

Artigo

## Matriz de Covariâncias dos Erros de Previsão Aplicada ao Sistema de Assimilação de Dados Global do CPTEC: Experimentos com Observação Única

Carlos Frederico Bastarz<sup>1</sup>, Dirceu Luis Herdies<sup>1</sup>, Luiz Fernando Sapucci<sup>1</sup><sup>1</sup>*Divisão de Modelagem e Desenvolvimento, Centro de Previsão de Tempo e Estudos Climáticos, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil*

Recebido em 22 de Setembro de 2016 – Aceito em 17 de Abril de 2017

### 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.

## Background Error Covariance Matrix Applied to the Global Data Assimilation System at CPTEC: Single Observation Experiments

### 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 **R**, e.g., 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 **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 (i.e., 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 soluções 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 da 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. Conseqüentemente, 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 expressá-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 *Gridpoint 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. Materiais 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 1.

### 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 (Lorenç, 1986).

No sistema 3DVar do GSI, a função custo representada (de forma simplificada) pela Eq. (1):

$$J(x) = \frac{1}{2} (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + \frac{1}{2} (y - \mathbf{H}x)^T \mathbf{R}^{-1} (y - \mathbf{H}x) \quad (1)$$

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 ( $p \times p$ ).

A função custo representada pela Eq. (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 (i.e., 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). A análise é calculada a partir do mínimo da

**Tabela 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	Holstlag e Boville
Restart	Superfície	Base da Camada Limite	Mellor e Yamada
Conservação de Massa	$ln(p)$	Modelo de Superfície	SSiB

função custo representada pela Eq. (1), quando  $\nabla J = 0$  implicando em  $x = x_a$ .

$$x_a = x_b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} [(\mathbf{H}^T \mathbf{R}^{-1})(y - \mathbf{H}x_b)] \quad (2)$$

Na Eq. (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, i.e.,  $\mathbf{R}^{-1}$  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. (1), pode ser escrito como:

$$\mathbf{H} = [0, \dots, 0, 1, 0, \dots, 0] \quad (3)$$

O gradiente da função custo, é então calculado aplicando-se o operador gradiente ( $\nabla J$ ) diretamente na Eq. (1) e igualando-se a expressão resultante a zero para se determinar o mínimo da função (Eq. (4)). Observe que neste caso, não se está considerando a hipótese de linearização do operador observação (i.e.,  $y - \mathbf{H}x + \mathbf{H}(x - x_b)$ ), o que acarretaria no peso obtido na expressão do lado direito da Eq. (2). Resolvendo a equação resultante para o incremento de análise  $x - x_b$  (Eq. (6)), obtemos:

$$\nabla J(x) = \mathbf{B}^{-1}(x - x_b) + \mathbf{H}^T \mathbf{R}^{-1}(y - \mathbf{H}x) = 0 \quad (4)$$

$$x - x_b = \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1}(y - \mathbf{H}x) \quad (5)$$

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:

$$\delta x \propto \mathbf{B} \mathbf{H}^T \quad (6)$$

ou seja, o incremento de análise  $\delta x$  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  $\delta x$  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 representam 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 h e 24 h 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 h) e 2013123000 (válido para uma previsão de 48 h). 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 h e 24 h.

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 h e 24 h 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 h e 24 h;
- Remoção de viés (em toda a coluna vertical);
- Cálculo das matrizes de balanço que permitirão as transformações entre função de corrente ( $\psi$ ) e as componentes desbalanceadas de velocidade potencial ( $\chi$ ), pressão em superfície ( $p$ ) e temperatura ( $T$ );
- Cálculo das variâncias dos erros de cada uma das variáveis de controle ( $\psi$ ,  $\chi$ ,  $q$ ,  $oz$ ,  $cw$ ,  $p$ );
- 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 (Fig. 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 (i.e., 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.

Na prática, a matriz de covariâncias do GSI possui as seguintes estruturas: amplitudes (representadas pelos desvios-padrão, variâncias e covariâncias), distâncias (repre-

sentadas 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. (7) e (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. (9)) é utilizada para contabilizar a correlação positiva entre divergência e vorticidade.

$$T_b = \mathbf{G}\phi \quad (7)$$

$$P_b = \mathbf{W}\phi \quad (8)$$

$$\chi_b = \mathbf{C}\phi \quad (9)$$

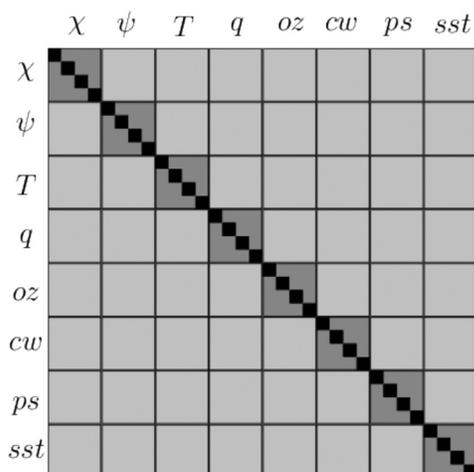
Esta é a forma utilizada pelo GSI para aplicar os incrementos de análise de forma filtrada, i.e., decompõe-se as variáveis de controle em uma parte balanceada (referente à parte lenta do fluxo atmosférico) e em uma parte não balanceada (referente à parte rápida e ruidosa do fluxo atmosférico) e então 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 Fig. 1) e, ao invés disso, filtros recursivos (Purser *et al.*, 2003a, 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:

$$\mathbf{B} = \mathbf{B}_z (V^1 \mathbf{B}_x^1 \mathbf{B}_y^1 \mathbf{B}_y^1 \mathbf{B}_x^1 V^1 + V^2 \mathbf{B}_x^2 \mathbf{B}_y^2 \mathbf{B}_y^2 \mathbf{B}_x^2 V^2) \mathbf{B}_z \quad (10)$$

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 (i.e., 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).



**Figura 1** - Representação idealizada da matriz  $\mathbf{B}$ : na diagonal principal (quadrados pretos), estão os elementos de variâncias de cada variável de estado; em cinza escuro, estão os elementos de autocovariância (i.e., variância da mesma variável em pontos de grade diferentes) e em cinza claro, os elementos de covariância multivariados (Adaptado de Petrie, 2012).

Um outro aspecto tratado durante o cálculo da matriz de covariância é o viés do modelo numérico. O viés ( $\beta^t$ ) é uma medida do erro sistemático do modelo de PNT e pode ser representada pelo modelo ( $\omega^f$ ) subtraído do erro real ( $\omega^t$ ) e de um erro aleatório, cuja esperança é zero, i.e.,  $\langle \varepsilon^t \rangle = 0$  (Chepurin *et al.*, 2005). A remoção de viés (Eq. 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.

$$\beta^t = \omega^f - \omega^t - \varepsilon^t \quad (11)$$

Os comprimentos de escala horizontais são importantes estruturas da matriz  $\mathbf{B}$  e são dependentes dos valores de variâncias calculados - Eq. (12), onde  $V_\psi$  e  $V_\zeta$ , 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).

$$L = \left( \frac{8V_\psi}{V_\zeta} \right)^{\frac{1}{4}} \quad (12)$$

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, velo-

cidade potencial, temperatura, umidade, pressão em superfície, encontram-se nas Tabelas 2 e 3.

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, i.e., 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 2) 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.

Os valores máximos e mínimos dos comprimentos de escala horizontais (Tabela 2) 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) 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.

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 Fig. 2 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”).

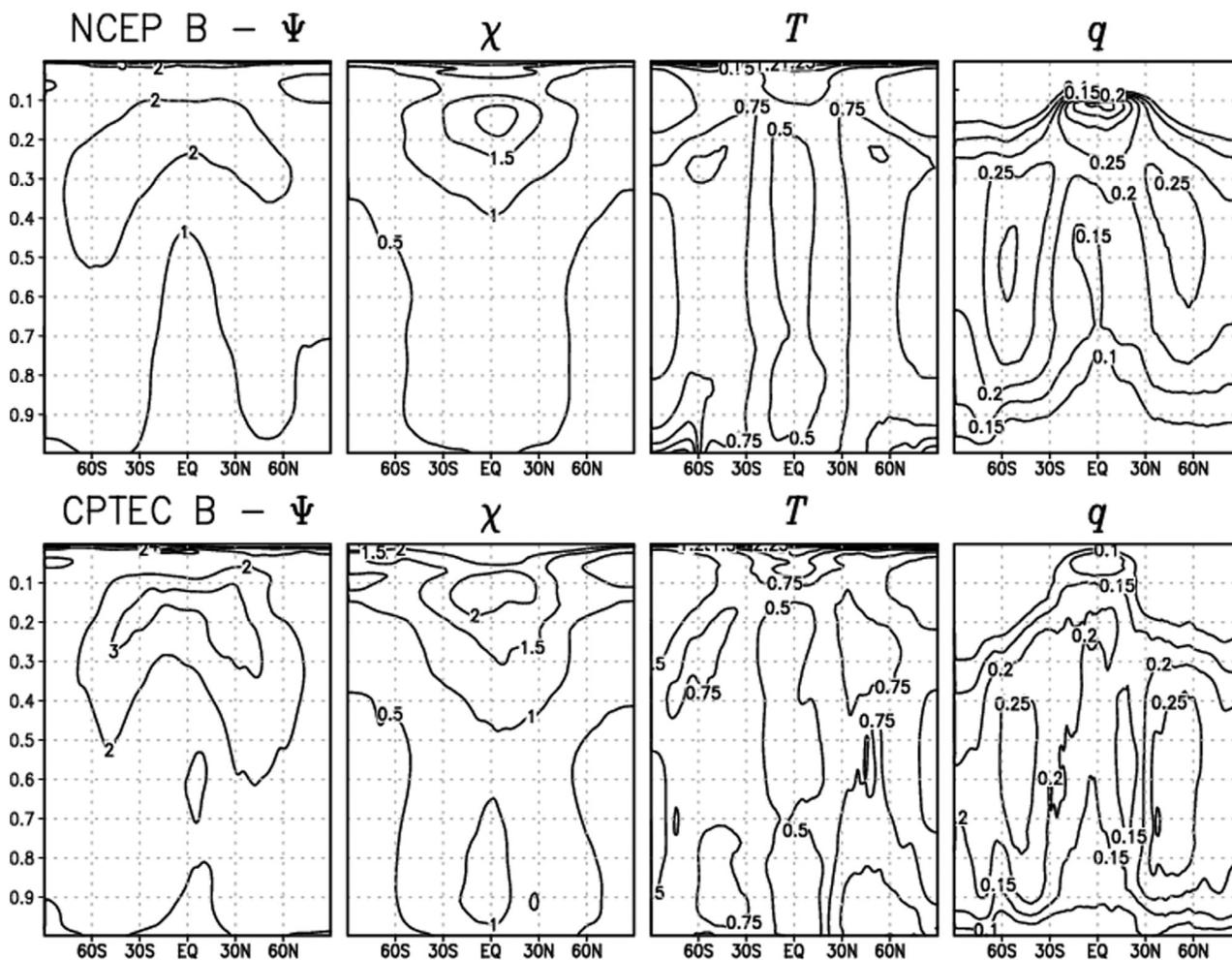
Comparando-se “B NCEP” e “B CPTEC” (as duas matrizes contendo os pares de previsões de todos os horários sinóticos padrão), observa-se que na matriz “B CPTEC” as amplitudes das variâncias dos erros da parte não divergente do fluxo atmosférico (representadas através da função de corrente, primeira coluna da Fig. 2) são mais

**Tabela 3** - 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
$\psi$	$8,3 \times 10^{-2}$	$9,1 \times 10^{-1}$	$8,41 \times 10^{-2}$	1,27
$\chi$	0	1,22	0	1,5
$t$	0	1,75	0	1,67
$q$	0	$1,3 \times 10^6$	0	$1,75 \times 10^6$

**Tabela 2** - 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
$\psi$	$4,3 \times 10^5$	$7,99 \times 10^6$	$4,6 \times 10^5$	$2,8 \times 10^7$	$2,2 \times 10^5$	$1,3 \times 10^6$	$4,55 \times 10^5$	$1,75 \times 10^6$
$\chi$	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	$2,22 \times 10^7$
$ps$	$4,88 \times 10^{-2}$	$1,01 \times 10^{-1}$	$6,35 \times 10^{-2}$	$1,48 \times 10^{-1}$	$1,19 \times 10^5$	$1,89 \times 10^5$	$1,83 \times 10^5$	$4,18 \times 10^5$



**Figura 2** - Distribuição das amplitudes da Função de Corrente ( $\psi$ ,  $\times 10^6$ ), Velocidade Potencial ( $\chi$ ,  $\times 10^6$ ), 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).

concentradas e menos suaves do que as amplitudes observadas na matriz “B NCEP” e espacialmente 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 h e 24 h provenientes do ciclo de assimilação de dados do G3DVAR. Consequentemente, a variância do erro sobre o Hemisfério Sul, sobretudo sobre a América do Sul (entre 0 e 60 S) 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 mo-

delos que foram utilizadas para o cálculo das matrizes de covariâncias. A matriz “B NCEP”, embora com uma grade horizontal de dimensões menores do que 900 x 450 (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 Fig. 2), assim como para a função de corrente, mostram que sobre a região tropical (entre 30 S e 30 N), 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 Fig. 2), a matriz “B CPTEC”, a maior parte da variância dos erros encontra-se concentrada sobre o Hemisfério Norte a partir de 30 N. Na matriz “B NCEP”, uma característica interessante pode ser assinalada sobre a latitude de 60 N, 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 60 S e 30 S e entre os níveis sigma 1 e 0,7 (aproximadamente entre 1024 e 716,8 hPa). Especula-se, entretanto, que tais diferenças possam estar relacionadas 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 Fig. 2) 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 Fig. 3 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 60 S e 60 N. Na matriz do CPTEC, a distribuição das amplitudes é assimétrica, sendo que os picos das variâncias dos erros são maiores sobre a latitude de 60 N do que sobre a latitude de 60 S. 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 30 N e 60 N.

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 utili-

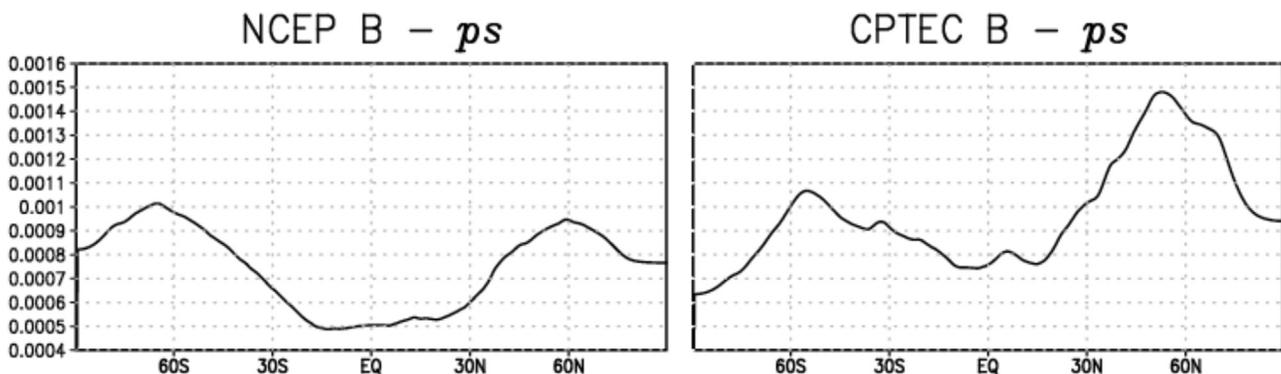


Figura 3 - Distribuição das amplitudes da Pressão em Superfície ( $ps$ ) ao longo das latitudes (em  $x$ ) e níveis sigma (em  $y$ ), calculadas utilizando-se o método NMC. A comparação é feita com o NCEP.

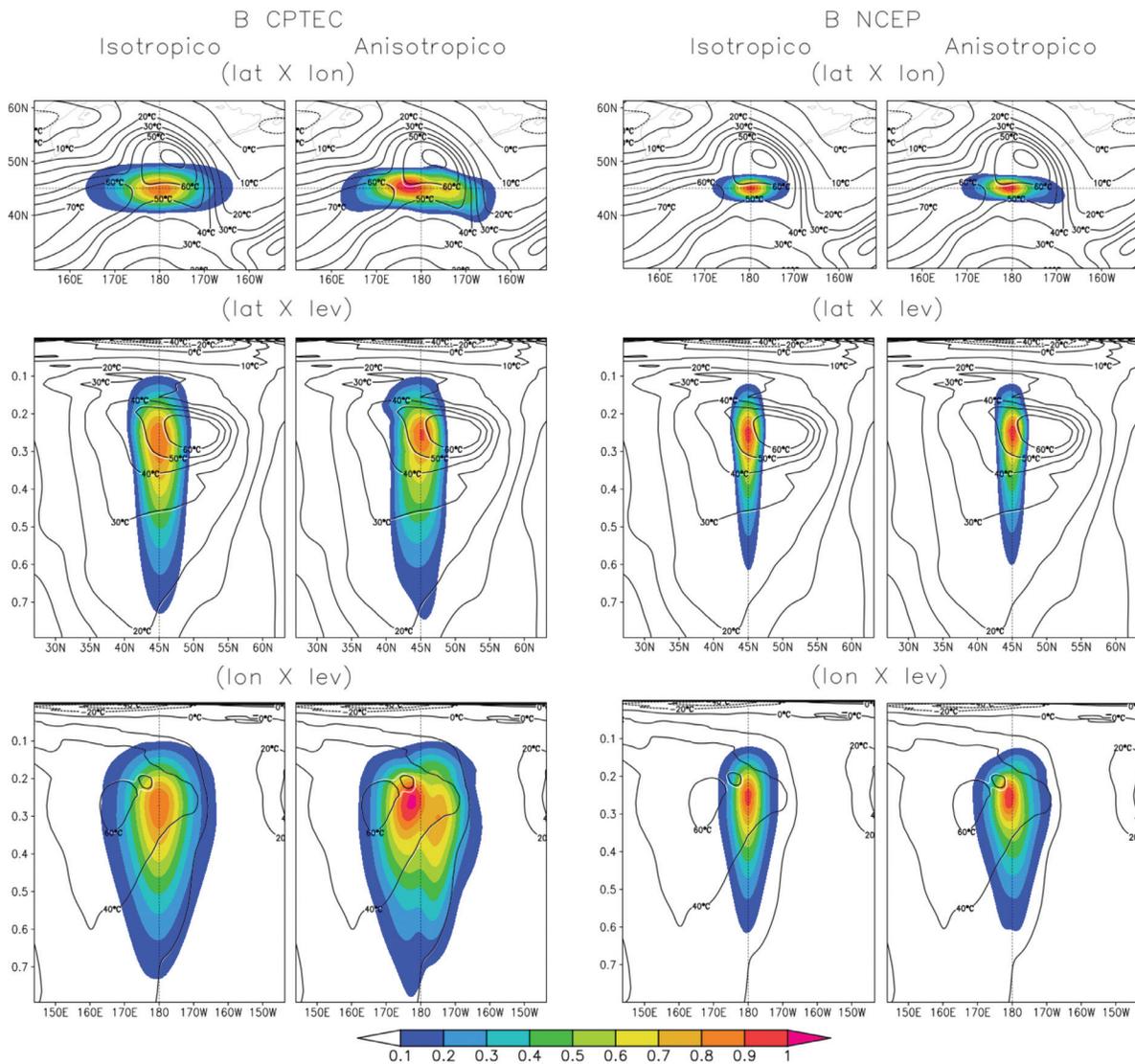
zados pelo modelo de transferência radiativa do GSI para o cálculo dos perfis de temperatura de brilho e dos perfis de radiancias simulados (perfis referentes ao modelo).

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 válidas 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 (i.e.,  $y - Hx$ ). A fim de que seja possível caracterizar o aspecto do incremento de análise resultante, as Figs. 3 e 4 apresentam



**Figura 4** - 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.

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 4 apresenta, portanto, os valores dos parâmetros principais utilizados para produzir os incrementos apresentados nas figuras desta seção.

A Fig. 4, 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 45 N e 180 (*lat X lon*, respectivamente). Além disso, ajustou-se o erro da observação para  $1 \text{ m s}^{-1}$ , bem como a magnitude da inovação trazida pela observação (previsão subtraída da observação, ou OMF). Na Fig. 4, 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 4). Esta opção é uma importante funcionalidade do sistema GSI, pois ela permite que a matriz de covariâncias seja ajustada com base na tendência temporal im-

pressa nos campos de *background* utilizados para o cálculo da análise.

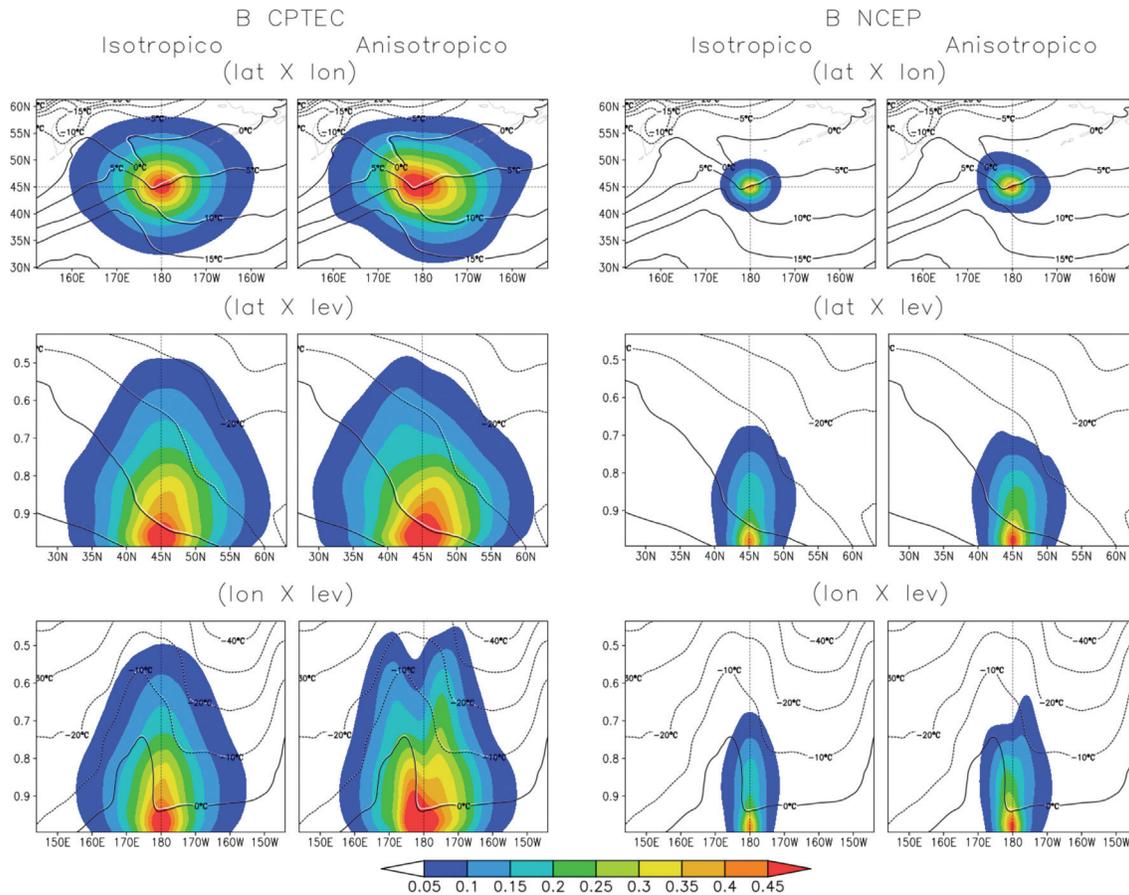
As linhas tracejadas na horizontal e na vertical nas figuras de ambos os painéis estão destacadas para auxiliar na identificação do ponto em que a observação sintética foi posicionada, além de indicar as seções latitudinais, longitudinais e verticais feitas para mostrar os diferentes aspectos do incremento de análise produzido. As isolinhas pretas indicam o *background* da componente zonal do vento (indicado em  $\text{ms}^{-1}$ ) 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 matrizes se ajustaram à isolinha de  $20 \text{ ms}^{-1}$ , a partir do nível sigma 0,6 (aproximadamente 600 hPa).

A Fig. 5 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  $1 \text{ }^\circ\text{C}$ , respectivamente. A observação foi também posicionada no ponto de coordenadas 45 N 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, i.e., 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.

As diferenças entre os incrementos de análises produzidos pelas matrizes **B** do CPTEC e do NCEP utilizando os

**Tabela 4** - Parâmetros de configuração utilizados no sistema GSI para a aplicação das matrizes de covariâncias do CPTEC e do NCEP.

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



**Figura 5** - 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.

parâmetros indicados na Tabela 4, 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 ( $\zeta$ ). 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 ( $\psi$ ) e velocidade potencial ( $\chi$ ), para o cálculo do incremento de análise do campo de vento, são então calculadas - através da Eq. (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 2). 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. (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 assi-

milação 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 (i.e., 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 intrasazonal, podem também ser utilizados com o objetivo de se melhorar a previsibilidade de fenômenos e oscilações (e.g., Madden-Julian) dentro desta escala de tempo.

## 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.

## Referências

- 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**, v. 134, n. 637, p. 1951-1970, 2008.
- 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.
- 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.
- DALEY, R. **Atmospheric Data Analysis**. Cambridge University Press, 1993.
- 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.

- DERBER, J.; ROSATI, A. A Global Oceanic Data Assimilation System. **Journal of Physical Oceanography**, v. 19, n. 9, p. 1333-1347, 1989.
- FIGUEROA, S.N.; BONATTI, J.P.; KUBOTA, P.Y.; GRELL, G.A.; MORRISON, H. *et al.* 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.
- 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 Meteorology 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. *et al.* Data Assimilation in the Ocean and in the Atmosphere: What Should be Next? **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**, v. 24, n. 6, p. 1691-1705, 2009.
- 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 de Doutorado, Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012-08-06 2012.
- 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**, v. 112, n. 474, p. 1177-1194, 1986.
- PANNEKOUCKE, O.; BERRE, L.; DESROZIERS, G. Background-Error Correlation Length-Scale Estimates and their Sampling Statistics. **Quarterly Journal of the Royal Meteorological Society**, v. 134, n. 631, p. 497-508, 2008.
- 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, 1992.
- PETRIE, R.E. **Background Error Covariance Modelling for Convective-Scale Variational Data Assimilation**. Ph.D. Thesis, 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**, v. 131, n. 8, p. 1536-1548, 2003.
- 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, 2003.
- 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.
- STENGEL, R. **Optimal Control and Estimation**. Dover Publications, 1994.
- 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.