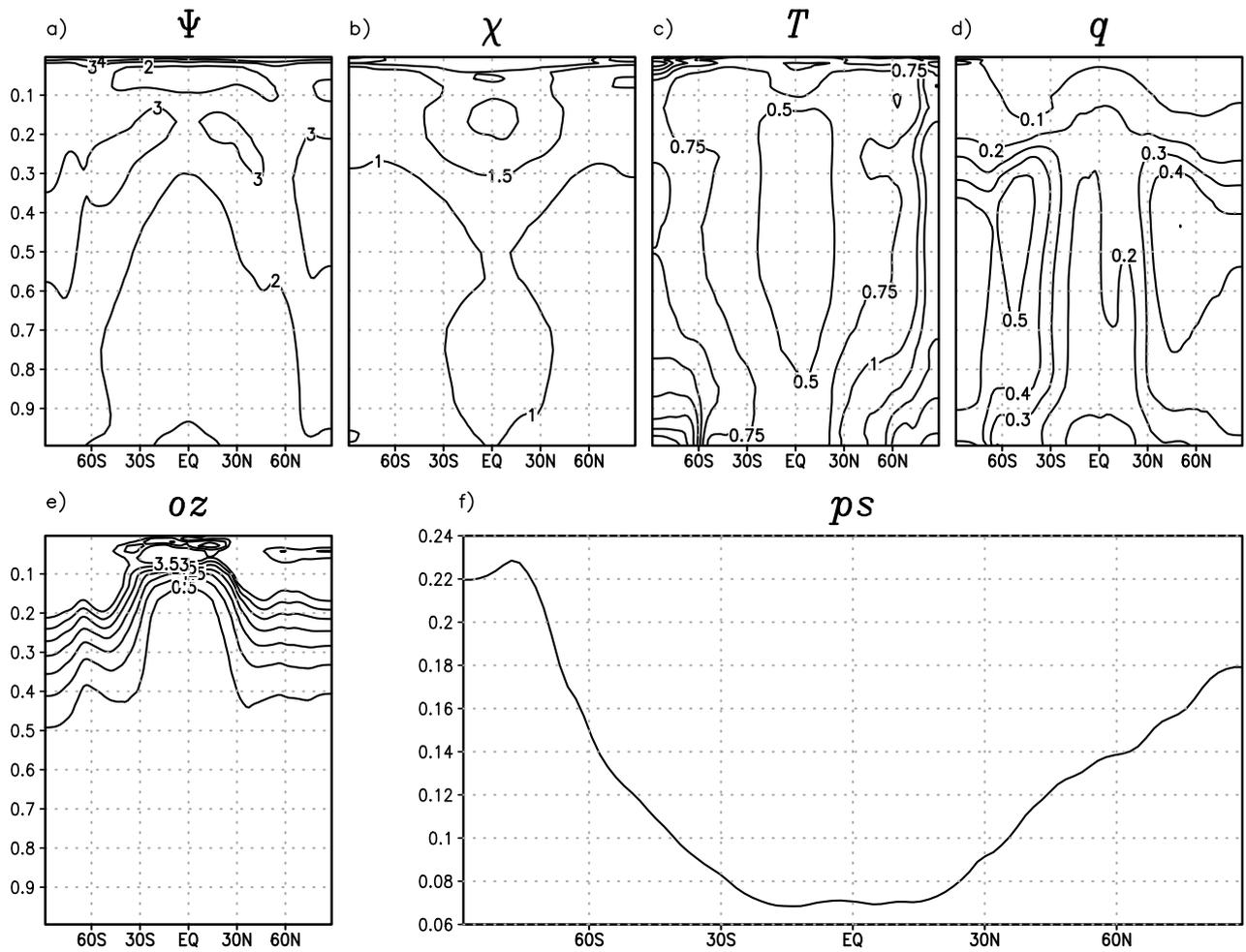
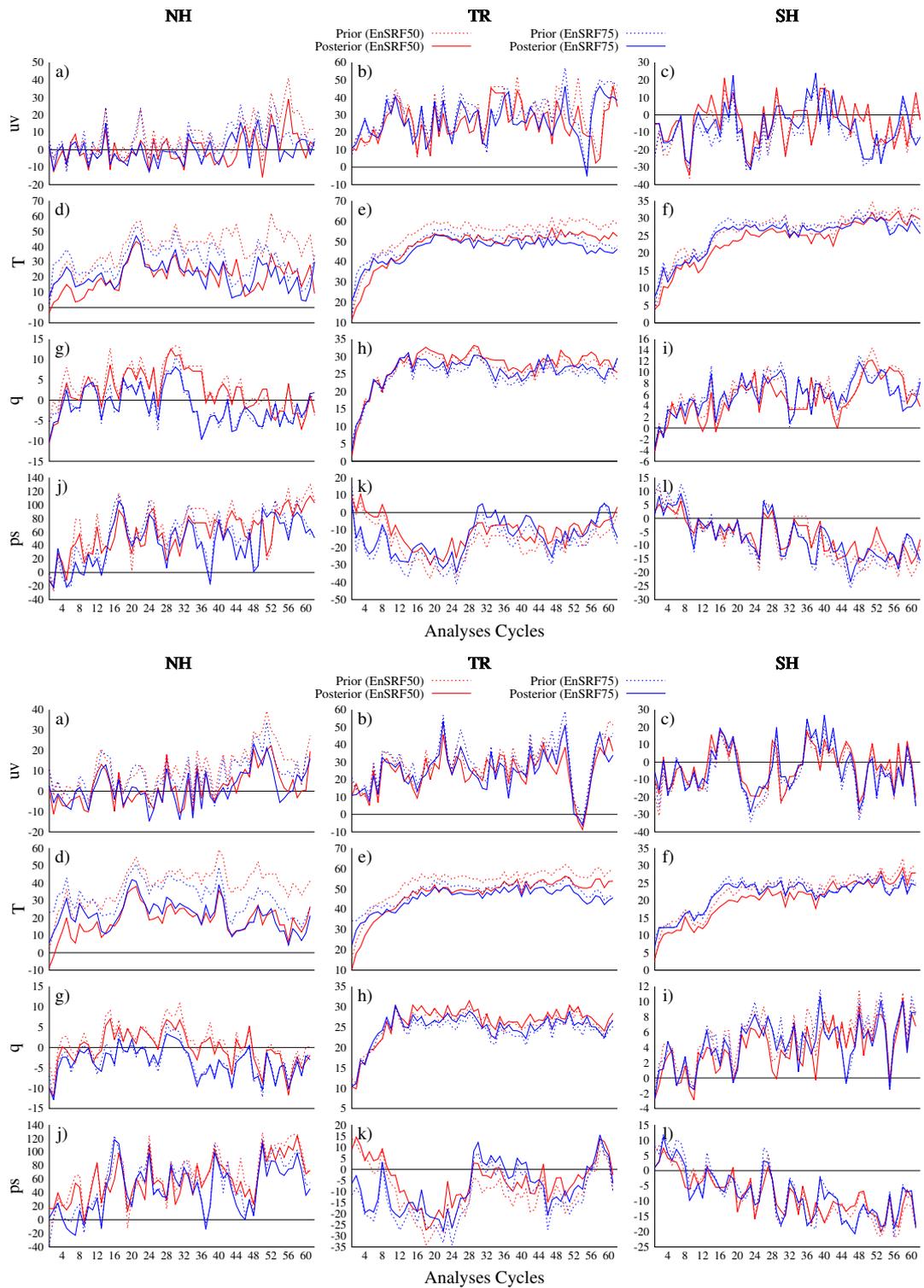


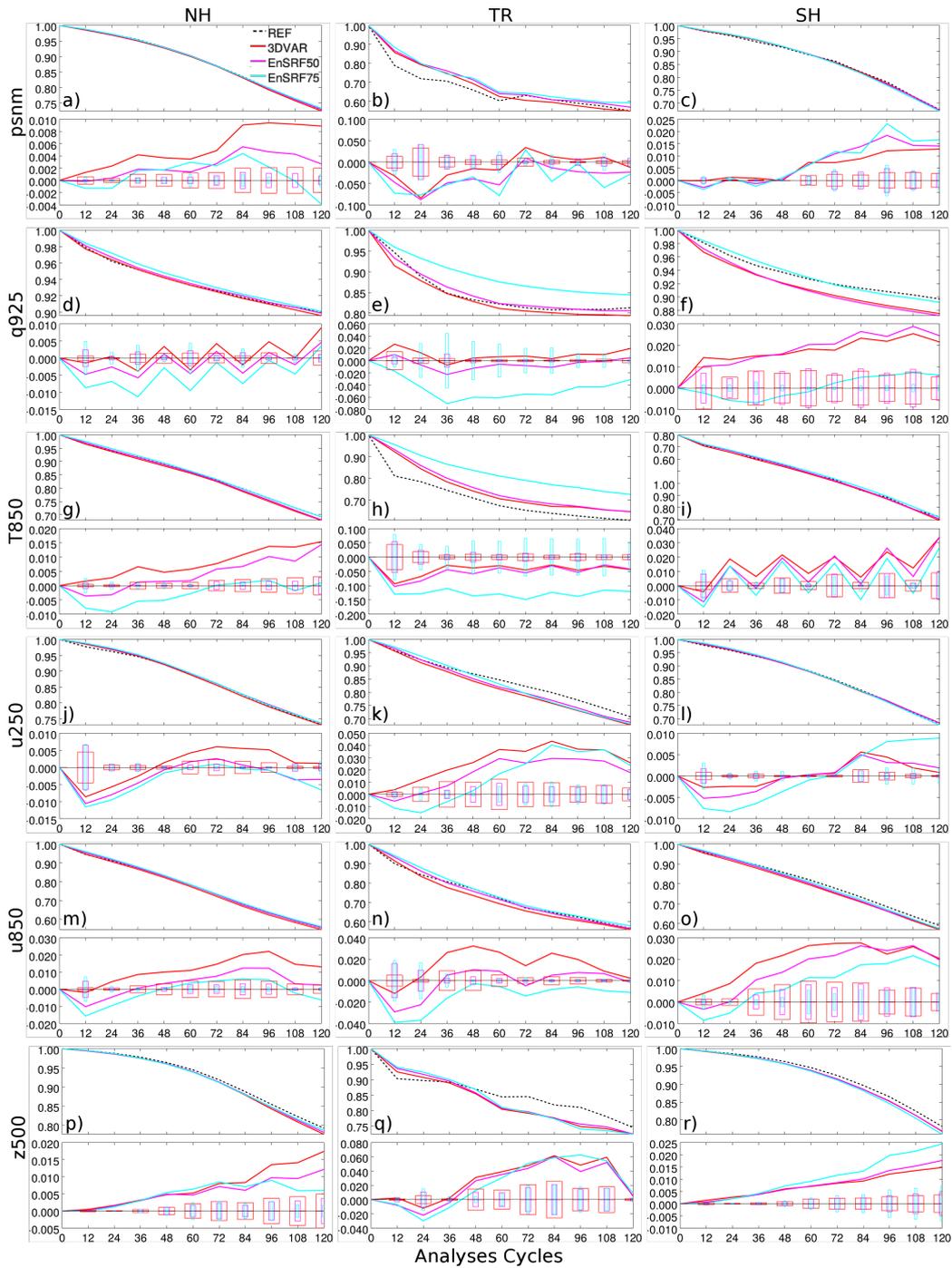
FIG. 1. Schematic diagram depicting the hybrid 3DVar analysis cycle for the BAMv0 model.



752 FIG. 2. Standard deviations of the variance errors distribution for ψ [$10^6 m^2 s^{-1}$], χ [$10^6 m^2 s^{-1}$], T [K], q , oz
 753 and ps [$\ln(hPa)$] throughout latitudes and vertical sigma levels.



754 FIG. 3. Ensemble innovation statistics for the EnSRF experiments: red (blue) color represent hybrid analyses
 755 with 50% (75%) of EnKF covariances contribution. Upper (bottom) panel plots are valid for 00UTC (12UTC)
 756 Dashed lines represents the priors and solid lines represents the posteriors. Left column represent the North
 757 Hemisphere (NH), middle the Tropics (TR) and right column, represent the South Hemisphere (SH).



758 FIG. 4. 5-day forecast skill for regions NH, TR and SH. Dashed black line represents the REF experiment
 759 analyses and forecasts; solid red line, the 3DVar experiment; magenta line, the EnSRF50 and cyan line, the
 760 EnSRF75. All figures are presented with their respective t-Student significance test, where a 95% confidence
 761 interval is hold.

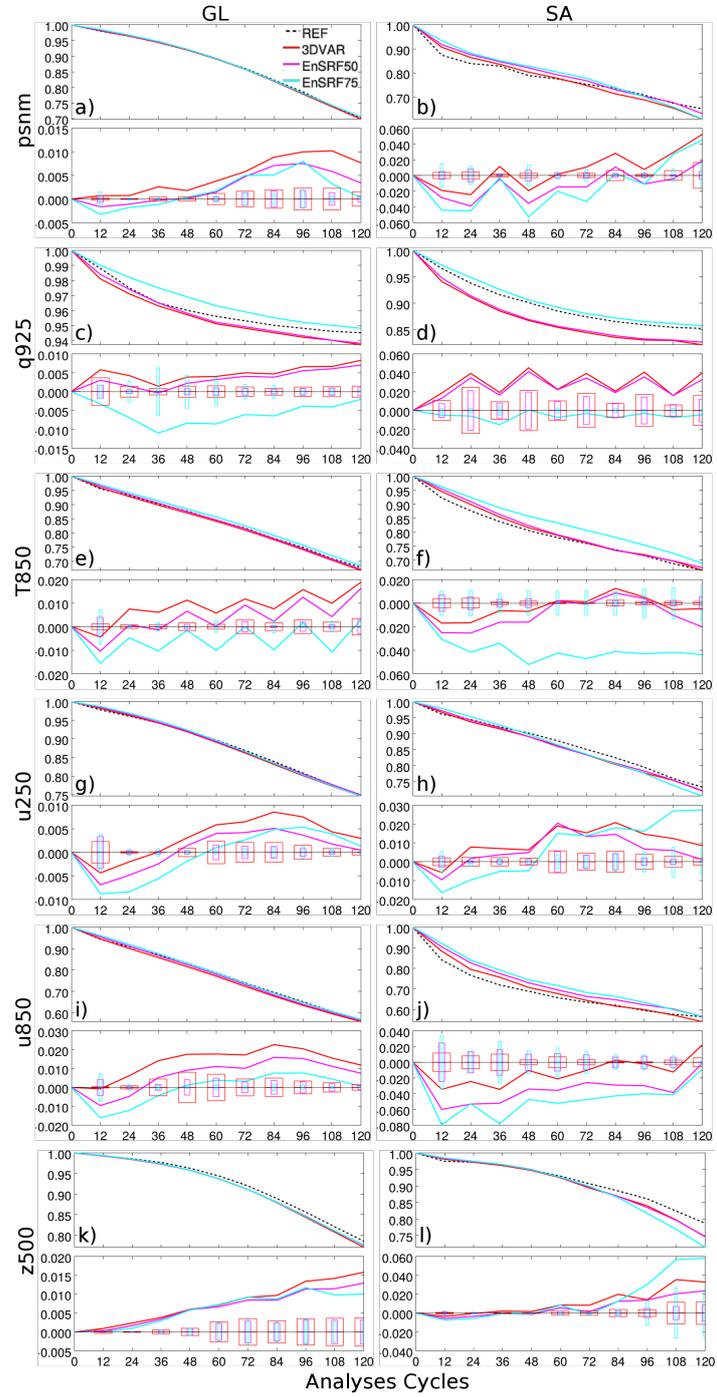
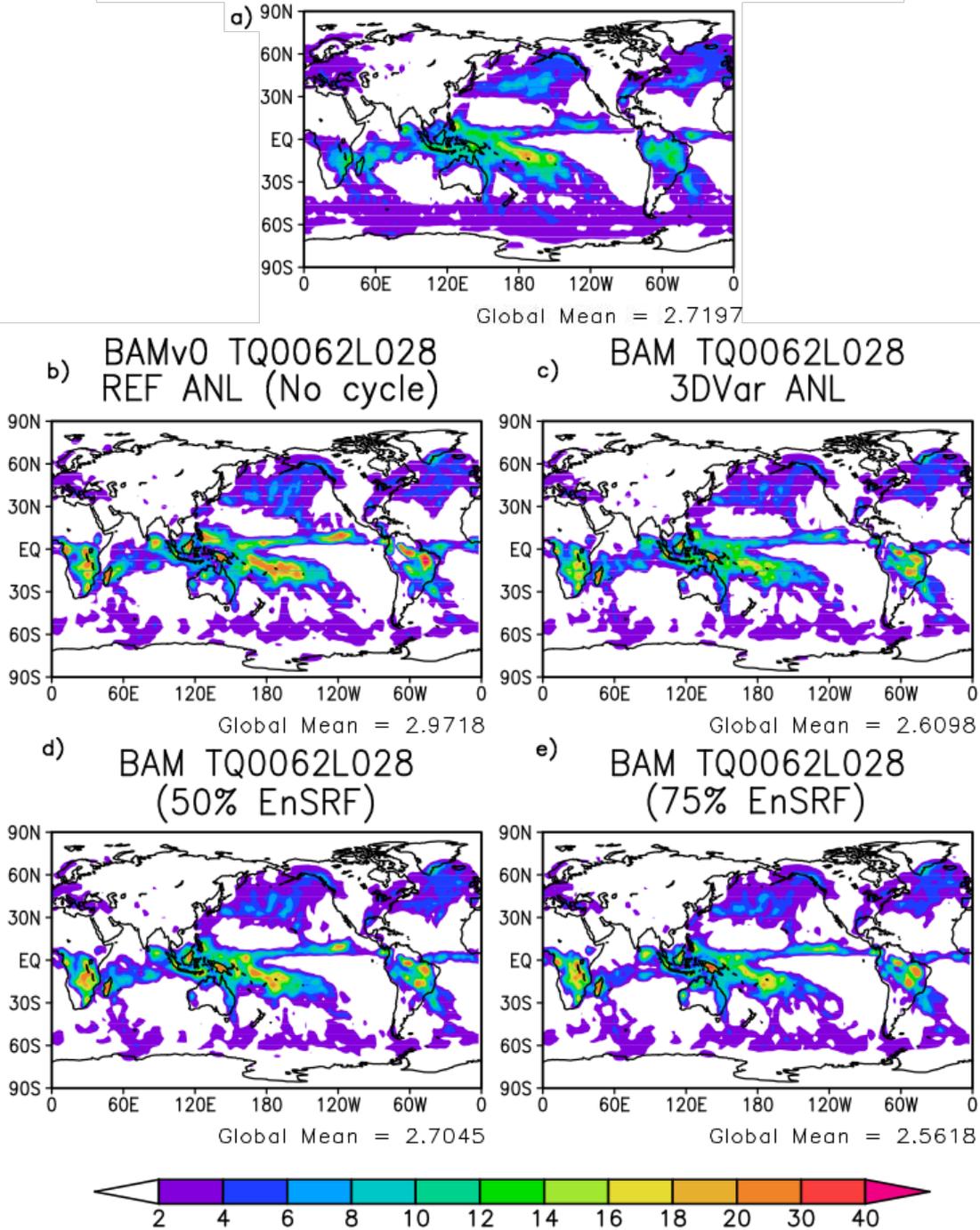


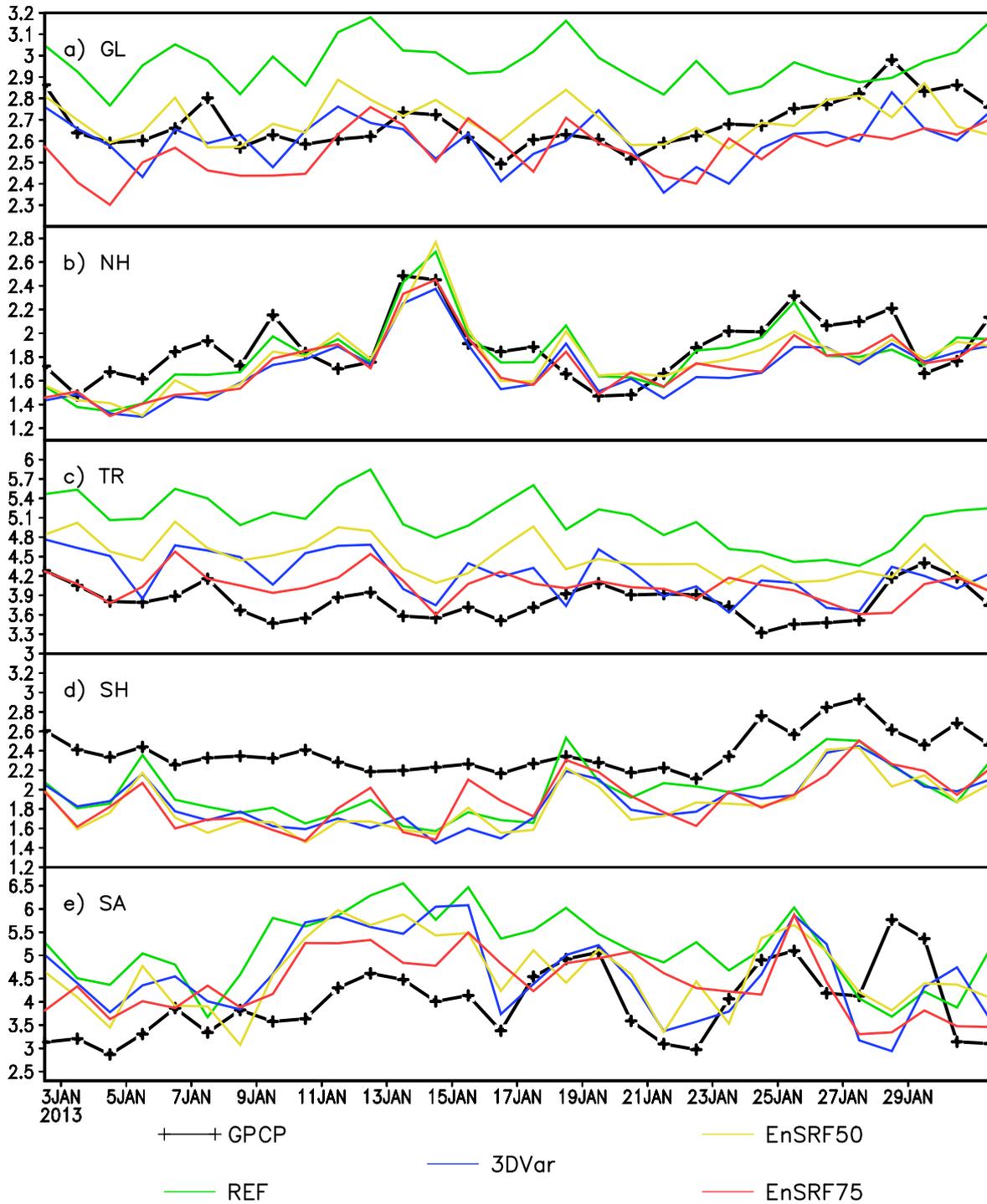
FIG. 5. Same as Figure 4, but for the GL and SA regions.

GPCP 2.5 24h Precip. Ave. (January 2013)



762 FIG. 6. Monthly average precipitation 24 hour forecast valid for January 2013 at 12UTC. Comparison is
 763 made between monthly mean precipitation from GPCPv2.2 and all experiments using different analyses. In a)
 764 GPCP v2.2; b) BAM with the REF analysis; c) BAM with 3DVar; d) BAM with hybrid 3DVar 50% and e) BAM
 765 with hybrid 3DVar 75%. Below each picture, it is annotated the spatial average of the monthly mean of the
 766 precipitation forecasts, in mm/month.

Area Average 24 hr Precip.
12z01Jan2013 12z31Jan2013



767 FIG. 7. Spatial averages of the 24 hour forecast precipitation (in mm/day) valid for January 2013 at 12Z. Panel
 768 a) precipitation for the whole globe (GL); b) same for the North Hemisphere (NH); c) for Tropics (TR); d) for
 769 South Hemisphere (SH) and e) for South America (SA). Precipitation from experiments are compared with the
 770 observed precipitation from GPCPv1.2.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o International Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.