

Assimilação de Dados por Redes Neurais Artificiais

Haroldo F. de Campos Velho – INPE

Helaine C. M. Furtado - UFOPA

Juliana A. Anochi – INPE

Roberto P. Souto – LNCC

Vinicius A. Albuquerque - UFRJ

Mini-curso: Assimilação de Dados por Redes Neurais

- O que é "assimilação de dados"?
 - O porque da necessidade e breve histórico
- Métodos de assimilação de dados
 - *Nudging* e Métodos Variacionais
 - Filtro de Kalman e filtro de Kalman por conjunto
- Redes neurais: breve descrição
 - Redes neurais: MLP, FBR, Deep Learning
 - Redes neurais para assimilação de dados
 - Modelo de circulação oceânica 2D
- Aplicações
 - Modelos de baixa ordem: Lorenz-63, *shallow water* 1D e 2D
 - Processamento paralelo para assimilação com redes neurais
 - Modelos atmosféricos 3D: WRF (regional), SPEED e FSU (globais)

Mini-curso: Assimilação de Dados por Redes Neurais

- O que é "assimilação de dados"?
 - O porque da necessidade e breve histórico
- Métodos de assimilação de dados
 - *Nudging* e Métodos Variacionais
 - Filtro de Kalman e filtro de Kalman por conjunto
- Redes neurais: breve descrição
 - Redes neurais: MLP, FBR, Deep Learning
 - Redes neurais para assimilação de dados
 - Modelo de circulação oceânica 2D
- Aplicações
 - Modelos de baixa ordem: Lorenz-63, *shallow water* 1D e 2D
 - Processamento paralelo para assimilação com redes neurais
 - Modelos atmosféricos 3D: WRF (regional), SPEED e FSU (globais)

Assimilação de dados por redes neurais

■ Instrutores:

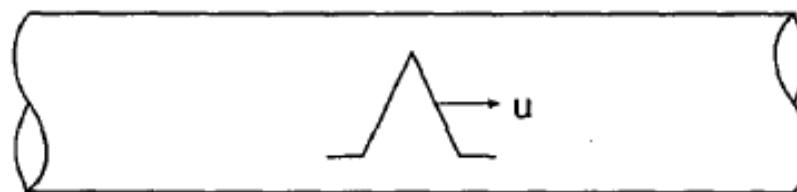
- Haroldo F. de Campos Velho – INPE
- Helaine C. M. Furtado - UFOPA
- Juliana A. Anochi – INPE
- Roberto P. Souto – LNCC
- Gerônimo Lemos – INPE
- Marcelo Paiva – INPE

O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações:

$$\left\{ \begin{array}{l} \frac{\partial \theta}{\partial t} + u \frac{\partial \theta}{\partial x} = 0 , \quad \theta(x,0) = \phi(x) \end{array} \right.$$

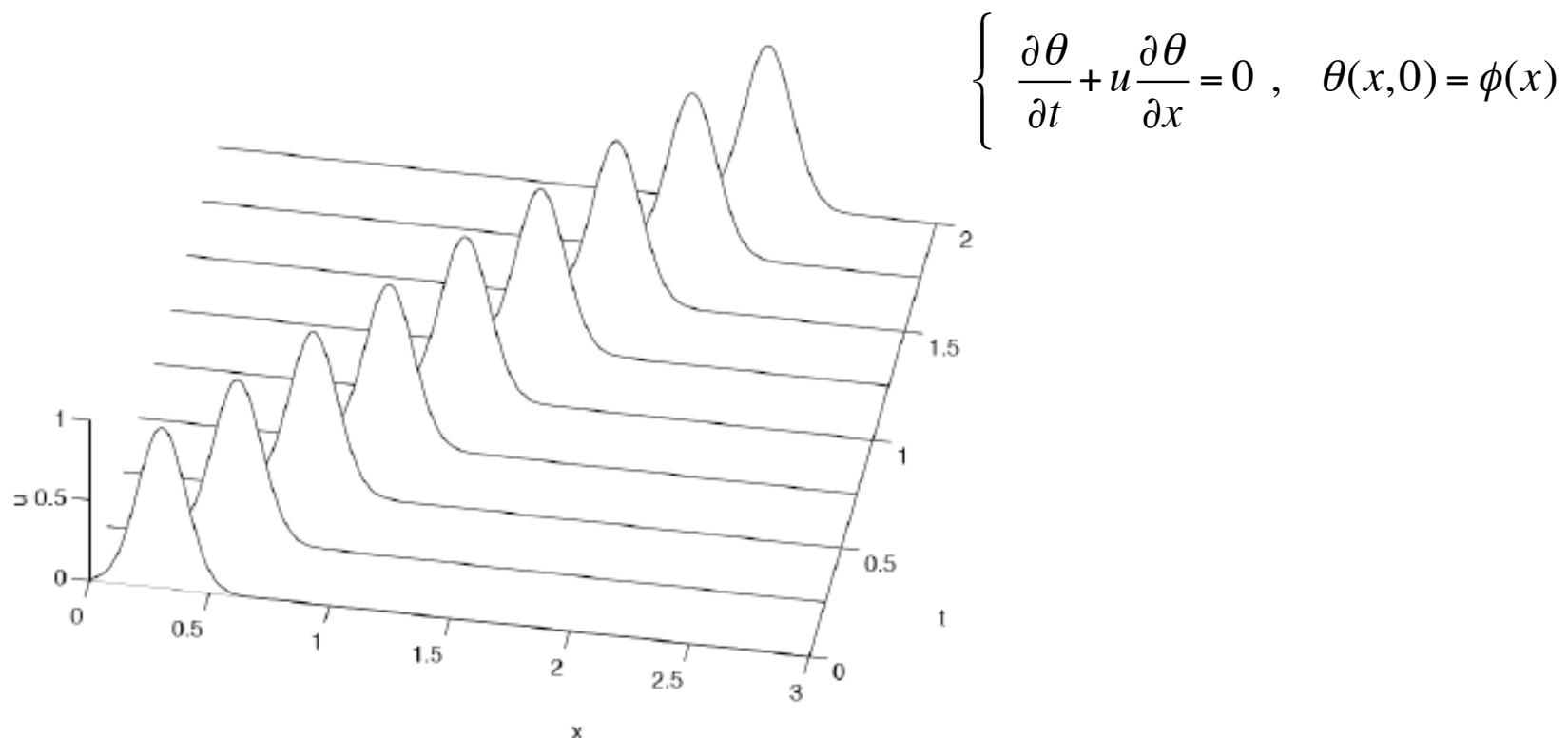
Incompressible Liquid



$$f_t + u f_x = 0, \quad f(x,0) = F(x), \quad f(x,t) = ?$$

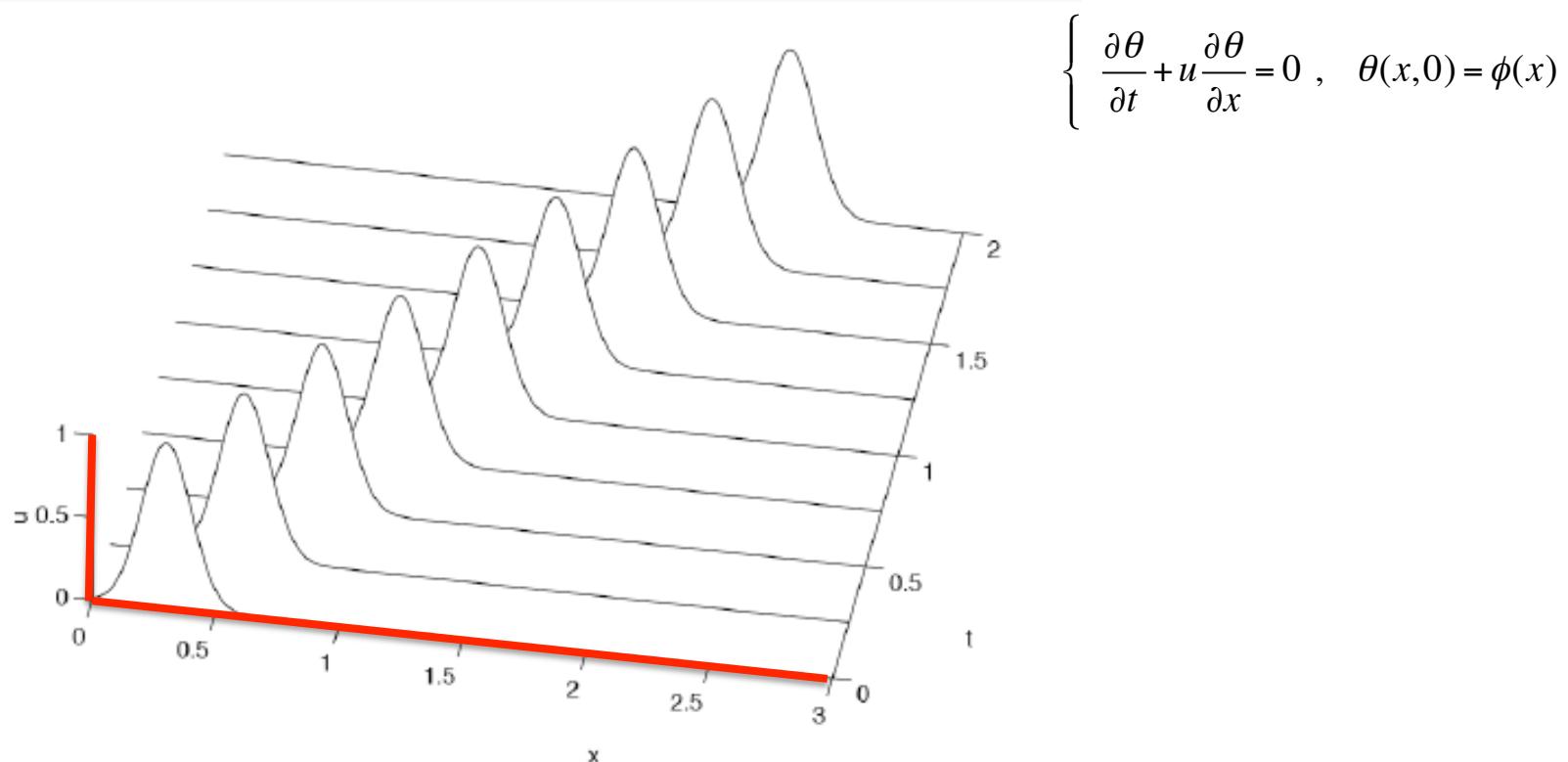
O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações:



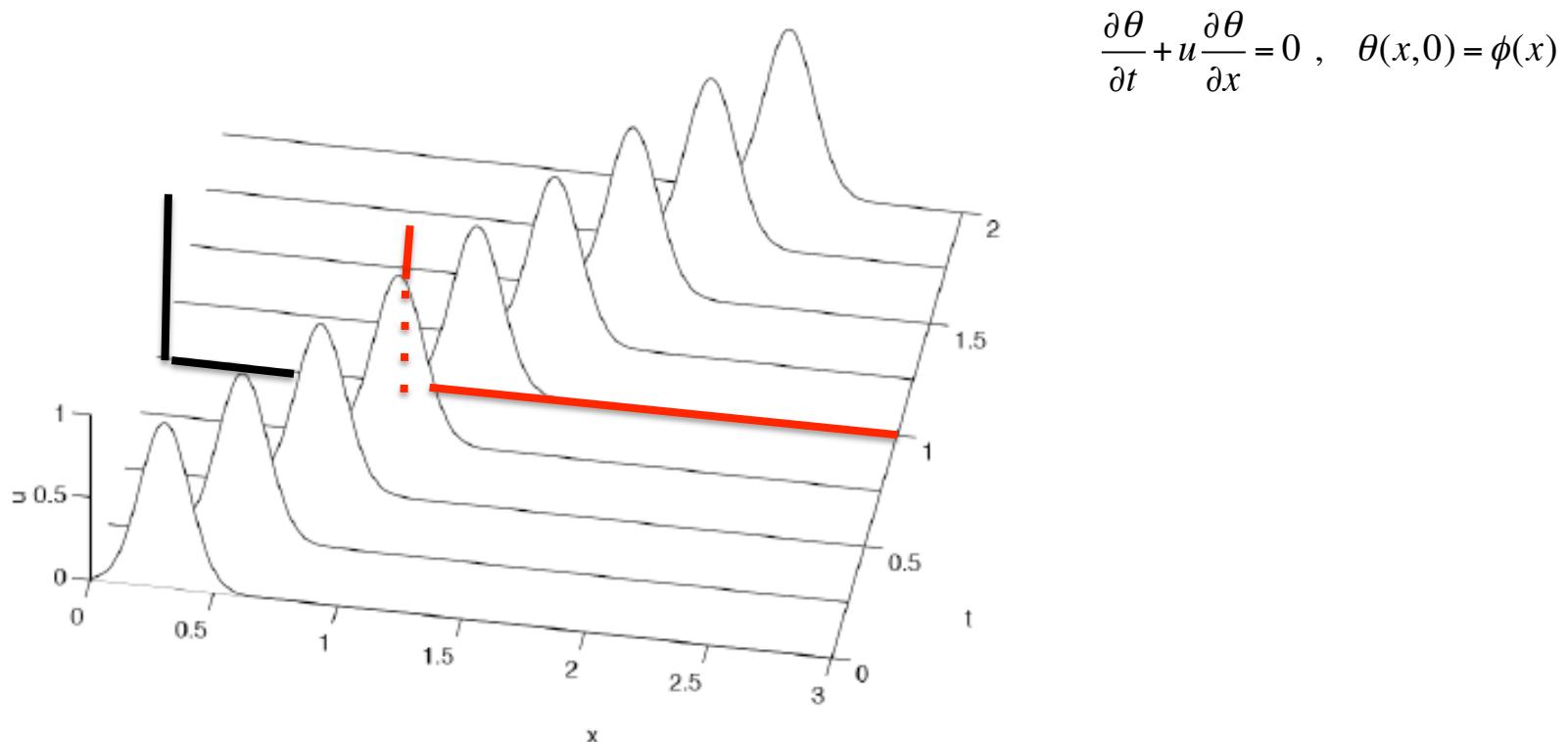
O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações:



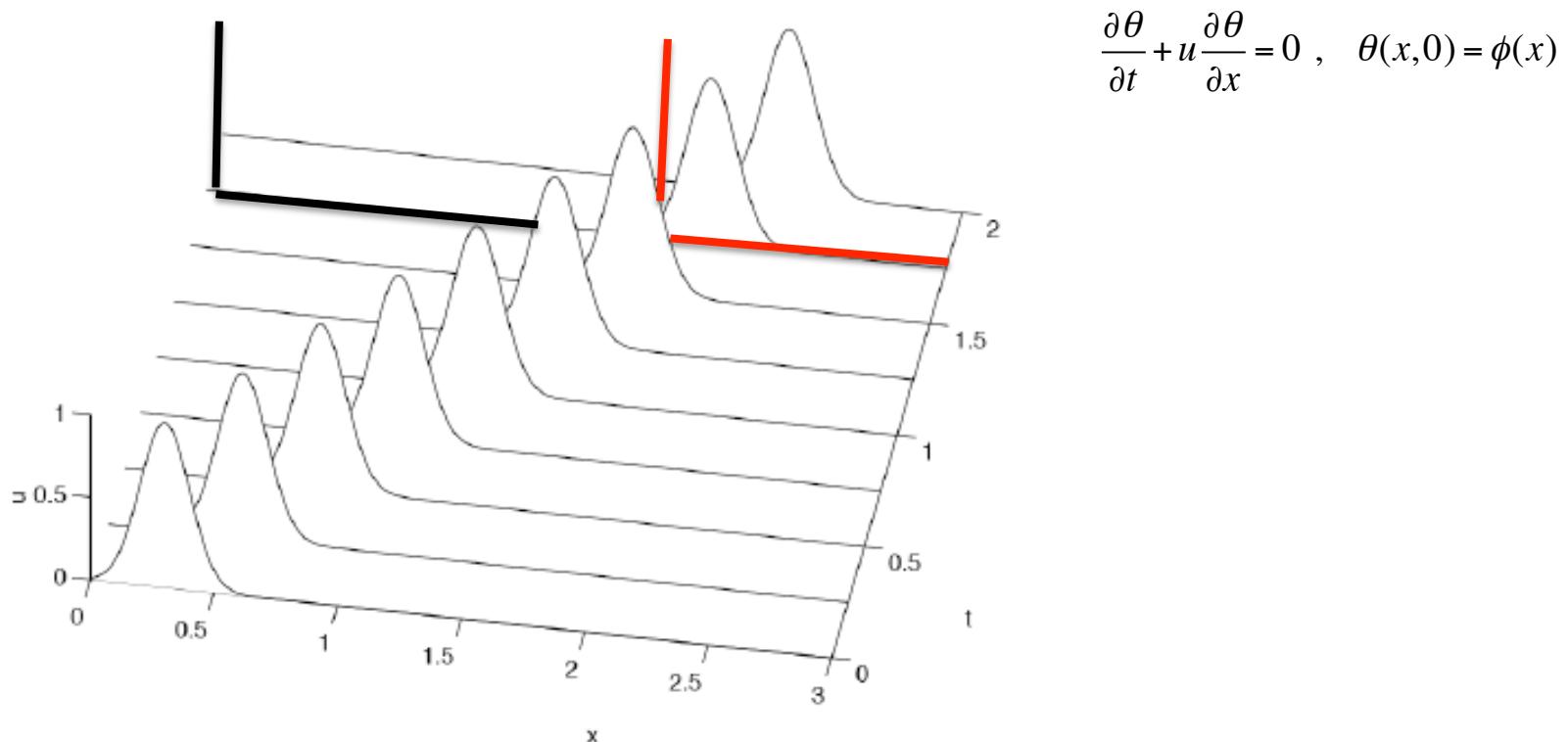
O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações:



O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações:



Equação da onda de 1a. Ordem (Eq. de advecção)

1. UPWIND: solução numérica da equação de transporte

$$\left(\frac{\partial f}{\partial t} \right)_i^n = -u \left(\frac{\partial f}{\partial x} \right)_i^n$$

$$\left(\frac{\partial f}{\partial t} \right)_i^n \approx \frac{f_i^{n+1} - f_i^n}{\Delta t}$$

$$\left(\frac{\partial f}{\partial x} \right)_i^n \approx \begin{cases} (f_{i+1}^n - f_i^n) / \Delta x & (\text{se: } c < 0) \\ (f_i^n - f_{i-1}^n) / \Delta x & (\text{se: } c > 0) \end{cases}$$

Equação da onda de 1a. Ordem (Eq. de advecção)

1. UPWING: solução numérica da equação da advecção

$$\frac{f_i^{n+1} - f_i^n}{\Delta t} + u \frac{f_i^n - f_{i-1}^n}{\Delta x} = 0 \quad (u > 0)$$



$$f_i^{n+1} = f_i^n - c(f_i^n - f_{i-1}^n) \quad \left[c = \frac{u\Delta t}{\Delta x} \right]$$

Equação da onda de 1a. Ordem (Eq. de advecção)

1. Método de Lax-Wendroff de passo único

$$f_i^{n+1} = f_i^n + (f_t)_i^{n+1} \Delta t + \frac{1}{2} (f_{tt})_i^{n+1} \Delta t^2 + O(\Delta t^3)$$

$$f_t = -u f_x \quad \Downarrow$$

$$f_{tt} = (f_t)_t = (-u f_x)_t = -u (f_t)_x = -u (-u f_x) = u^2 f_{xx}$$

$$\Downarrow$$

$$f_i^{n+1} = f_i^n - (u f_x)_i^{n+1} \Delta t + \frac{u^2}{2} (f_{xx})_i^{n+1} \Delta t^2 + O(\Delta t^3)$$

Equação da onda de 1a. Ordem (Eq. de advecção)

1. Método de Lax-Wendroff de passo único

$$f_i^{n+1} = f_i^n - u \left(\frac{f_{i+1}^n - f_{i-1}^n}{2\Delta x} \right) \Delta t + \frac{u^2}{2} \left(\frac{f_{i+1}^n - 2f_i^n + f_{i-1}^n}{\Delta x^2} \right) \Delta t^2$$



$$f_i^{n+1} = f_i^n - \frac{c}{2} (f_{i+1}^n - f_{i-1}^n) + \frac{c^2}{2} (f_{i+1}^n - 2f_i^n + f_{i-1}^n)$$

Equação da onda de 1a. Ordem (Eq. de advecção)

1. Método de Lax-Wendroff de passo único

$$\begin{aligned} f_t + u f_x = & -\frac{1}{2} f_{tt} \Delta t - \frac{1}{6} f_{ttt} \Delta t^2 - \frac{1}{24} f_{tttt} \Delta t^3 \dots - \frac{1}{6} u f_{xxx} \Delta x^2 - \dots \\ & + \frac{1}{2} u^2 f_{xx} \Delta t - \frac{1}{24} u^2 f_{xxxx} \Delta x^2 \Delta t + \dots \end{aligned}$$

2. Quando: $\Delta t \rightarrow 0$ e $\Delta x \rightarrow 0$ (método é consistente)
3. Substituindo (EDM-2): $f_{tt} = u^2 f_{xx}$
4. Nota-se que: $O(\Delta t^2), O(\Delta x^2)$.

Equação da onda de 1a. Ordem (Eq. de advecção)

1. Lax-Wendroff de passo único: estabilidade

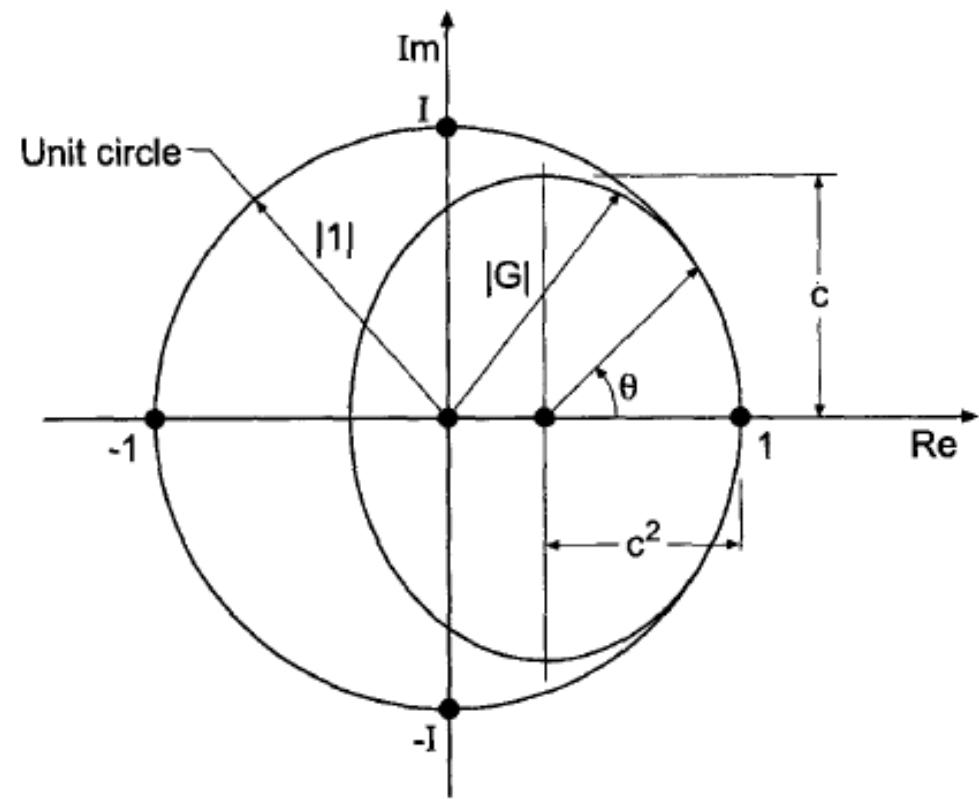
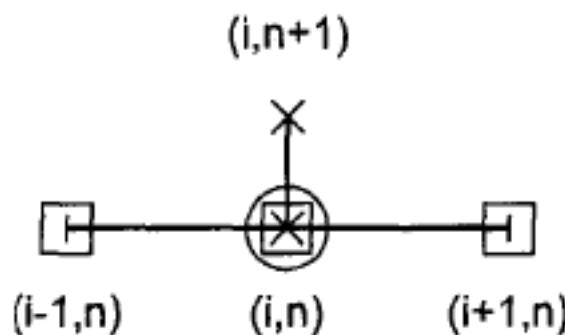
$$G = \left[(1 - c^2) + c^2 \cos \theta \right] - Ic \sin \theta$$

2. A equação acima é uma elipse no plano complexo. A elipse estará contida no círculo unitário sse:

$$c = \frac{u\Delta t}{\Delta x} \leq 1$$

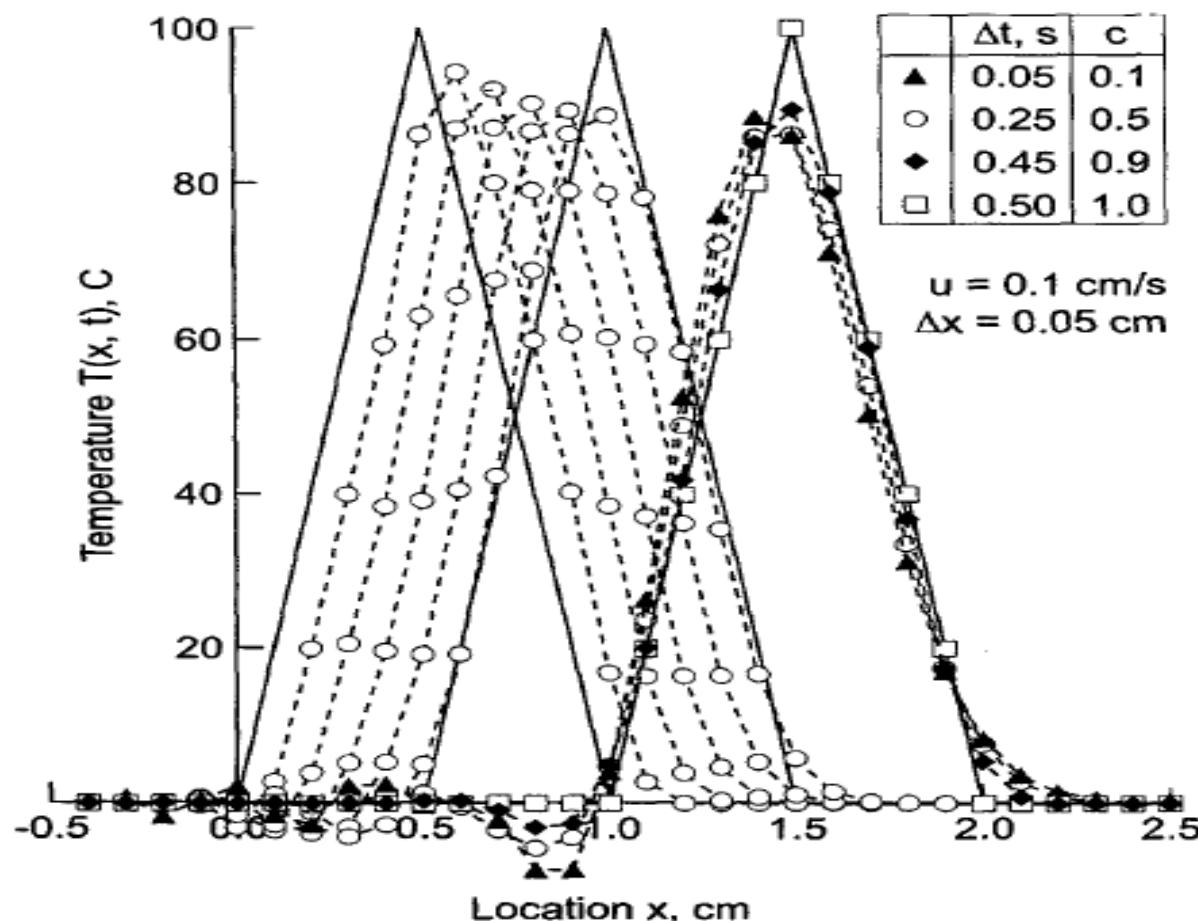
Equação da onda de 1a. Ordem (Eq. de advecção)

1. Lax-Wendroff de passo único: estabilidade



Equação da onda de 1a. Ordem (Eq. de advecção)

1. Lax-Wendroff de passo único



Equação da onda de 1a. Ordem (Eq. de advecção)

1. Difusão e dispersão numéricas

- Método de Lax-Wendroff (passo único): EDF

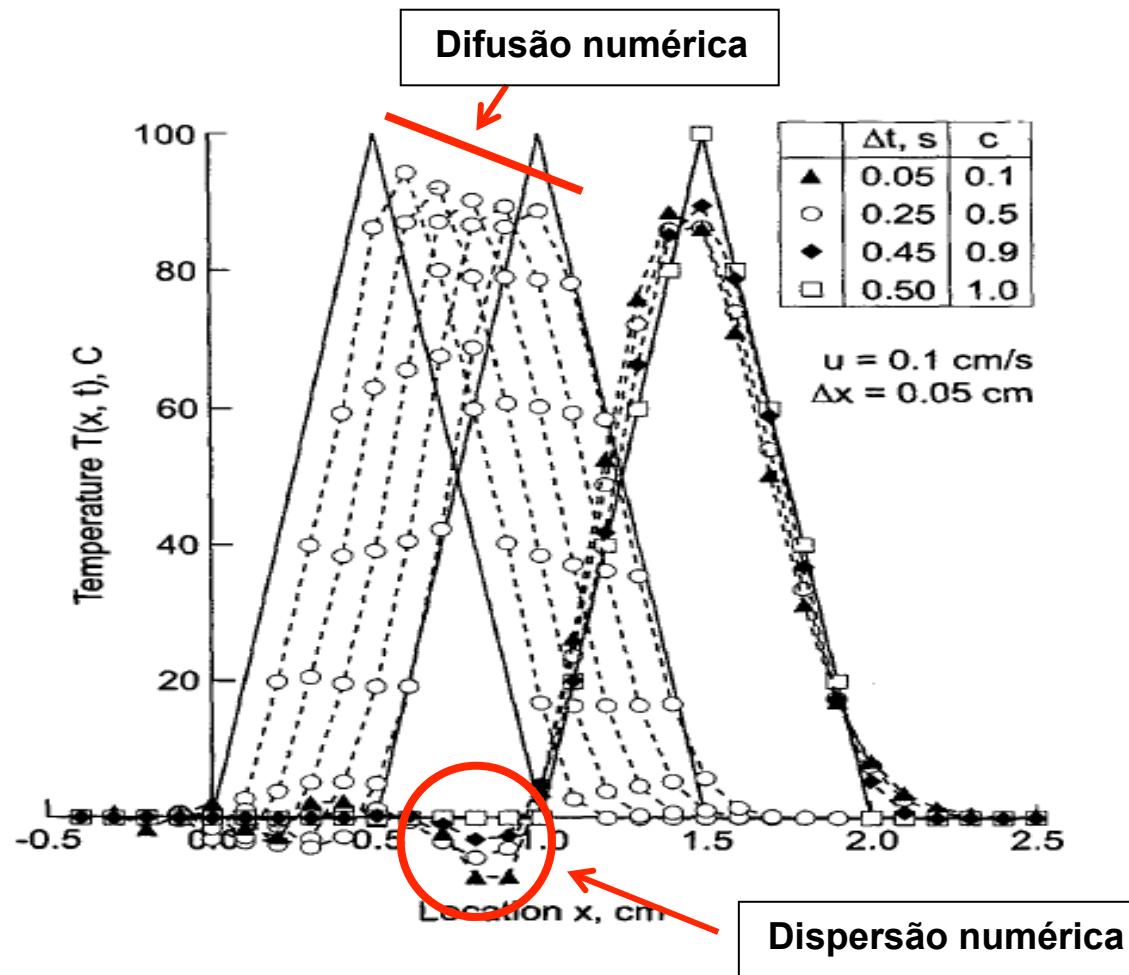
$$f_i^{n+1} = f_i^n - \frac{c}{2} (f_{i+1}^n - f_{i-1}^n) + \frac{c^2}{2} (f_{i+1}^n - 2f_i^n + f_{i-1}^n)$$

- Análise de consistência (o que de fato a EDF resolve)

$$f_t + u f_x = \underbrace{\frac{u}{6} (u^2 \Delta t^2 - \Delta x^2)}_{\text{↑ Coeficientes numéricos!}} f_{xxx} + \underbrace{\frac{u \Delta t}{8} (u^3 \Delta t^3 - \Delta x^2)}_{\begin{array}{l} \text{Derivada de ordem par (4a. ordem): difusão} \\ \text{Derivada de ordem ímpar (3a. ordem): dispersão} \end{array}} f_{xxxx} + \dots$$

Equação da onda de 1a. Ordem (Eq. de advecção)

1. Lax-Wendroff de passo único



O que é “Assimilação de Dados”?

- Antes, vamos mencionar o contexto.
- O foco é em sistemas de previsão operacional ...
- ... baseados em integração temporal de equações diferenciais ou integro-diferenciais.
- Problemas:
 - O modelo de equações diferenciais é uma aproximação!
 - Por vezes, não conhecemos o valor dos parâmetros da equação.
 - Não conhecemos a(s) condição(ões) inicial(is) das equações.

O que é “Assimilação de Dados”?

- Tentando mitigar os problemas ...
- O modelo matemático é só uma aproximação.
 - Não há muito o que fazer.
 - É necessário um estudo contínuo para melhorar parte das representações contidas no modelo que simula a realidade.
 - Novas teorias (parametrizações (*)) são incorporadas ao modelo na medida que são desenvolvidas, testadas e então um novo modelo é aplicado para previsão.
 - (*) Parametrizações: Ao invés de descrever a representação de algum processo em toda sua complexidade, usamos simplificações, que tentam (mesmo assim) gerar resultados satisfatórios.

O que é “Assimilação de Dados”?

- Tentando mitigar os problemas ...
- Não conhecemos os valores dos parâmetros ...
 - Este é problema de ajustar os parâmetros do modelo aos dados medidos – campanhas experimentais.
 - Isto é por vezes chamado de **calibração** de modelos.
 - De forma mais geral, o problema de calibração é também chamado de **problema inverso**.
 - **Problema Direto:** modelo matemático que temos conhecimento de características do sistema, das condições iniciais e de contorno, das fontes/sumidouros do sistema.
 - **Problem Inverso:** conhecemos o problema direto, mas queremos estimar alguma(s) das propriedades listadas acima.

O que é “Assimilação de Dados”?

- Tentando mitigar os problemas ...
- Não conhecemos a condição inicial.
 - Sabemos que o modelo é só uma aproximação da realidade.
 - Desta forma, mesmo que a previsão apresente uma boa aderencia com a realidade, a medida que o período de tempo da previsão vai ficando cada vez maior, o resultado da simulação (previsão) terá um erro cada vez maior. Até que a previsão está completamente dissociada da realidade.
 - “Solução aparente”: faz-se medidas da realidade e então se substitui valores simulados por valores observacionais.
 - A prática mostrou que a “solução aparente” não funciona: ao invés de melhorar a previsão, o procedimento degrada a previsão ...

O que é “Assimilação de Dados”?

- Tentando mitigar os problemas ...
- Não conhecemos a condição inicial.
 - A prática mostrou que a “solução aparente” não funciona: ao invés de melhorar a previsão, o procedimento degrada a previsão ...
 - **Como resolver?**
 - Solução adotada para a prática da previsão:
 - Combina-se de alguma forma valores da previsão anterior com valores de observação.
 - Com a estimativa da condição inicial (previsão prévia + dados de observação), é iniciado novo ciclo de previsão.

O que é “Assimilação de Dados”?

- Respondendo a pergunta:
- É um processo para determinar condição inicial.
- O processo é uma **metodologia** adequada para combinar dados de uma previsão anterior, como dados de observação.
- Assim, o processo de previsão envolve 2 sistemas:
 - Modelo matemático com evolução do tempo calculado por métodos computacionais.
 - Um sistema de observação (diferentes sensores) que coletam as medidas que são enviadas regular e sistematicamente para os Centros de Previsão.

O que é “Assimilação de Dados”?

- Respondendo a pergunta:
- É um processo para **determinar condição inicial**.
- Quanto melhor a condição inicial, mais precisa será a previsão.
- Ou seja, assimilação de dados é um **problema inverso de determinação da condição inicial**.
- É um problema inverso de tal relevância que tem uma atenção específica.

Porque assimilação de dados é importante?

- Antes, vamos ver Centros de Previsão e o que fazem.
- Centros operacionais: Instituições que fazem previsões periódicas de interesse da sociedade.
- Centros de previsão:
 - Tempo (*weather*) e clima
 - Circulação oceânica
 - Previsão ambiental
 - Conteúdo eletrônico da atmosfera
 - Epidemias, ...
- Os períodos de previsão podem ser:
 - Algumas horas à frente (*nowcasting*).
 - Dias à frente (curto ou médio prazo)
 - Mais de 30 dias (período que já se considera na escala de *clima*)

Porque assimilação de dados é importante?

- Antes, vamos ver Centros de Previsão e o que fazem.
- Previsão de tempo e clima é importante porque:
 - Defesa civil (catástrofes provocadas por condições meteorológicas extremas: inundações, secas profundas, furacões/tufões), **agricultura**, indústria do turismo, produção/administração de energia (usinas hidro-elétricas – Brasil), precificação de seguros, aviação, etc.
- Previsão de circulação oceânica é importante porque:
 - Navegação de navios, poluição, recursos naturais, etc.
- Previsão ambiental é importante porque:
 - Seres vivos precisam de ambientes limpos
- Previsão de conteúdo eletrônico é importante porque:
 - Quando sinais de GNSS (GPS, Glonass, Galileo, ...) são confiáveis?
- Previsão de epidemia/pandemia: Não precisa explicar.

O que é “Assimilação de Dados”?

- É um processo para determinar condição inicial.
- Assimilação de dados é um **problema inverso** de **determinação da condição inicial**.
- Vale a pena olhar métodos de solução de problemas inversos ...

O que é “Assimilação de Dados”?

- É um processo para determinar condição inicial.
- Um conjunto de técnicas importantes de um ramo especializado da **análise de dados**.
- É uma das disciplinas da **teoria de estimação**.
- É um tipo especial e importante de **problema inverso**.
- É um conjunto de técnicas empregadas para realizar adequadamente a **inserção de dados de observação num sistema operacional de previsão**.

Scientific challenges

1. Before the XX century:

We want to know the nature laws (mechanics, thermodynamics, electromagnetism, life evolution, social behaviour, transfinite numbers)

2. During the XX century:

We know the laws (equations), but we want to solve them.

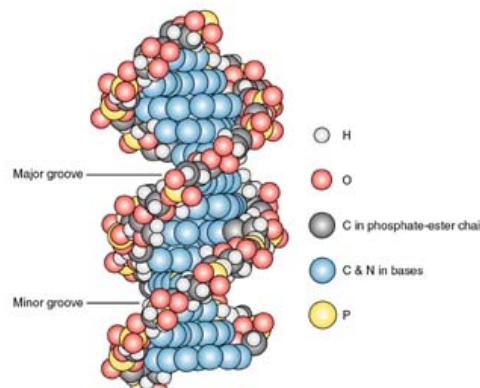
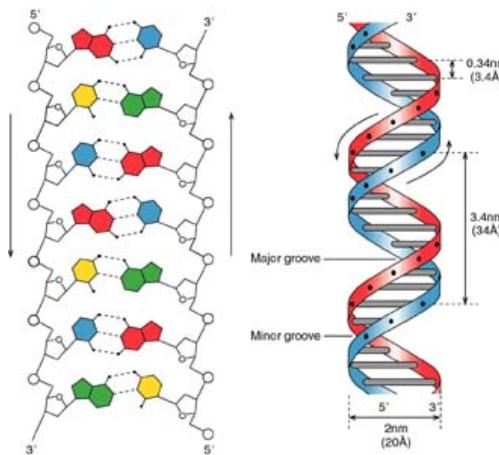
Remarkable conquest: modern numerical weather prediction!

3. After the XX century (our century!):

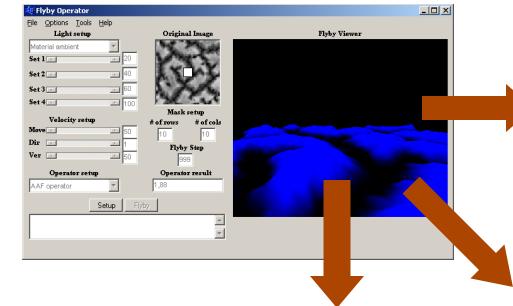
Starting this new century, data analysis is occupying a central role in the science (genomic, data mining, background cosmic radiation in microwave, data assimilation).

Scientific challenges: data analysis

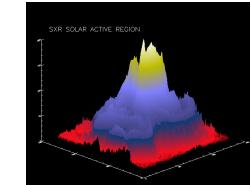
Genome projects



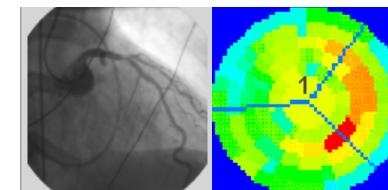
Complex system characterization



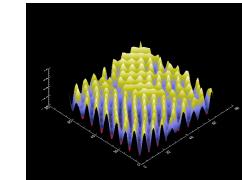
Solar physics



Heart disease



Plane waves

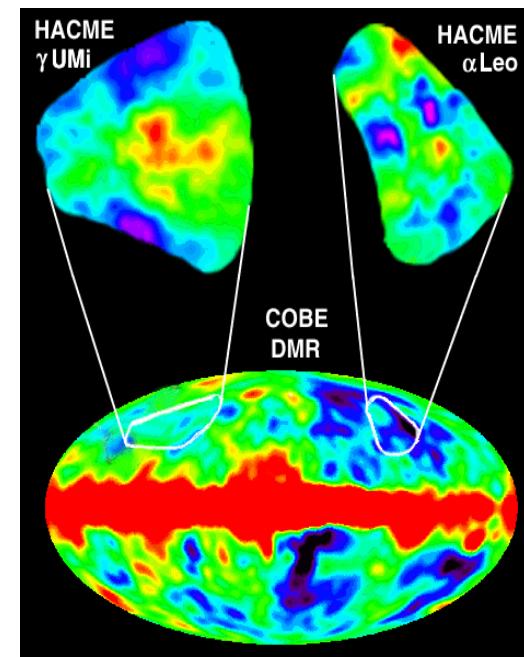
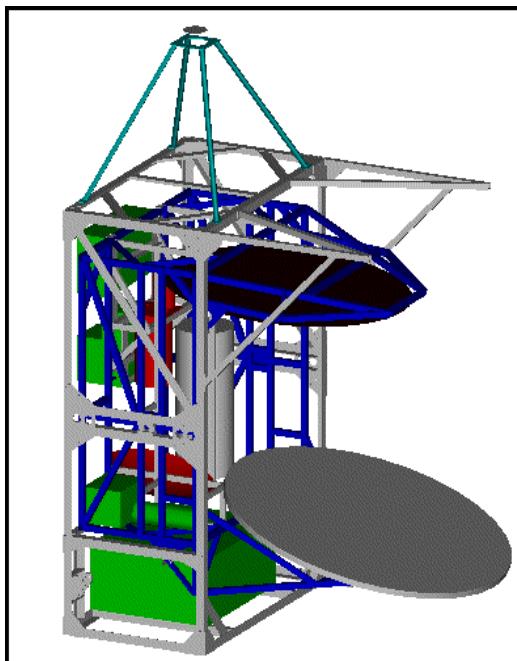


Scientific challenges: data analysis

Background cosmic radiation in microwave

$$J(\mathbf{a}) = \sum_{i=1}^N \frac{1}{\sigma_i^2} \left[\frac{\Delta T_i}{T} - \sum_{k=1}^M a_k Y_{k,i} \right]$$

vector $\mathbf{a} = [a_1 \ a_2 \dots a_M]^T$ denotes the expansion coefficients on spherical harmonics Y_k and ΔT_i is the sky temperature.



Scientific challenges: data analysis



Applications:

Air monitoring (P. Zannetti: Air Pollution Modeling, 1990)

Meteorology (R. Daley: Atmospheric Data Analysis, 1991)

(E. Kalnay: Atmospheric Modeling, Data assimilation and Predictability, 2002)

Oceanography (A.F. Bennet: Inverse Methods in Physical Oceanography, 1992)

Data assimilation: graphical scenario



Forecasts Scores

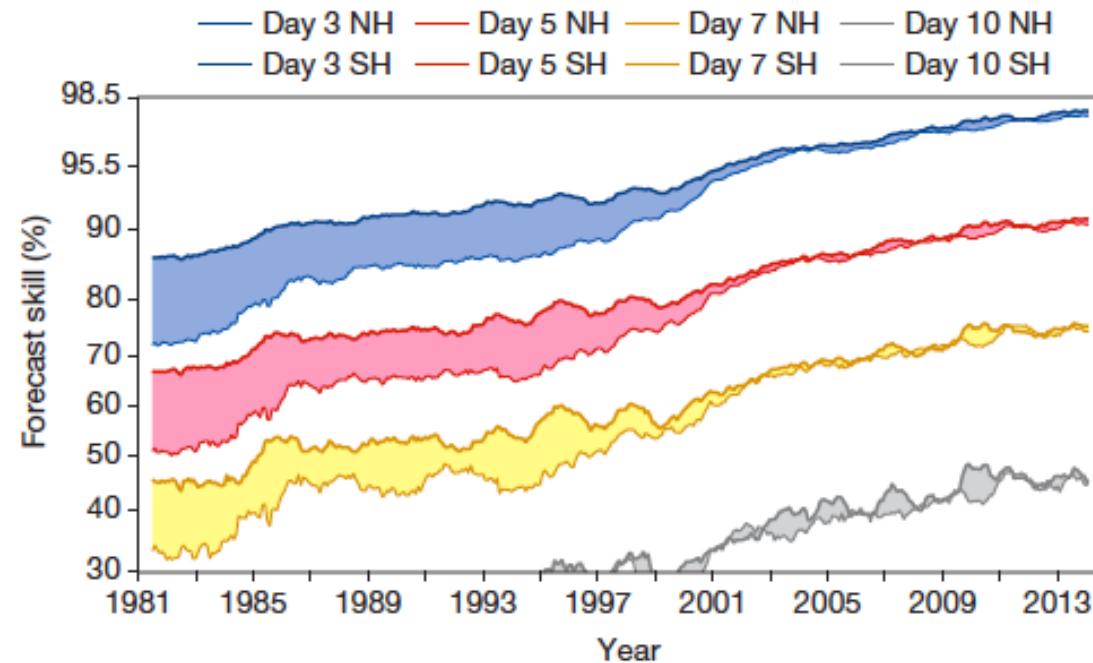
ECMWF

REVIEW

doi:10.1038/nature14956

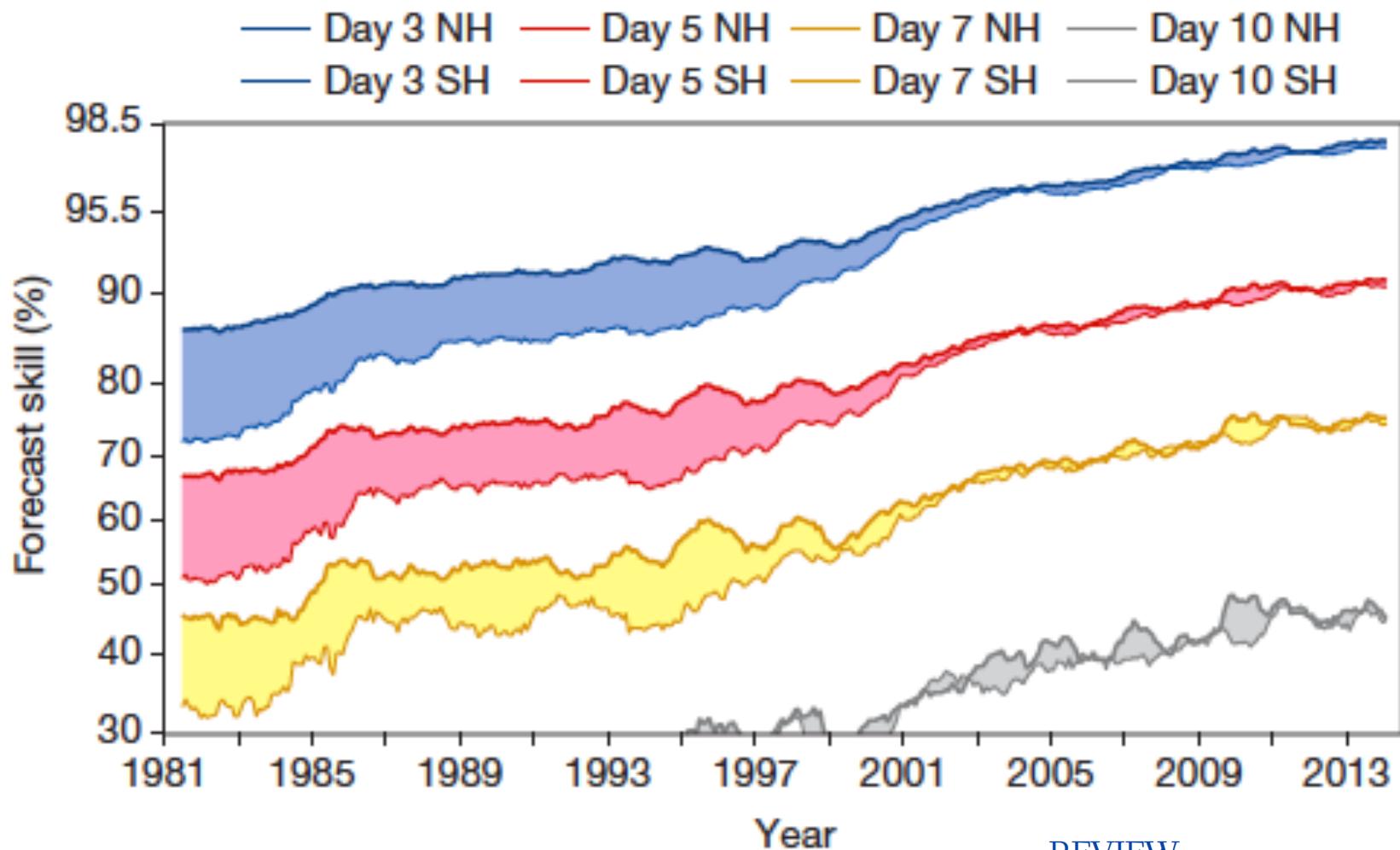
The quiet revolution of numerical weather prediction

Peter Bauer¹, Alan Thorpe¹ & Gilbert Brunet²



Forecasts Scores

ECMWF



REVIEW

The quiet revolution of numerical weather prediction
Peter Bauer¹, Alan Thorpe¹ & Gilbert Brunet²

doi:10.1038/nature14956

Métodos para assimilação de dados

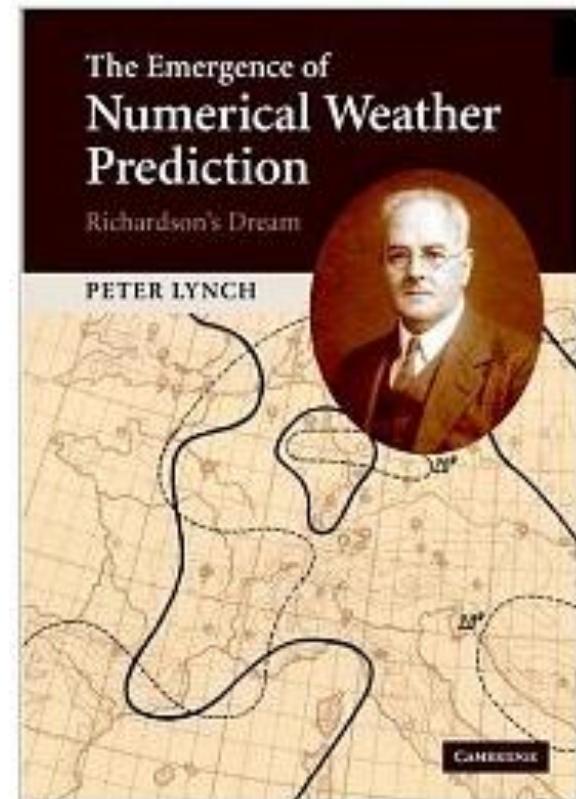
- Relaxação Newtoniana (*nudging*)
- Interpolação (“ótima”) estatística
- Filtro de Kalman
- Métodos Variationais: 3D e 4D
- Novos métodos em assimilação de dados:
 - *Ensemble Kalman filter* (EnKF)
 - Filtro de Partículas
 - Métodos híbridos
 - Redes neurais artificiais

História da assimilação de dados

- Tem início com a história da previsão numérica do tempo (PNT).
- Teorema de (Vilhelm) Bjerks (1904):
 - Previsão do tempo é resolver um problema de valor inicial.
 - Equação de Navier-Stokes da dinâmica de fluidos ...
- PNT: Lewis Fry Richardson (1922):
 - Livro: “Weather Prediction by Numerical Process”
 - A previsão descrita no livro falha ...
- John von Neumann e o “Projeto Meterorologia”
 - Jule Charney + Ragnar Fjortoft + von Neumann: *Tellus*, 1950.

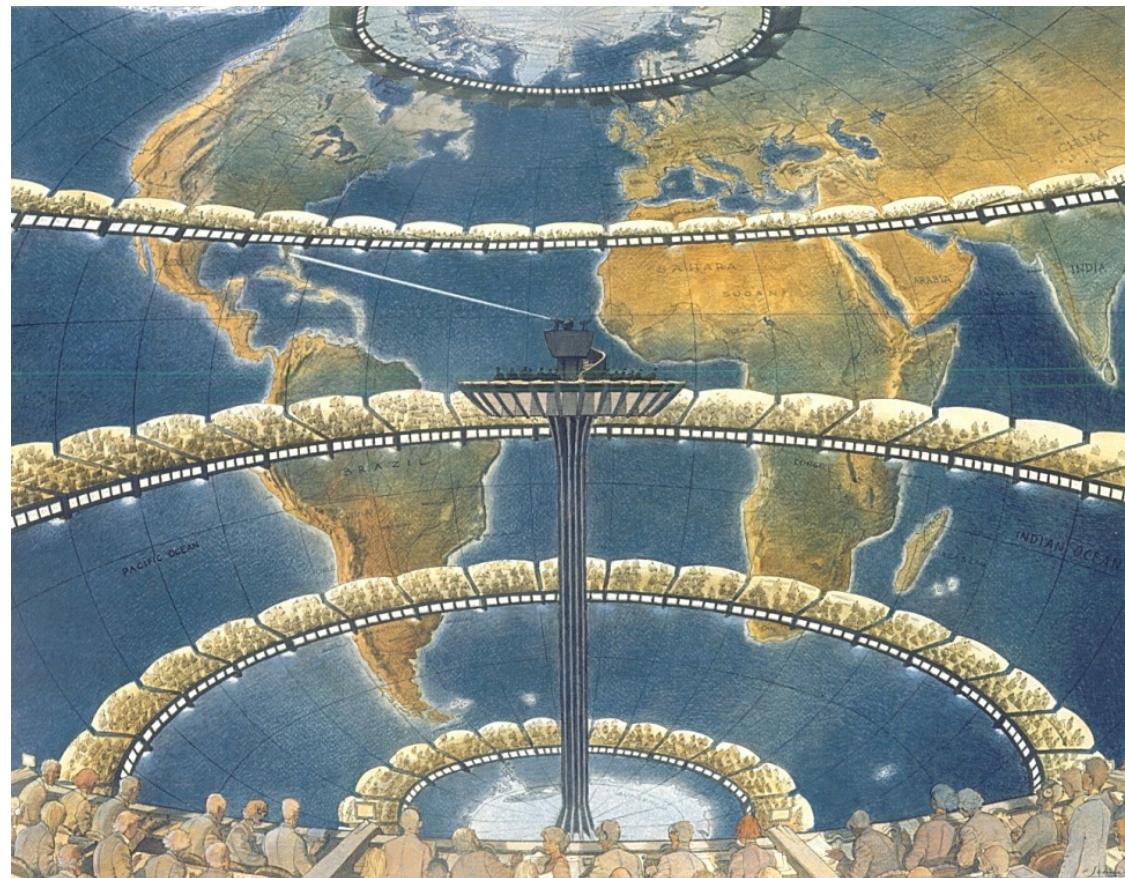
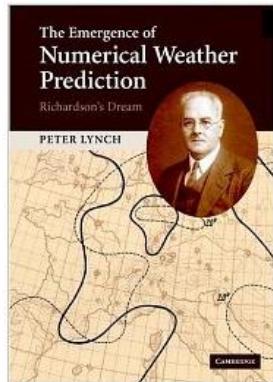
História da assimilação de dados

- Tem início com a história da previsão numérica do tempo (PNT).
- Descrição técnica de parte da história da PNT: ver livro do Dr. Peter Lynch:



História da assimilação de dados

- Descrição técnica de parte da história da PNT:
ver livro do Dr. Peter Lynch:



História da assimilação de dados

- Relaxação Newtoniana (*nudging*)
- Interpolação (“ótima”) estatística
- Filtro de Kalman
- Métodos Variationais: 3D e 4D
- Novos métodos em assimilação de dados:
 - *Ensemble Kalman filter* (EnKF)
 - Filtro de Partículas
 - Métodos híbridos
 - Redes neurais artificiais

Assimilação de Dados

- É um processo para determinar condição inicial.
- É um tipo de **problema inverso**.
- Problema inverso é parte da **teoria de estimação**.
- Ramo especializado da **análise de dados**.
- Combina **dados previsão com de observação**

Sociedade Brasileira de Matemática
Aplicada e Computacional 



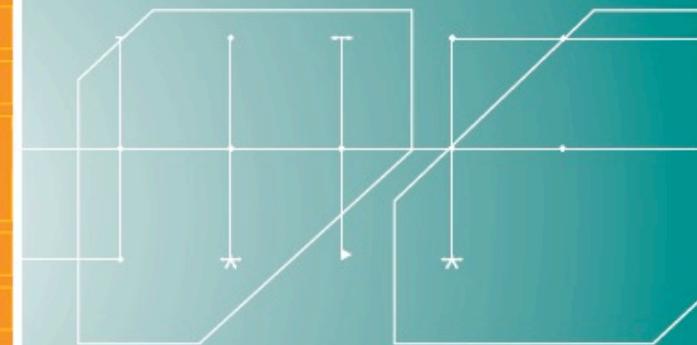
 FAPESP

PROBLEMAS INVERSOS EM PESQUISA ESPACIAL

Campos Velho

JDMAc

PROBLEMAS INVERSOS EM PESQUISA ESPACIAL



Uma publicação da Sociedade Brasileira de Matemática Aplicada e Computacional 

Haroldo Fraga de Campos Velho

Aplicações: mais informações no artigo

Inverse Problems in Space Science and Technology

Haroldo F. de Campos Velho¹, Fernando M. Ramos¹, Ezzat S. Chalhoub¹, Stephan Stephany¹, João C. Carvalho², Fabiano L. de Sousa³

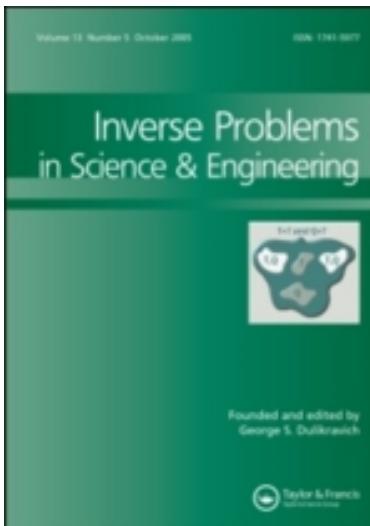
¹Laboratory for Computing and Applied Mathematics (LAC)

²Earth Observation Division (OBT)

³Mechanical Engineering Division (DEM)

National Institute for Space Research (INPE)

E-mail: [haroldo, fernando, ezzat, stephan]@lac.inpe.br, jcarlos@ltid.inpe.br, fabiano@dem.inpe.br
12245-970 – São José dos Campos (SP) – Brazil



[Inverse Problems in Science and Engineering](#), Volume [15](#) Issue 4 2007

Mini-Symposium on Inverse Problems Methods and Applications, 5th International Congress on Industrial and Applied Mathematics Sydney, Australia July 7–11, 2003

Increase in Pages in 2007!

ISSN: 1741-5985 (electronic) 1741-5977 (paper)

Foreword

[Special Issue-IPSE/ICIAM](#)

[291](#)

DOI: [10.1080/17415970600838954](https://doi.org/10.1080/17415970600838954)

[Inverse problems in space science and technology](#)

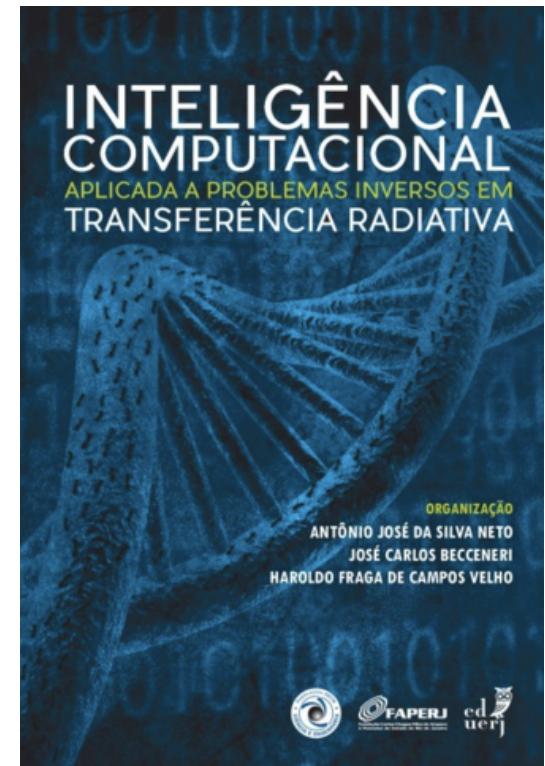
[359 – 372](#)

Authors: Haroldo F. De Campos Velho; Fernando M. Ramos; Ezzat S. Chalhoub; Stephan Stephany; João C. Carvalho; Fabiano L. De Sousa

DOI: [10.1080/17415970600839127](https://doi.org/10.1080/17415970600839127)

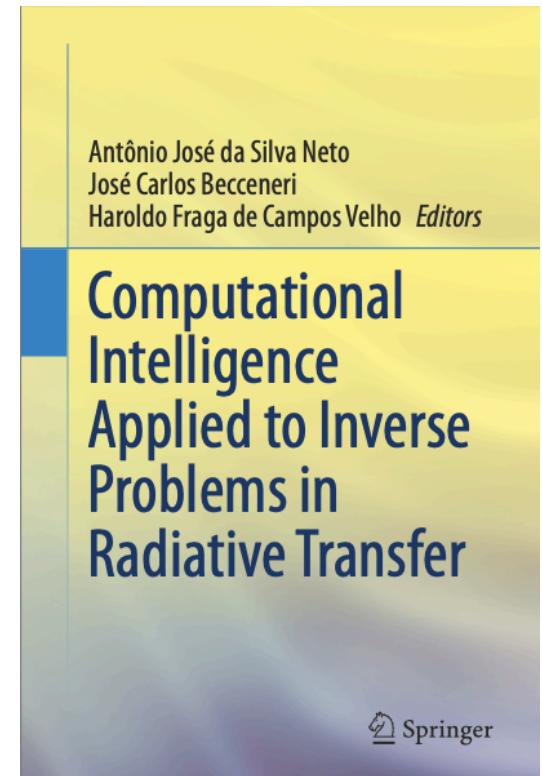
Meta-heuristics for inverse problems

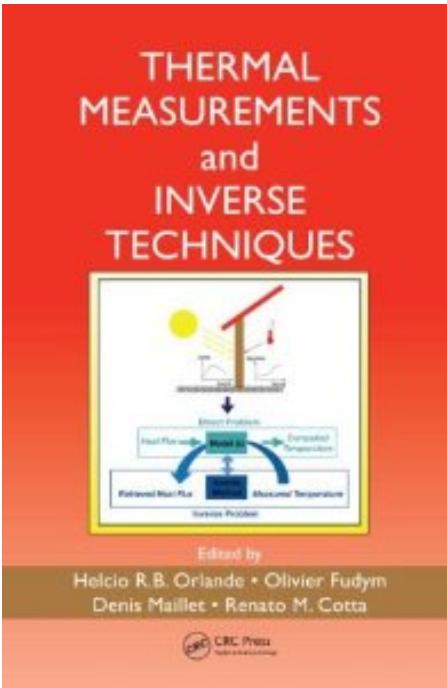
- Ant Colony Optimization (ACO)
- Multi-Particle Collision Algorithm (MPCA)
- Genetic algorithm (GA)
- And ... much more!!



Meta-heuristics for inverse problems

- Ant Colony Optimization (ACO)
- Multi-Particle Collision Algorithm (MPCA)
- Genetic algorithm (GA)
- And ... much more!!





8

Inverse Problems and Regularization

Haroldo F. de Campos Velho

CONTENTS

8.1	Introduction	283
8.1.1	Estimating Initial Condition in Heat Transfer	285
8.2	Regularization Method: Mathematical Formulation	288
8.3	Determining the Regularization Parameter	289
8.3.1	Generalized Discrepancy Principle.....	290
8.4	Tikhonov Regularization	290
8.5	Entropic Regularization	292
8.6	The Unified Regularization: The Principle of Maximum Nonextensive Regularization.....	296
8.7	Intrinsic Regularization	300
8.8	Statistical Methods	301
8.9	Regularized Neural Networks	306
8.10	Conclusion.....	309
	Appendix 8.A: Some Properties for Nonextensive Thermostatics.....	310
	Nomenclature	311
	References.....	312

Assimilação de Dados

- É um processo para determinar condição inicial.
- É um tipo de **problema inverso**.
- Problema inverso é parte da **teoria de estimação**.
- Ramo especializado da **análise de dados**.
- Combina **dados previsão com de observação**

Assimilação de Dados

- É um processo para determinar condição inicial.

$$\frac{\partial^2 T(x,t)}{\partial x^2} = \frac{1}{\alpha} \frac{\partial T(x,t)}{\partial t} \quad \text{for } 0 \leq x \leq L \text{ and } t > 0$$

$$\frac{\partial T(x,t)}{\partial x} = 0 \quad \text{at } x = 0 \text{ and } x = L$$

$$T(x,t) = f(x) \quad \text{at } t = 0$$

Goal: estimation of $f(x)$!

Assimilação de Dados

- **Problema inverso:** determinar condição inicial.
- Teoria da regularização: minimizar o funcional abaixo:

$$J_\alpha(f) = \|A(f) - \theta^\delta\|_2^2 + \alpha \Omega[f]$$

$A(f)$: Solução do problema direto

“ f ”: Condição inicial (função) candidata

θ^δ : temperatura medida em vários pontos da barra

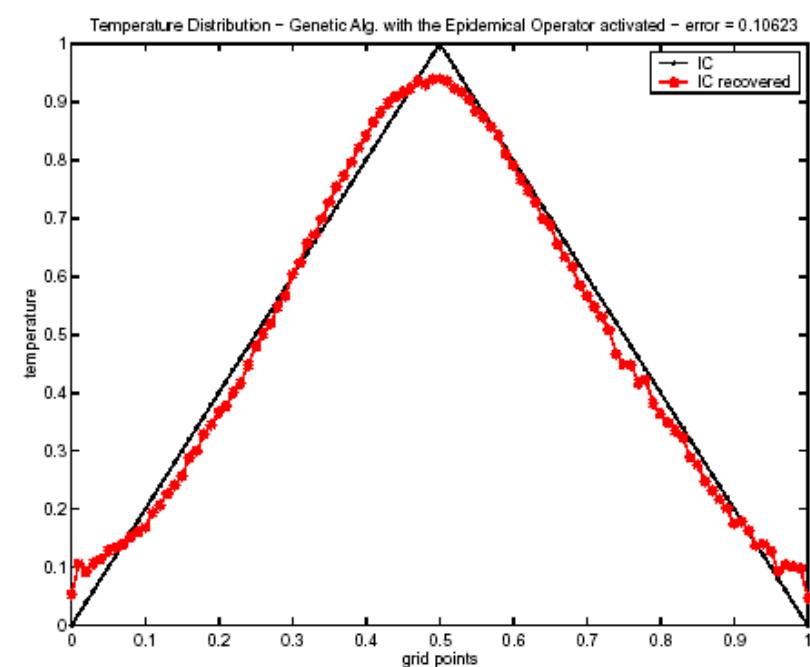
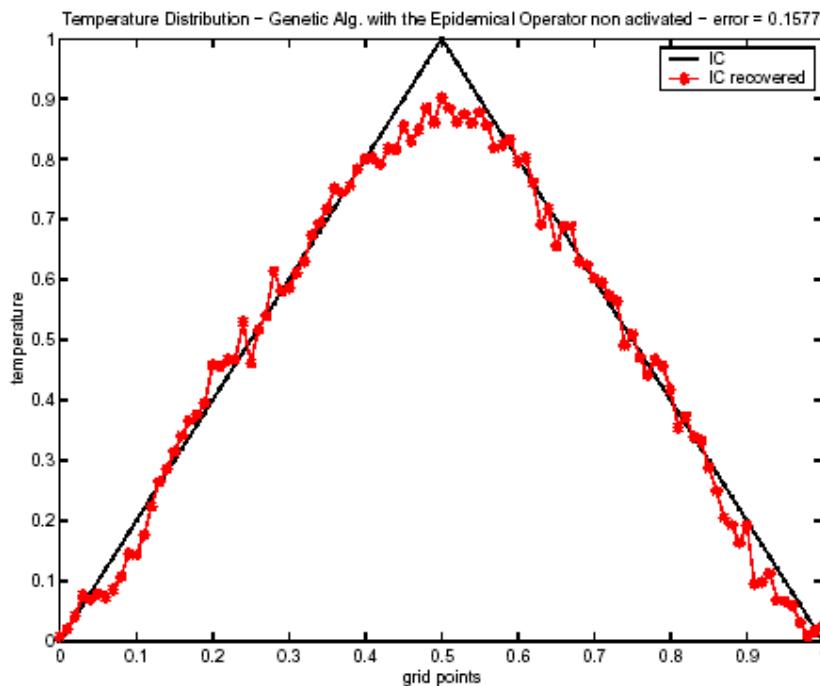
$\Omega[f]$: Operador de regularização

α : Parâmetro de regularização

Assimilação de Dados

- **Problema inverso:** determinar condição inicial

$$\text{Min} \left\{ J_\alpha(f) = \left\| A(f) - \theta^\delta \right\|_2^2 + \alpha \Omega[f] \right\}$$



Assimilação de Dados

■ **Problema inverso:** determinar condição inicial.



ELSEVIER

Journal of Computational and Applied Mathematics 103 (1999) 145–163

JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICS

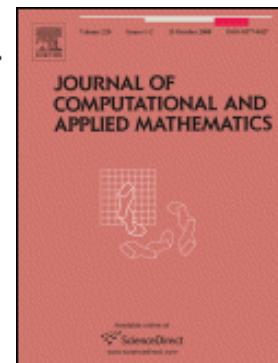
A comparison of some inverse methods for estimating the
initial condition of the heat equation

Wagner Barbosa Muniz^{a,*}, Haroldo F. de Campos Velho^b, Fernando Manuel Ramos^b

^a Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

^b Laboratório Associado de Computação e Matemática Aplicada, Instituto Nacional de Pesquisas Espaciais, CP 515,
12201-970 São José dos Campos, SP, Brazil

Received 29 October 1997; received in revised form 23 December 1997

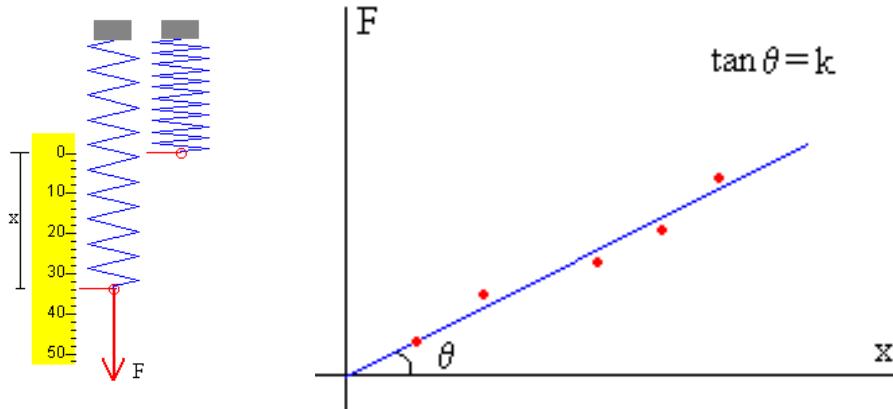


Assimilação de Dados

- É um processo para determinar condição inicial.
- É um tipo de **problema inverso**.
- Problema inverso é parte da **teoria de estimação**.
- Ramo especializado da **análise de dados**.
- Combina **dados previsão com de observação**

Estimation theory: Least square

The goal is to identify the “law” associated to the process under studying



Hooke's law: $F = k x$

Which is correct stiffness k ?

Estimation theory: Least square

Consider the linear model: $F(x) = \beta_0 + \beta_1 x$

We desire to estimate β_0 and β_1 according with some criteria, for example, minimizing the residuals (quadratic functional):

$$\min J(\beta_0, \beta_1) = \|F^{\text{Exp}} - F^{\text{Mod}}\|^2 = \|F^{\text{Exp}} - (\beta_0 + \beta_1 x)\|^2$$

The variation of a functional is given by

$$\delta J(f) = \nabla J \delta x = 0 \Rightarrow \boxed{\nabla J = 0}$$

Least square history

The Carl Frederic Gauss has developed the method at 1795 (by age of eighteen)



The expression “Least squares” was introduced by Legendre's (appendix: Nouvelles méthodes pour la détermination des orbites des comètes: “Sur la Méthode des moindres quarrés (On the method of least squares) – published in 1805).



Least square theory

The Robert Adrain (an american mathematician) in his paper “Research concerning the probabilities of errors which happen in making observations”, published in 1808, also derived the least square method.



Another and more rigorous proof was given by Gauss in the 1809 paper: “Theoria motus corporum coelestium.”

Least squares was developed independently by three different mathematicians, each contributed important parts of the theory.

Estimation theory: Least square

Consider the linear model: $F(x) = \beta_0 + \beta_1 x$

We desire to estimate β_0 and β_1 according with some criteria, for example, minimizing the residuals (quadratic functional):

$$\min J(\beta_0, \beta_1) = \|F^{\text{Exp}} - F^{\text{Mod}}\|^2 = \|F^{\text{Exp}} - (\beta_0 + \beta_1 x)\|^2$$

The variation of a functional is given by

$$\delta J(f) = \nabla J \delta x = 0 \Rightarrow$$

$$\nabla J = 0$$

Least square estimation

1. Linear regression

$$\nabla J \equiv \frac{\partial J(\beta_0, \beta_1)}{\partial (\beta_0, \beta_1)} \equiv \begin{bmatrix} \partial J / \partial \beta_0 \\ \partial J / \partial \beta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

3. Carrying out the derivatives:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^m [y^{\text{Exp}} - (\beta_0 + \beta_1 x_i)]$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^m x_i [y^{\text{Exp}} - (\beta_0 + \beta_1 x_i)]$$

Least square estimation

1. Linear regression

$$\begin{cases} \beta_0 m + \beta_1 \sum_{i=1}^m x_i = \sum_{i=1}^m y_i^{\text{Exp}} \\ \beta_0 \sum_{i=1}^m x_i + \beta_1 \sum_{i=1}^m x_i^2 = \sum_{i=1}^m x_i y_i^{\text{Exp}} \end{cases}$$

3. Matrix formula:

$$\begin{bmatrix} m & \sum_{i=1}^m x_i \\ \sum_{i=1}^m x_i & \sum_{i=1}^m x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m y_i^{\text{Exp}} \\ \sum_{i=1}^m x_i y_i^{\text{Exp}} \end{bmatrix}$$

Least square estimation

The linear system has a simple solution:

$$\beta_1 = \frac{\sum_i (x_i - \bar{x})(y_i^{\text{Exp}} - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$\begin{aligned}\beta_0 &= \frac{1}{m} \left(\sum_{i=1}^m y_i^{\text{Exp}} - \beta_1 \sum_{i=1}^m x_i \right) = \langle y^{\text{Exp}} \rangle - \beta_1 \langle x \rangle \\ &= \bar{y}^{\text{Exp}} - \beta_1 \bar{x}\end{aligned}$$

Least square estimation

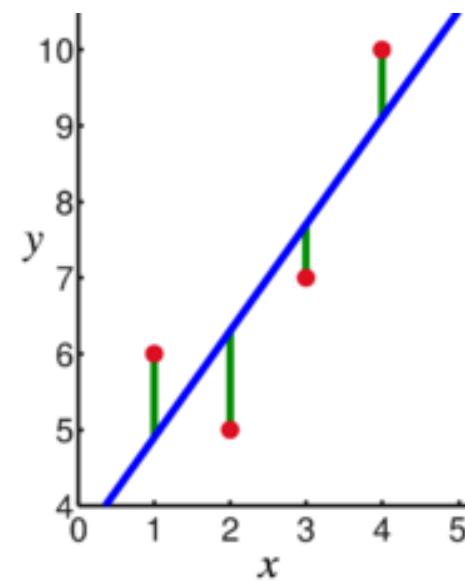
Linear regression example:

1. Considering the set points (see figure):

(1,6), (2,5), (3,7), (4,10)

2. Linear equations:

$$\begin{cases} \beta_0 + 1\beta_1 = 6 \\ \beta_0 + 2\beta_1 = 5 \\ \beta_0 + 3\beta_1 = 7 \\ \beta_0 + 4\beta_1 = 10 \end{cases}$$



Example from the Wikipedia



Least square estimation

Linear regression example:

1. The functional to be minimized:

$$\begin{aligned} J(\beta_0, \beta_1) = & [6 - (\beta_0 + 1\beta_1)]^2 + [5 - (\beta_0 + 2\beta_1)]^2 \\ & + [7 - (\beta_0 + 3\beta_1)]^2 + [10 - (\beta_0 + 4\beta_1)]^2 \end{aligned}$$

2. Solution:

$$\begin{cases} \beta_0 = 3.5 \\ \beta_1 = 1.4 \end{cases}$$

Example from the Wikipedia



Least square estimation

Another looks at least square solution:

1. Equivalent statement: find a minimum

$$J(\beta) = (\mathbf{A}\beta - \mathbf{y})^T (\mathbf{A}\beta - \mathbf{y}) \equiv \mathbf{e}^T \mathbf{e}$$

2. Solution (normal equations):

$$\mathbf{A}^T \mathbf{A}\beta = \mathbf{A}^T \mathbf{y} \Rightarrow \beta = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

3. Note ($m=4$; $n=2$):

$$\mathbf{A}_{m \times n} \Rightarrow (\mathbf{A}^T \mathbf{A})_{n \times n}$$

Example from the Wikipedia



Least square estimation

Another looks at least square solution:

1. Algebraic solution:

$$\mathbf{A}^T \mathbf{A} \boldsymbol{\beta} = \mathbf{A}^T \mathbf{y} \Rightarrow \boldsymbol{\beta} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}$$

2. The right hand side matrix is called the Moore-Penrose generalized inverse:

$$\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \Rightarrow \boldsymbol{\beta} = \mathbf{A}^+ \mathbf{y}$$

Example from the Wikipedia



Least square estimation

How good is the least square solution?

Least square estimation

How good is the least square solution?

The answer for the above question is addressed by the Gauss-Markov theorem:

*“For a linear model in which the error have zero mean and are uncorrelated and have equal variances, a **best linear unbiased estimator (BLUE)** of the coefficients is given by the least-square estimator.”*

Least square estimation

The Gauss-Markov theorem

1. Assumptions:

- a. True linear model: $y_i = \sum_j \beta_j x_i + \varepsilon_i$
- b. $E\{\varepsilon_i\} = 0$
- c. $V\{\varepsilon_i\} = E\{(\varepsilon_i - \bar{\varepsilon})^2\} = \sigma^2 < \infty$
- d. Proof:

Let be $\tilde{\beta}$ another linear estimator of β :

$$\tilde{\beta} = [(x^T x)^{-1} x^T + D] y \equiv C y$$



Least square estimation

The Gauss-Markov theorem – proof

1. Expected value for the $\tilde{\beta}$:
2. Remembering: $\tilde{\beta} = [(x^T x)^{-1} x^T + D]y$
3.
$$\begin{aligned} E\{Cy\} &= E\{[(x^T x)^{-1} x^T + D](x\beta + \varepsilon)\} \\ &= [(x^T x)^{-1} x^T + D]x\beta + \underbrace{[(x^T x)^{-1} x^T + D]E\{\varepsilon\}}_{\text{zero}} \\ &= [(x^T x)^{-1} x^T x]\beta + Dx\beta = (I + Dx)\beta \end{aligned}$$

$$E\{\tilde{\beta}\} = \beta \quad \text{iff} \quad Dx = 0.$$

Least square estimation

The Gauss-Markov theorem – proof

1. Variance of $\tilde{\beta}$:

$$2. V\{\tilde{\beta}\} = V\{Cy\} = CV\{y\}C^T = \sigma^2 CC^T$$

$$= \sigma^2 [(x^T x)^{-1} x^T + D][x(x^T x)^{-1} + D^T]$$

$$= \sigma^2 [(x^T x)^{-1} x^T x(x^T x)^{-1} + (x^T x)^{-1} x^T D^T + D x(x^T x)^{-1} + D D^T]$$

Note: $V\{X\} = E\{(X - \mu)(X - \mu)^T\}$

Least square estimation

The Gauss-Markov theorem – proof

1. Variance of $\tilde{\beta}$:

$$\begin{aligned} 2. \quad V\{\tilde{\beta}\} &= V\{Cy\} = CV\{y\}C^T = \sigma^2 CC^T \\ &= \sigma^2(x^T x)^{-1} + \sigma^2(x^T x)^{-1} \underbrace{(Dx)^T}_{\text{zero}} \\ &\quad + \sigma^2 \underbrace{Dx(x^T x)^{-1}}_{\text{zero}} + \sigma^2 DD^T \\ &= \underbrace{\sigma^2(x^T x)^{-1}}_{V(\beta)} + \sigma^2 DD^T \end{aligned}$$

Least square estimation

The Gauss-Markov theorem – proof

1. Conclusion:
2. The variance of new estimation $V\{\tilde{\beta}\}$ exceeds the previous one $V\{\beta\}$ by a positive semidefinite matrix (DD^T).

Least square estimation

Multiple linear regression method

1. True model:

$$y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

2. Computing the estimation:

$$\boldsymbol{b}^T = [b_0 \ b_1 \ b_2 \ \dots \ b_n]$$

3. Residuals functional (to be minimized):

$$J(\boldsymbol{b}) = \|e\|^2 = \|y^{\text{Exp}} - y^{\text{Mod}}\|^2$$

Least square estimation

Multiple linear regression method

1. Experimental data set ($m > n$)

$$\mathbf{y}^{\text{Exp}} = [(y_1)^{\text{Exp}} \quad (y_2)^{\text{Exp}} \quad \dots \quad (y_m)^{\text{Exp}}]^T$$

2. Relation between model and experiments:

$$\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} y_1^{\text{Exp}} \\ y_2^{\text{Exp}} \\ \vdots \\ y_m^{\text{Exp}} \end{bmatrix} \Leftrightarrow \mathbf{X}\mathbf{b} = \mathbf{y}^{\text{Exp}}$$

Least square estimation

Multiple linear regression method

1. Expressing the normal equations ($m > n$)

$$\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}^T \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}^T \begin{bmatrix} y_1^{\text{Exp}} \\ \vdots \\ y_m^{\text{Exp}} \end{bmatrix}$$



$$\mathbf{X}^T \mathbf{X} \mathbf{b} = \mathbf{X}^T \mathbf{y}^{\text{Exp}}$$

Least square estimation

Weighted least square

1. For linear solution of least squares, the vector \mathbf{b} is BLUE (*best linear unbiased estimator*), due to Gauss-Markov theorem.
2. If the measurements are uncorrelated, but have different uncertainties, a modified approach should be applied, keeping feature BLUE.
3. Considering de new functional:

$$J(\mathbf{b}) = \sum_{i=1}^n W_{ii} e_i^2 , \quad W_{ii} = \frac{1}{\sigma_i^2}$$

Least square estimation

Weighted least square

- For this situation, the normal equations become:

$$\sum_{i=1}^n \sum_{k=1}^m X_{ij} W_{ii} X_{ik} b_k = \sum_{i=1}^n X_{ij} W_{ii} y_i^{\text{Exp}} , \quad j = 1, \dots, m$$



$$(\mathbf{X}^T \mathbf{W} \mathbf{X}) \mathbf{b} = \mathbf{X}^T \mathbf{W} \mathbf{y}^{\text{Exp}}$$

Least square estimation

Parameter error, and confidence limits

1. Co-variance observations errors: matrix M .
2. Co-variance parameters errors: matrix M^b .

$$\mathbf{M}^b = [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}] \mathbf{M} [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}]^T$$

3. If: $\mathbf{W} = \mathbf{M}^{-1}$, the expression above is simplified

$$\mathbf{M}^b = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}$$

Least square estimation

Parameter error, and confidence limits

1. When the unit weights are used ($W=I$), the experimental errors are uncorrelated and all equal: $M = \sigma^2 I$.
2. In this case:

$$\sigma^2 \approx \frac{J_{\min}(\mathbf{b})}{m - n} \Rightarrow \mathbf{M}^{\mathbf{b}} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}$$

Least square estimation

Parameter error, and confidence limits

1. Under the Normal distribution assumption (with zero mean, and σ^2 variance), the following confidence limits can be derived (*):

68% confidence limit $b \pm \sigma$

95% confidence limit $b \pm 2\sigma$

99% confidence limit $b \pm 2.5\sigma$

- (*) The Gaussian assumption make sense if $m \gg n$. If the experimental errors are Gaussian, the parameter errors belongs to a t-Student with $m-n$ degrees of freedom.

Least square estimation

Recurssive Least Square

1. Supposing that we have a first set of data. For this set the least squares solution is

$$\mathbf{b}_0 = (\mathbf{X}_0^T \mathbf{W}_0 \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{W}_0 \mathbf{y}^{\text{Exp}}$$

3. The error ($\mathbf{b} - \mathbf{b}_0$) has zero mean and its covariance matrix is

$$P_0 = E\{(\boldsymbol{\beta} - \mathbf{b}_0)(\boldsymbol{\beta} - \mathbf{b}_0)^T\} = (\mathbf{X}_0^T \mathbf{W}_0 \mathbf{X}_0)^{-1}$$

Least square estimation

Recurssive Least Square

1. Question: If more experimental data are available (we could be dealing of data pipeline), can the best estimate for the combined system

$$\mathbf{X}_0 \mathbf{b}_0 = \mathbf{y}_0^{\text{Exp}} \quad \mathbf{X}_1 \mathbf{b}_1 = \mathbf{y}_1^{\text{Exp}}$$

be computed from \mathbf{b}_0 and $(\mathbf{y}_1)^{\text{Exp}}$ without restarting the calculation from $(\mathbf{y}_0)^{\text{Exp}}$?

Least square estimation

Recurssive Least Square

1. The goal here is to estimate \mathbf{b}_1 from the previous estimation \mathbf{b}_0 and from the new data $(\mathbf{y}_1)^{\text{Exp}}$.
2. The final result should be the same as computing from the standard procedure. The co-variance new matrix is

$$\mathbf{W} = \begin{bmatrix} W_0 & 0 \\ 0 & W_1 \end{bmatrix} \text{ is the covariance of the errors : } \begin{bmatrix} e_0 \\ e_1 \end{bmatrix}$$

Least square estimation

Recurssive Least Square

1. The matrix \mathbf{W} is block-diagonal because e_1 is independent of e_0 .
2. The coefficient matrix for the equation \mathbf{b}_1 is

$$P_1^{-1} = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}^T \begin{bmatrix} W_0 & 0 \\ 0 & W_1 \end{bmatrix}^{-1} \begin{bmatrix} X_0 \\ X_1 \end{bmatrix} = X_0^T W_0^{-1} X_0 + X_1^T W_1^{-1} X_1$$

Least square estimation

Recurssive Least Square

1. The solution b_1 is not based only on $(y_1)^{\text{Exp}}$, it is the best for the combined system $X_0 b_0 = (y_0)^{\text{Exp}}$ $X_1 b_1 = (y_1)^{\text{Exp}}$.
2. The normal equations are:

$$(X^T W^{-1} X)b = X^T W^{-1} y^{\text{Exp}}$$

5. The matrix on the left hand side is exactly P^{-1} .

Least square estimation

Recurssive Least Square

1. The optimal b_1 is:

$$b_1 = P_1 \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}^T W^{-1} \begin{bmatrix} y_0^{\text{Exp}} \\ y_1^{\text{Exp}} \end{bmatrix}^T = P_1 \underbrace{(X_0^T W_0^{-1})}_{P_0^{-1}} y_0^{\text{Exp}} + X_1^T W_1^{-1} y_1^{\text{Exp}}$$

3. In the solution above, we want to compute by recursion procedure (using the already computed b_0 in place of $(y_0)^{\text{Exp}}$).

Least square estimation

Recurssive Least Square

- Therefore, we up date the matrix P along with the estimate b :

$$P_1^{-1} = P_0^{-1} + X_1^T W_1^{-1} X_1$$

- The actual estimate b_1 has to be based on the actual $(y_0)^{\text{Exp}}$ and $(y_1)^{\text{Exp}}$. We can re-write the formula:

$$\begin{aligned} b_1 &= P_1(P_0^{-1}b_0 + X_1^T W_1^{-1} y_1^{\text{Exp}}) \\ &= P_1(P_1^{-1}b_0 - X_1^T W_1^{-1} X_1 b_0 + X_1^T W_1^{-1} y_1^{\text{Exp}}) \\ &= b_0 + K_1(b_0 - X_1 b_0) \end{aligned}$$

Least square estimation

Recurssive Least Square

1. The matrix K_1 is called the gain matrix.

$$K_1 = P_1 X_1^T W_1^{-1}$$

2. with such manipulation the formula become recursive, it is b_0 instead $(y_0)^{\text{Exp}}$.
3. The recursive formula:

$$P_k^{-1} = P_{k-1}^{-1} + X_k^T W_k^{-1} X_k$$

$$b_k = b_{k-1} + K_k (y_k^{\text{Exp}} - X_k b_{k-1}) \quad \text{with : } K_k = P_k X_k^T W_k^{-1}$$

Kalman filter

1. This algorithm is considered one of the most important advance for forecasting and filtering.
To estimate the unknown b_0 vector at each step we have two equations:
2. Measurements $(y_0)^{\text{Exp}}, \dots, (y_m)^{\text{Exp}}$ have been taken up to the time $t=m$. If these were exact they would be related to the true values by .
But, the measurements are not exact, and the matrices are rectangular.

Kalman filter

1. From the dynamical law

$$b_{i+1} = F_i b_i \quad (*)$$

That governs the change in state as time goes forward.

2. This time it is an error ε_i in the model, instead of an error e_i in the measurement.
3. Thus our equations are the two kinds: diagnosis and forecasting.

Kalman filter

1. Estimation and prediction equations

$$X_i b_i = y_i \quad (\text{with error } e_i)$$

$$-F_i b_i + b_{i+1} = 0 \quad (\text{with error } \varepsilon_i)$$

3. We want to combine these equations into a single system. Suppose we know y_1, y_2, y_3 ; we are already to estimate b_2 .

Kalman filter

1. Putting the equations together

$$\begin{bmatrix} X_0 \\ -F_0 & I \\ & X_1 \\ & -F_1 & I \\ & & X_2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} y_0 \\ 0 \\ y_1 \\ 0 \\ y_2 \end{bmatrix}$$

3. The goal of filtering is to find the best solution to this system.

Kalman filter

1. Following Gauss, we should solve the last system by least squares (with his weighting matrix).
2. According to Kalman, we should solve the system recursively: the work should not repeat the previous step, and it should be put to use at the next step.
3. The difference between Kalman strategy, and the recursive least squares is the matrix P is not evolving – this not occur in the Kalman filter.

Kalman filter

Difference between Kalman filter and recursive least square:

1. The transition matrix between two states is not dependent on time for recursive least squares.
2. The modeling error is not taken into account for the recursive least squares.

Kalman filter

1. Kalman filter is a procedure with 2 steps:

$$\text{Forecast: } \mathbf{x}_{n+1}^f = F \left\{ \mathbf{x}_n^f, t_n \right\}$$

$$\text{Analysis: } \mathbf{x}_{n+1}^a = \mathbf{x}_{n+1}^f + \mathbf{d}_{n+1}$$

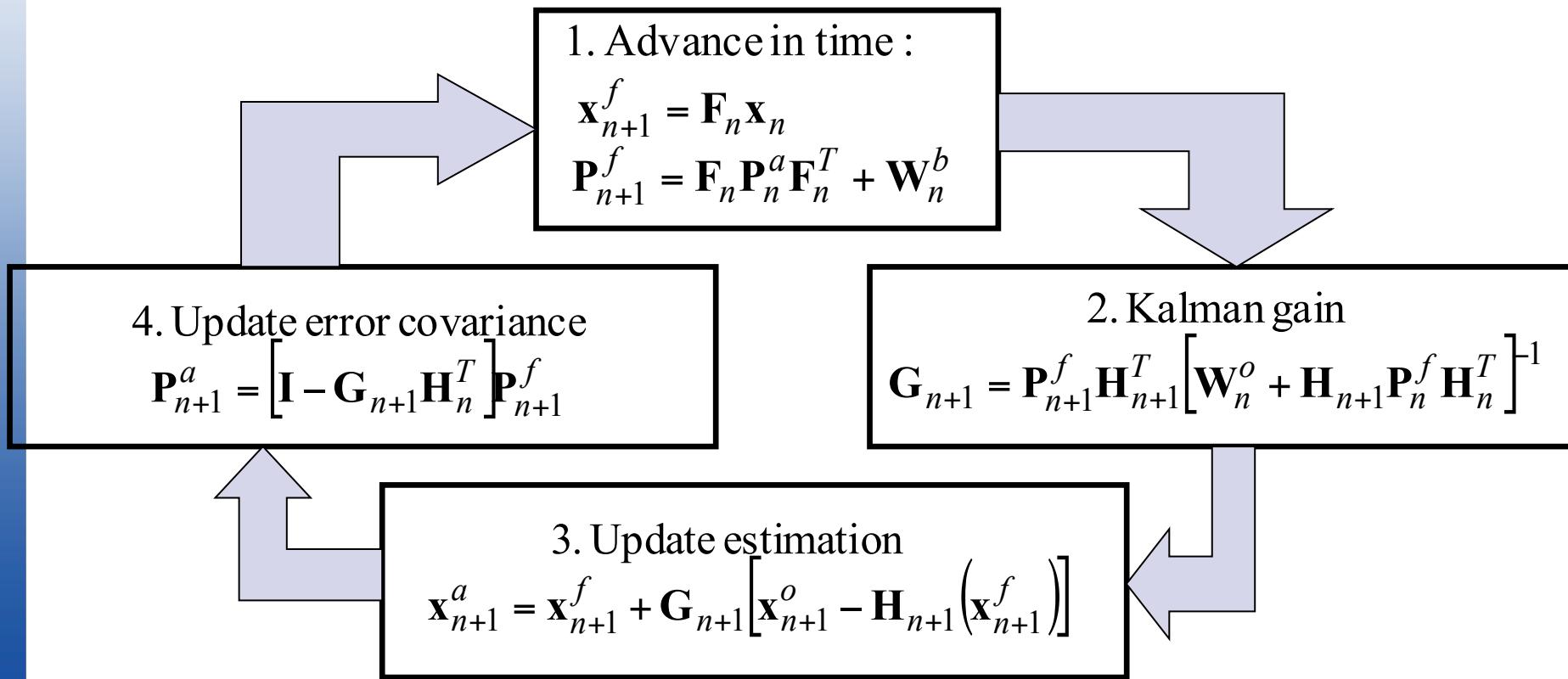
- where \mathbf{d}_{n+1} is the innovation:

$$\mathbf{d}_n = \mathbf{G}_n p \left[\mathbf{x}_n^o - H_n \left(\mathbf{x}_n^f \right) \right] \quad \begin{cases} \mathbf{G}_n: \text{the gain matrix} \\ p[.]: \text{discrepancy function} \end{cases}$$

- being \mathbf{x}_n^o the observed state, and H_n represents the observation system.

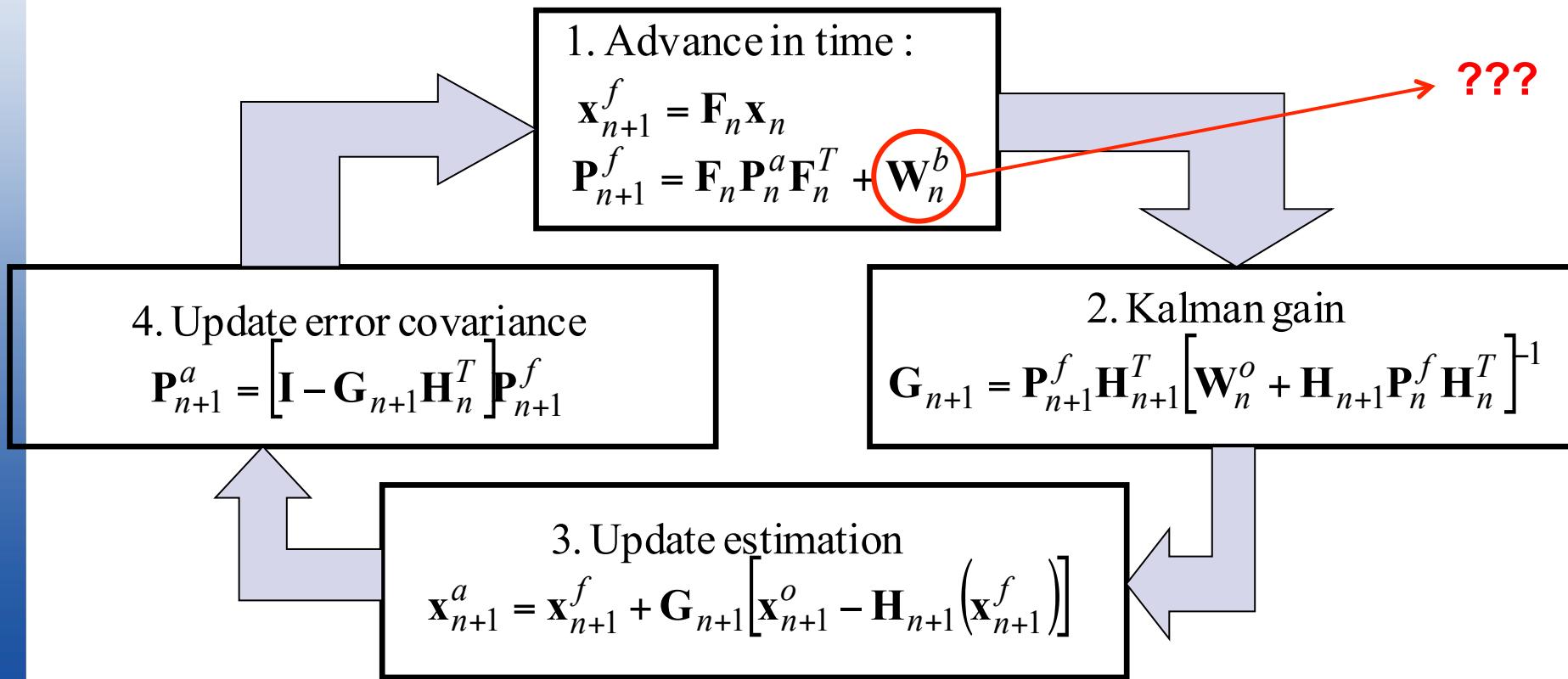
Kalman filter

Three versions: Linear, Extended, Adaptive (next slide)



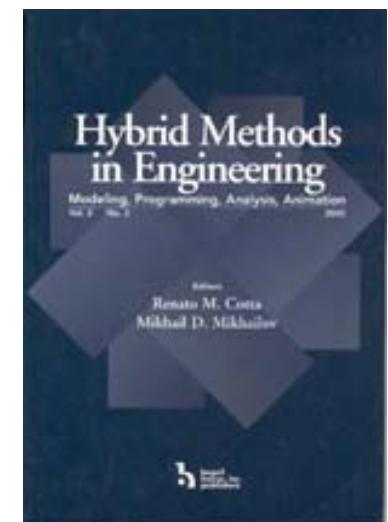
Kalman filter

Three versions: Linear, Extended, Adaptive (next slide)



Kalman filter

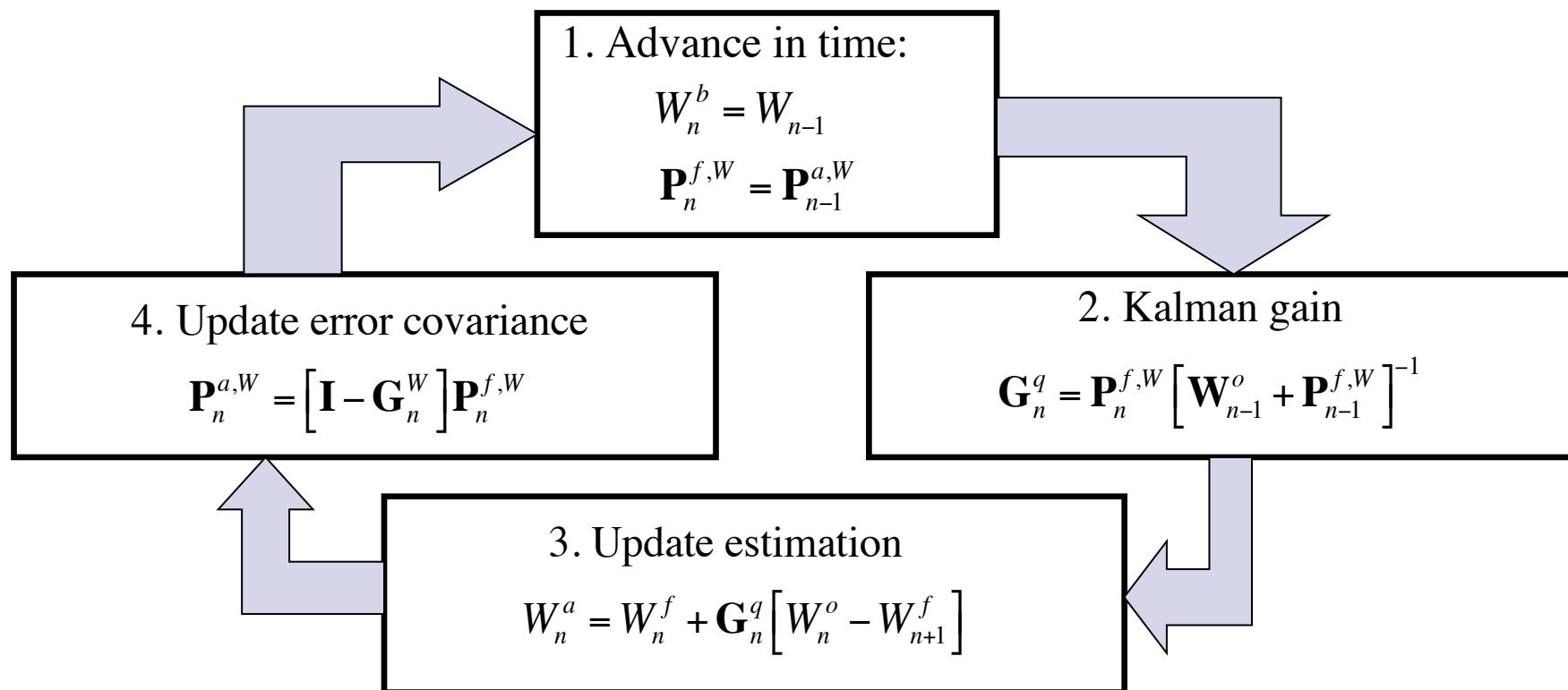
- **Identifying modeling error co-variance matrix**
- Three schemes:
 - Estimating W^b by parameterization
 - Estimating W^b by Fokker-Planck equation
 - Estimating W^b by ensemble strategy



Kalman filter

$$\mathbf{x}_{n+1} = F[\mathbf{x}_n, t_n] \approx \mathbf{F}_n + \frac{\partial F}{\partial \mathbf{x}} \Big|_{t=t_n} \mathbf{x}_n + O(\Delta t^2) \approx \mathbf{E}_n \mathbf{x}_n$$

Three versions: Linear, Extended, **Adaptive**



A.G. Nowosad, A. Rios Neto, H.F. de Campos Velho: Data Assimilation Using an Adaptive Kalman Filter and Laplace Transform, Hybrid Methods in Engineering, 2(3), 291-310, 2000.

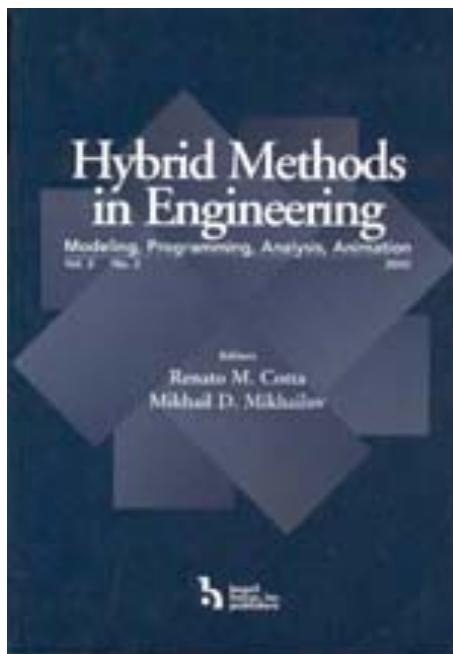
Data assimilation: adaptive Kalman filter

Data Assimilation Using an Adaptative Kalman Filter and Laplace Transform

A.G. Nowosad^a (DCM)

A. Rios Neto^b

H.F. de Campos Velho^a (LAC)



^a Instituto Nacional de Pesquisas Espaciais (INPE)
Caixa Postal 515
12201-970 – São José dos Campos (SP), BRAZIL,
E-mail: alex@met.inpe.br - haroldo@lac.inpe.br

^b Instituto de Pesquisa e Desenvolvimento (IP&D)
Universidade do Vale do Paraíba (UNIVAP)
Av. Shishima Hifumi, 2.911 - Urbanova
12245-720 - São José dos Campos (SP), BRAZIL
E-mail: atair@univap.br

Hybrid Methods in Engineering: (2000) 2(3): 291-310

Data assimilation in geophysical fluid dynamics: multi-step process

- Collecting data (observational system: ground stations, radiosondes, satellite, radar, aircrafts, ...).
- Observation data: quality control (data checking).
- Data assimilation (objective analysis):
 - Data fusion: observations + mathematical model = analysis.
- Initialization:
 - Filtering out high frequencies.

O sistema de Lorenz



AMERICAN METEOROLOGICAL SOCIETY
AMS Journals Online

AMS Home Journals Home Journal Archive Subscribe For Authors Help Advanced Search

Abstract View

[Volume 20, Issue 2 \(March 1963\)](#)

Journal of the Atmospheric Sciences
Article: pp. 130–141 | [Abstract](#) | [PDF \(1.02M\)](#)

Deterministic Nonperiodic Flow

Edward N. Lorenz

Massachusetts Institute of Technology

(Manuscript received November 18, 1962, in final form January 7, 1963)

DOI: [10.1175/1520-0469\(1963\)020<0130:DNF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2)

O sistema de Lorenz

Lorenz's attractor



Edward N. Lorenz

Simplest chaotic system:
the Lorenz Attractor.
A metaphor for weather
unpredictability

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

$$\frac{dz}{dt} = xy - \beta z$$

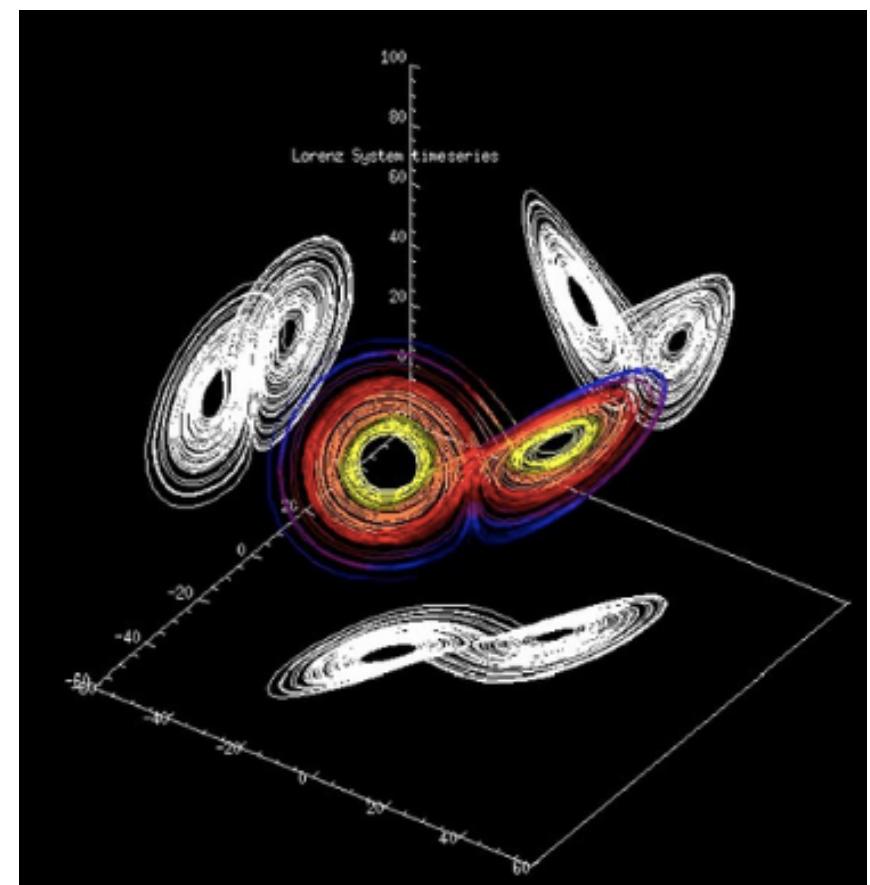
O sistema de Lorenz

Lorenz's attractor

$$\frac{dx}{dt} = \sigma(y - x)$$

$$\frac{dy}{dt} = x(\rho - z) - y$$

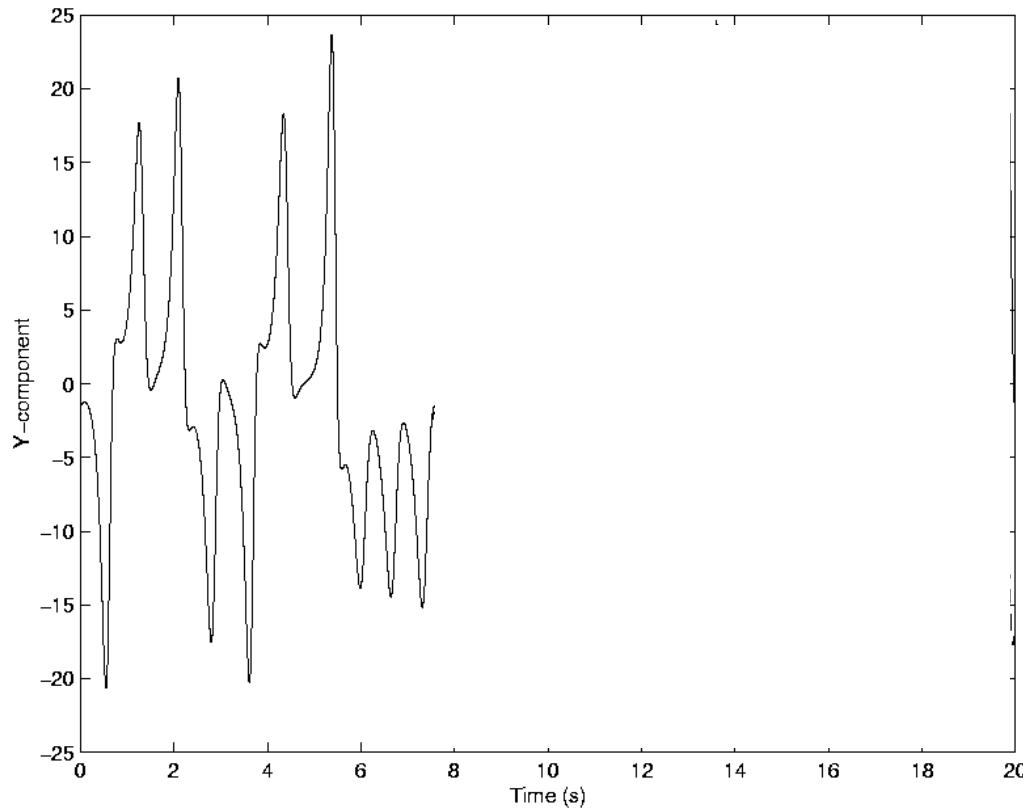
$$\frac{dz}{dt} = xy - \beta z$$



O sistema de Lorenz

Dinâmica caótica no sistema de Lorenz

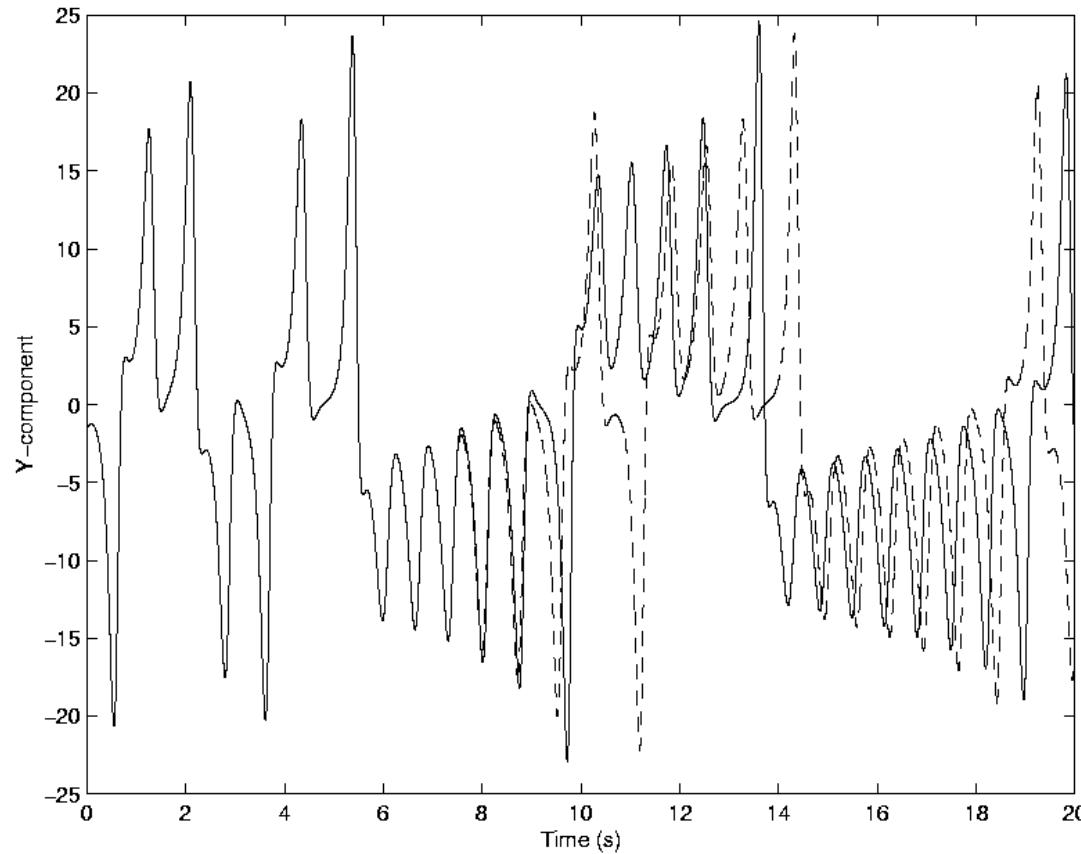
- Two initial conditions (Y component): (w_0) and $(w_0 + \Delta w)$



O sistema de Lorenz

Dinâmica caótica no sistema de Lorenz

- Two initial conditions (Y component): (w_0) and $(w_0 + \Delta w)$



Relaxação newtoniana (nudging)

$$\frac{dA_m(t)}{dt} \approx \frac{A_m(t + \Delta t) - A_m(t - \Delta t)}{2\Delta t} = F_m(A, t) + N[A_m^{\text{Obs}}(t) - A_m(t)]$$



- N : relaxation coefficient (from numerical experimentation)
- The idea is to integrate the model up to a future time (observations), considering the forcing (last rhs term).
- The procedure have had less use, because sometimes it is not able to follow a chaotic dynamics.

Nudging: resultado de convergência

- Modelo:
$$\begin{cases} \frac{dX}{dt} = F(X) , & 0 < t < T \\ X(0) = X_0 \end{cases}$$

- Nudging:

$$\begin{cases} \frac{dX}{dt} = F(X) + \mathbf{K}(\mathbf{Y}_{obs} - \mathbf{H}(X)) , & 0 < t < T \\ X(0) = X_0 \end{cases}$$

Nudging: resultado de convergência

- Modelo com nudging:

$$\frac{dX}{dt} = FX + K(Y_{obs} - HX)$$

- Discretização:

$$\frac{X^{n+1} - X^n}{\Delta t} = FX^{n+1} + K(Y_{obs}^{n+1} - HX^{n+1})$$

Nudging: resultado de convergência

- A solução X deve satisfazer a equação:

$$X^{n+1} - X^n = \Delta t F X^{n+1} + \Delta t H^T R^{-1} (Y_{obs}^{n+1} - H X^{n+1})$$

- Assumindo:

$$Y_{obs} = H X_{true}$$

Nudging: resultado de convergência

- Modelo:

$$\begin{cases} \frac{dX}{dt} = FX + K(Y_{obs} - HX), & 0 < t < T \\ X(0) = X_0 \end{cases}$$

- Que pode ser escrita como:

$$\begin{cases} \frac{dX}{dt} = FX + KH(X_{true} - HX), & 0 < t < T \\ X(0) = X_0 \end{cases}$$

Nudging: resultado de convergência

- Equação do erro: $E = X - X_{true}$

$$\begin{cases} \frac{dE}{dt} = \frac{dX}{dt} - \frac{dX_{true}}{dt} = FX + KH(X_{true} - X) - FX_{true}, & 0 < t < T, \\ E(0) = X(0) - X_{true}(0) = X_0 - X_{true,0} \end{cases}$$

$$\begin{cases} \frac{dE}{dt} = F(X - X_{true}) + KH(X_{true} - X) = (\mathbf{F} - \mathbf{K}\mathbf{H})E, & 0 < t < T, \\ E(0) = E_0. \end{cases}$$

Nudging: resultado de convergência

- Equação do erro: $E = X - X_{true}$

$$\frac{dE}{dt} = (F - KH)E$$

- Integrando: $E(t) = \exp^{(F - KH)t} E(0)$
- E o erro tende para zero quando:

$$\max Re(Sp(F - KH)) < 0$$

Relaxação newtoniana (nudging)

Revista Brasileira de Meteorologia, v.17, n.2, 125-140, 2002

A FIRST-GUESS FIELD PRODUCED BY MERGING DIGITAL FILTER AND NUDGING TECHNIQUES⁽¹⁾

VALDIR INNOCENTINI⁽²⁾, ERNESTO CAETANO⁽³⁾ AND FABRICIO PEREIRA HARTER⁽²⁾

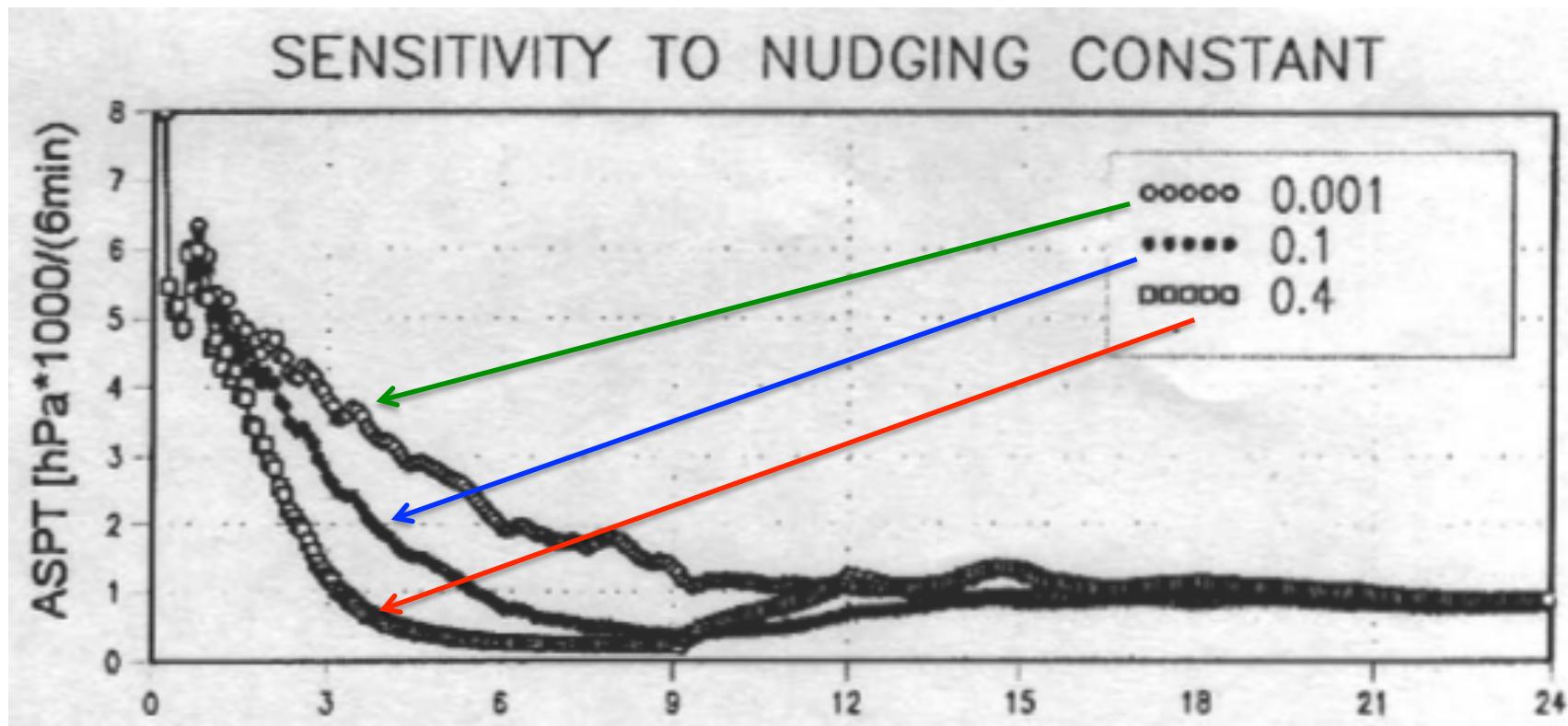
⁽²⁾ Instituto Nacional de Pesquisas Espaciais, MCT, São José dos Campos, Brasil

⁽³⁾ Centro de Ciencias de la Atmósfera, UNAM

Av. Circuito Exterior s/n, Ciudad Universitaria 04510, México D.F., México

caetano@servidor.unam.mx

Relaxação newtoniana (nudging)



Relaxação newtoniana (nudging)

Inverse Problems
in Science & Engineering

Volume 21 Numbers 3-4 April-June 2013

ISSN: 1741-5977

CONTENTS

(Number 3 includes a Special Session)
Special Session:
 Special Issue on Inverse problems in geosciences
 Haroldo F. de Campos Velho, Valéria C. Barbosa and Steven Cocke 355

A multi-objective hierarchical calibration procedure for land surface/ecosystem models
 Christopher R. Wicks, Steven H. Cook and Carla Cristina Souza Gomes 357

Estimating and reducing uncertainty in parameterizations in data assimilation
 Takeshi Miyoshi, Eugenia Kalnay and Hong Li 367

An evolution of the back and forth nudging for geophysical data assimilation: application to the ocean circulation estimation
 Didier Arribalzaga, Pierre Brousset and Jeanne Aloum 389

A coupled soil moisture initialization scheme for the FSU/COAPS climate model
 Steven Cocke, Marie Boisserie and Dong-Wook Shin 420

Calibrating an hydrological model by an evolutionary strategy for multi-objective optimization
 Andréia da S. Araújo, Haroldo F. de Campos Velho, Edmar F.P. Luz, Saúlo R. Freitas 438

Precipitation estimation to determine the precipitation field on South America
 Ariane F. da Senna, Haroldo F. de Campos Velho, Edmar F.P. Luz, Saúlo R. Freitas, Georg Gröll and Manoel A. Gao 451

Review Articles:

Numerical solution of a 2D inverse heat conduction problem
 Zhi Qian and Xiali Feng 467

Simultaneous determination of a time-dependent heat source and the initial temperature in an inverse heat conduction problem
 Jin Wen, Masahiro Yamamoto and Ting Wei 485

(Continued on inside back cover)

Taylor & Francis
Taylor & Francis Group

MEX
Printed in Mexico
y 100%
FSC® C004639

Volume 21 Numbers 3-4 April-June 2013

INCLUDES THE SPECIAL SECTION
Inverse Problems in Geosciences

Special Issue Guest Editors: Haroldo F. de Campos Velho, Valéria C. Barbosa and Steven Cocke

T=? and Q=?

Founded and edited by
 George S. Dulikravich

Taylor & Francis
Taylor & Francis Group

Steven Cocke, Marie Boisserie, and Dong-Wook Shin (2013): A coupled soil moisture initialization scheme for the FSU/COAPS climate model. **Inv. Prob. Sic. Eng.**, **21**(3), 420-437.

In this journal issue, there is a Special Session on Inverse Problems in Geosciences: Special issue Guest Editors: Haroldo F. De Campos Velho, Valéria C. Barbosa, and Steven Cocke.

Relaxação newtoniana (nudging)



Available online at www.sciencedirect.com

ScienceDirect

[Advances in Space Research 54 \(2014\) 22–36](http://Advances in Space Research 54 (2014) 22–36)

**ADVANCES IN
SPACE
RESEARCH**
(a COSPAR publication)
www.elsevier.com/locate/asr

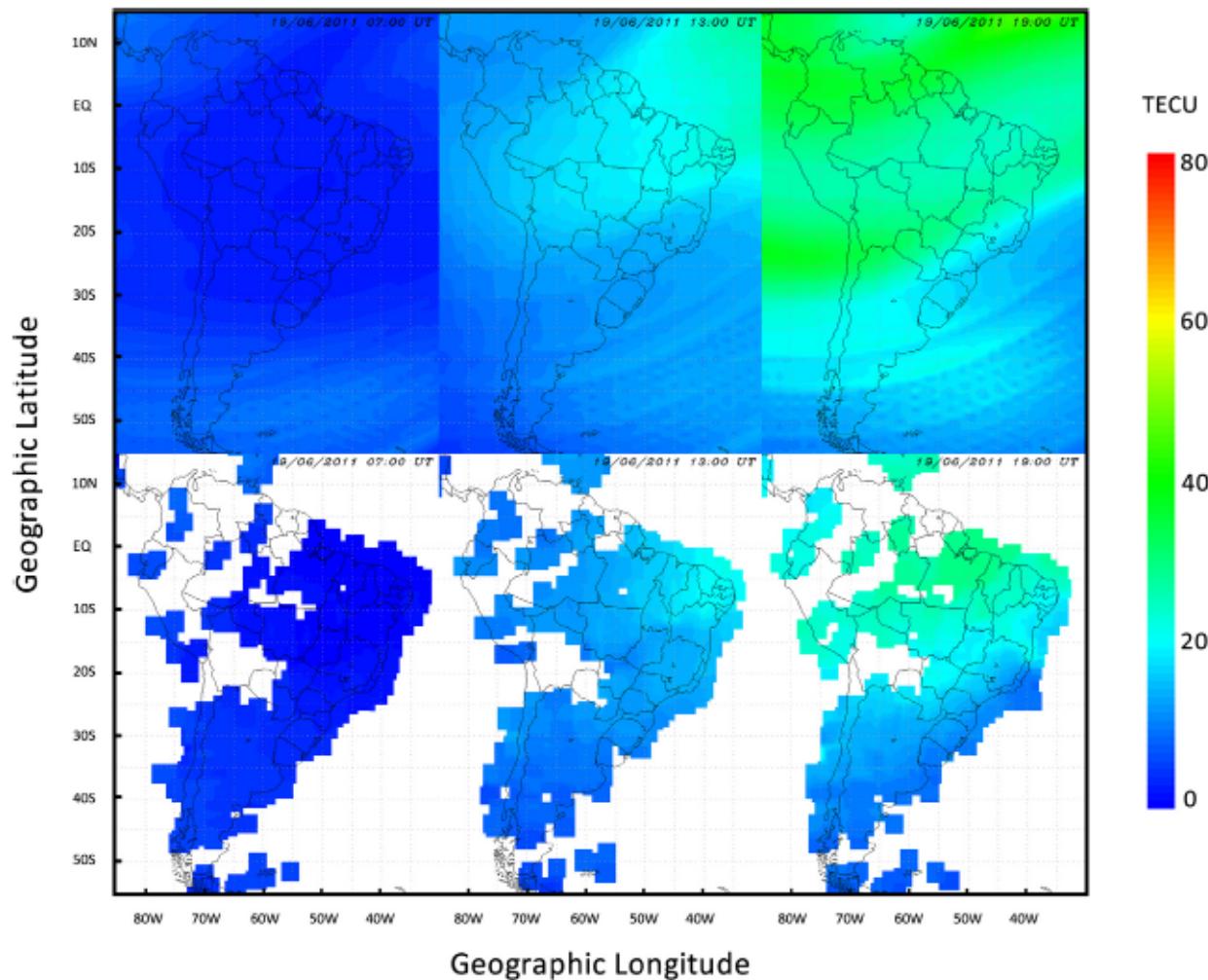
First results of operational ionospheric dynamics prediction
for the Brazilian Space Weather program

Adriano Petry^{a,*}, Jonas Rodrigues de Souza^{b,1}, Haroldo Fraga de Campos Velho^{c,2},
André Grahl Pereira^{d,3}, Graham John Bailey^e



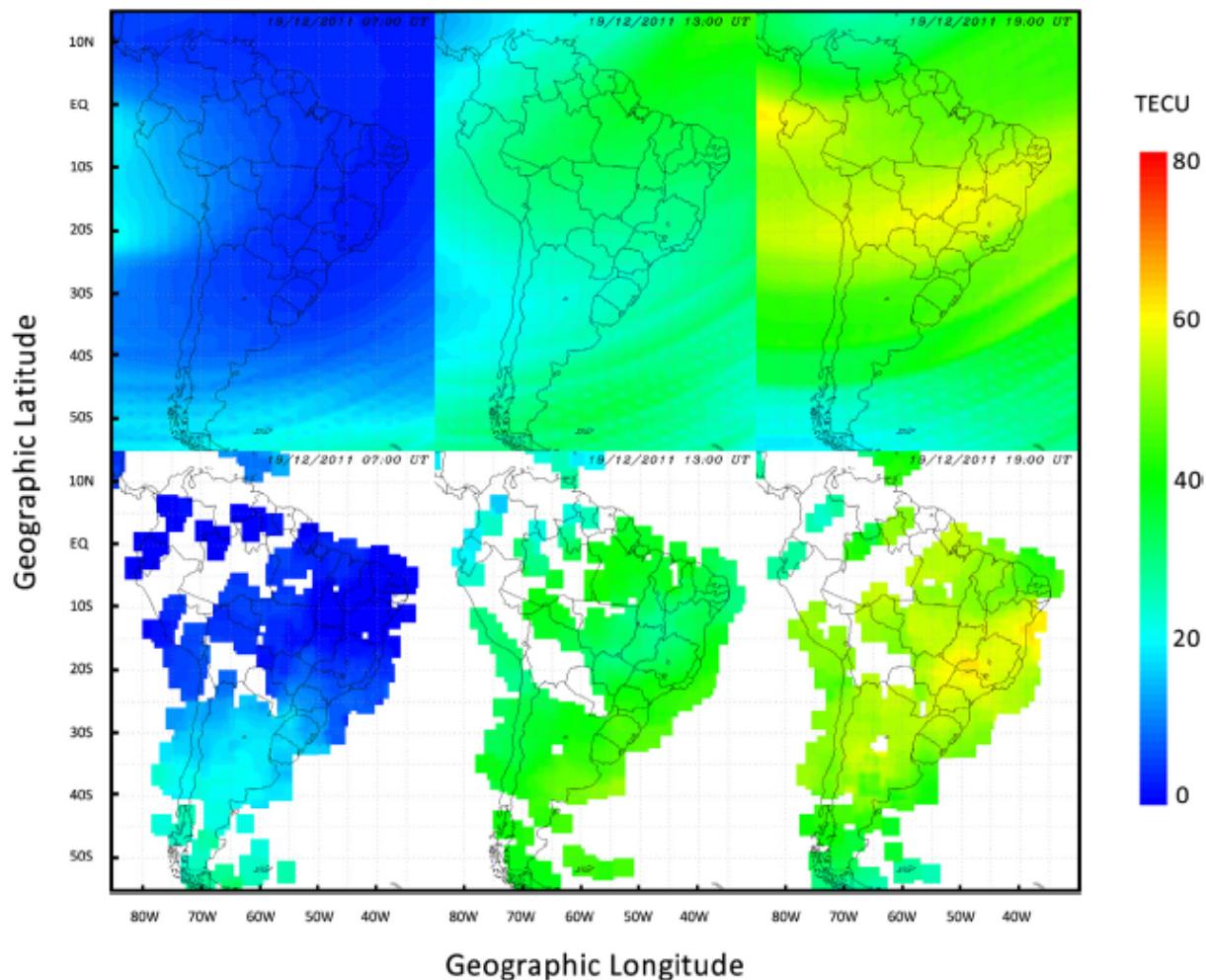
SUPIM: Space weather prediction

- 7, 13, 19 UT: June 19th, 2011



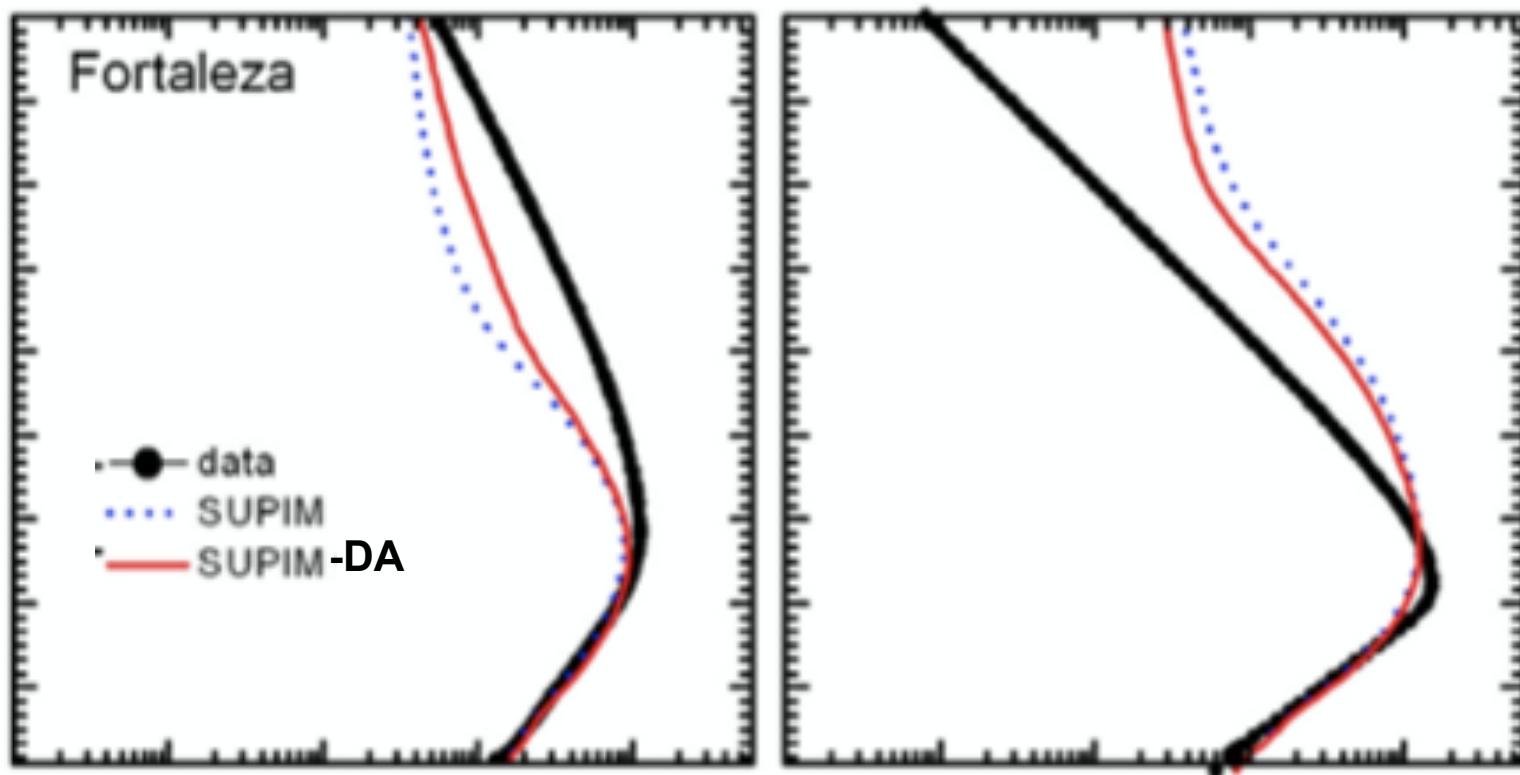
SUPIM: Space weather prediction

- 7, 13, 19 UT: December 19th, 2011



SUPIM: Space weather prediction

- Assimilation: Data from Fortaleza-sensor



Correções sucessivas

- Assimilação de dados por métodos tradicionais, pode ser vista com um processo em 2 etapas:
- Previsão (*forecasting*):

$$\mathbf{x}_n^f = F(\mathbf{x}_{n-1}, t)$$

- Análise (*analysis*):

$$\mathbf{x}_n^a = \mathbf{x}_n^f + \mathbf{W} \left[\mathbf{x}_n^o - H(\mathbf{x}_n^f) \right]$$

Correções sucessivas

- A seguinte notação é empregada:

$\mathbf{x}_n \equiv \mathbf{x}(t_n)$: vetor de estado no instante t_n

\mathbf{x}_n^f : previsão para o vetor de estado

\mathbf{x}_n^o : vetor de estado observado (medido)

$H(\cdot)$: operador de observação

\mathbf{W} : matriz de ponderação

Correções sucessivas

Matriz de ponderação

- Lembrando da equação:

$$\mathbf{f}_a = \left[\mathbf{W}_b^{-1} + \mathbf{W}_o^{-1} \right]^{-1} \left[\mathbf{W}_b^{-1} \mathbf{f}_b + \mathbf{W}_o^{-1} \mathbf{f}_o \right]$$

$$\mathbf{f}_a - \mathbf{f}_b = \mathbf{W}_b \left[\mathbf{W}_b^{-1} + \mathbf{W}_o^{-1} \right]^{-1} \left[\mathbf{f}_o - \mathbf{f}_b \right]$$

- Há várias alternativas para o cálculo da matriz de ponderação

Correções sucessivas

Matriz de ponderação

$$\mathbf{f}_a - \mathbf{f}_b = \mathbf{W}_b [\mathbf{W}_b + \mathbf{W}_o]^{-1} [\mathbf{f}_o - \mathbf{f}_b]$$

- Alternativa-1:

$$\mathbf{W} = \mathbf{W}_b [\mathbf{W}_b + \mathbf{W}_o]^{-1}$$

- Alternativa-2:

$$\mathbf{W} = \alpha \mathbf{I}$$

Correções sucessivas

- Historicamente a análise foi obtida por

$$\mathbf{x}_n^a(\mathbf{r}_i) = \frac{E_b^{-2} \mathbf{x}_n^f(\mathbf{r}_i) + E_o^{-2}(k) w(\mathbf{r}_k - \mathbf{r}_i) \left[\mathbf{x}_n^f(\mathbf{r}_i) + \mathbf{x}_n^o(\mathbf{r}_i) - \mathbf{x}_n^f(\mathbf{r}_i) \right]}{E_b^{-2} + E_o^{-2}(k) w(\mathbf{r}_k - \mathbf{r}_i)}$$

$$E_{b,o}^2 = \langle \varepsilon_{b,o}^2(\mathbf{r}) \rangle \text{ - assumido independente da posição}$$

$$w(\mathbf{r}_k - \mathbf{r}_i) = \begin{cases} 1 & \text{para } \mathbf{r}_k = \mathbf{r}_i \\ 0 & \text{para } (\mathbf{r}_k - \mathbf{r}_i) \rightarrow \infty \end{cases}$$

Correções sucessivas

- A equação da análise pode ser escrita como:

$$\mathbf{x}_n^a(\mathbf{r}_i) - \mathbf{x}_n^b(\mathbf{r}_i) = \frac{E_o^{-2}(k)w(\mathbf{r}_k - \mathbf{r}_i) \left[\mathbf{x}_n^o(\mathbf{r}_k) - \mathbf{x}_n^f(\mathbf{r}_k) \right]}{E_b^{-2} + E_o^{-2}(k)w(\mathbf{r}_k - \mathbf{r}_i)}$$

$$\mathbf{x}_n^a(\mathbf{r}_i) - \mathbf{x}_n^b(\mathbf{r}_i) = \mathbf{W}_{ik} \left[\mathbf{x}_n^o(\mathbf{r}_k) - \mathbf{x}_n^f(\mathbf{r}_k) \right]$$

$$\mathbf{W}_{ki} = \frac{E_b^2 w(\mathbf{r}_k - \mathbf{r}_i)}{E_b^2 w(\mathbf{r}_k - \mathbf{r}_i) + E_o^2(k)} : \text{ peso "a posteriori"}$$

Correções sucessivas

- Tem surgido na literatura várias formas para a função peso “a priori” $w(\mathbf{r}_k - \mathbf{r}_i)$:

$$w(\mathbf{r}_k - \mathbf{r}_i) = w(\hat{\mathbf{r}}) = \begin{cases} \left(R^2 - \hat{\mathbf{r}}^2 \right) / \left(R^2 + \hat{\mathbf{r}}^2 \right) & \hat{\mathbf{r}} \leq R \\ 0 & \hat{\mathbf{r}} > R \end{cases}$$

$$w(\mathbf{r}_k - \mathbf{r}_i) = w(\hat{\mathbf{r}}) = \exp\left(\frac{-\hat{\mathbf{r}}^2}{2R^2}\right)$$

Correções sucessivas

- Ciclo de iteração: $\left(\varepsilon_o^2 \equiv E_o^2 / E_b^2 \right)$

$$\mathbf{x}_n^{a,1}(\mathbf{r}_i) = \mathbf{x}_n^b(\mathbf{r}_i) + \mathbf{W}_i^T \left[\mathbf{x}_n^o(\mathbf{r}_i) - \mathbf{x}_n^f(\mathbf{r}_i) \right]$$

$$W_{ki} = \frac{w(\mathbf{r}_k - \mathbf{r}_i)}{\sum_{k=1}^{K_i} w(\mathbf{r}_k - \mathbf{r}_i) + \varepsilon_o^2}$$

$$\mathbf{x}_n^{a,m+1}(\mathbf{r}_i) = \mathbf{x}_n^{a,m}(\mathbf{r}_i) + \mathbf{W}_i^T \left[\mathbf{x}_n^o(\mathbf{r}_i) - \mathbf{x}_n^{a,m}(\mathbf{r}_i) \right]$$

Interpolação estatística

- Estimação de variância mínima:
- Seja s o valor verdadeiro e ε_n o erro associado as medidas s_n ($n=1,2, \dots, N$). O erro é suposto ser não tendencioso e não correlacionado:

$$\langle \varepsilon_n \rangle = 0 \quad \langle \varepsilon_m \varepsilon_n \rangle = 0 \quad \text{para } m \neq n$$

$$\sigma_n^2 = \langle \varepsilon_n^2 \rangle$$

Interpolação estatística

- Considerando a seguinte estimativa linear

$$s_e = \sum_n c_n s_n \quad \text{com } c_n \geq 0$$

- O erro da estimativa é ε_e e a tendência/viés (bias):

$$\begin{aligned} \langle \varepsilon_e \rangle &= \langle s_e \rangle - \langle s \rangle + \left[\left\langle \sum_n c_n s \right\rangle - \left\langle \sum_n c_n s \right\rangle \right] = \sum_n c_n \langle s_n - s \rangle - \langle s \rangle + \left\langle \sum_n c_n s \right\rangle \\ &= \sum_n c_n \langle \varepsilon_n \rangle - \langle s \rangle \left[1 - \sum_n c_n \right] \end{aligned}$$

Interpolação estatística

- Como $\langle \varepsilon_n \rangle = 0$ segue o resultado:

$$\langle s_e \rangle = 0 \quad \text{se : } \sum_n c_n = 1$$

- A estimativa é *linear não-tendenciosa* SE: $\sum_n c_n = 1$

- Definindo a variância

$$\langle \varepsilon_e^2 \rangle = \langle (s_e - s)^2 \rangle = \left\langle \left(\sum_n c_n s_n - \sum_n c_n s \right)^2 \right\rangle = \left\langle \left(\sum_n c_n \varepsilon_n \right)^2 \right\rangle \leq \sigma_{\max}^2 \sum_n c_n^2$$

Interpolação estatística

- Como $c_n \geq 0$ é claro que:

$$c_n \leq 1 \Rightarrow c_n^2 \leq c_n \quad \text{e} \quad \sum_n c_n^2 \leq 1$$

- Então, $\langle \varepsilon_e^2 \rangle \leq \sigma_{\max}^2$
- Estamos interessados estimar s_a que minimiza a variância:

$$\langle \varepsilon_e^2 \rangle = \sum_n c_n^2 \sigma_n^2 \quad \text{sujeito a restrição} \quad \sum_n c_n = 1$$

Interpolação estatística

- A otimização com restrições pode ser re-escrita:

$$J = \langle \varepsilon_e^2 \rangle + \lambda \left(1 - \sum_n c_n \right) = \sum_n c_n^2 \sigma_n^2 + \lambda \left(1 - \sum_n c_n \right)$$

- Então,
- Ou: $\frac{\partial J}{\partial c_n} = 2c_n \sigma_n^2 - \lambda = 0$
- Ou: $c_n = \frac{\lambda}{2\sigma_n^2} \Rightarrow c_n = \frac{\sigma_n^{-2}}{\sum_n \sigma_n^{-2}}$

Interpolação estatística

- Definindo $\sigma_{\min}^2 \equiv \min\{\sigma_1^2, \dots, \sigma_N^2\}$, então:

$$\frac{1}{\langle \varepsilon_a^2 \rangle} = \sum_n \sigma_n^{-2}$$

- Ou (estimativa tem erro menor do que a melhor das medidas):

$$\frac{1}{\langle \varepsilon_a^2 \rangle} \geq \sigma_n^{-2}; \quad \text{e: } \langle \varepsilon_a^2 \rangle \leq \sigma_{\min}^2$$

Interpolação estatística

- O método prevê uma estimativa de variância mínima:

$$\mathbf{x}_n^o = \mathbf{x}_n^f + \mathbf{W} \left[\mathbf{x}_n^o - H(\mathbf{x}_n^f) \right]$$

$$\mathbf{W} = W_b H^T \left[H W_b H^T + W_o \right]^{-1}$$

Análise de correção

- É inspirado na interpolação estatística, combinado com o método das correções sucessivas:

$$\mathbf{x}_n^{o,m+1} = \mathbf{x}_n^{f,m} + \mathbf{KQ} \left[\mathbf{x}_n^{o,m} - H(\mathbf{x}_n^{f,m}) \right]$$

$$\mathbf{x}_n^{o,m+1} = \mathbf{x}_n^{o,m} - \mathbf{Q} \left[\mathbf{x}_n^{o,m} - H(\mathbf{x}_n^{f,m}) \right]$$

$$K = W_b H^T R^{-1}$$

$$Q = (HW + I)^{-1}$$

Estimativa por mínimos quadrados

- Ou seja, com medidas com erros diversos, deseja-se encontrar a estimativa de mínimos quadrados para:

$$\mathbf{W}\mathbf{H}\mathbf{x} = \mathbf{W}\mathbf{y}$$

- E a estimativa será

$$\mathbf{x}^a = (\mathbf{H}^T \mathbf{C} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C} \mathbf{y} \quad - \text{ ou : } \mathbf{x}^a = \mathbf{L} \mathbf{y}$$

- Escolha da matriz \mathbf{C} como: $\mathbf{C} = \mathbf{R}^{-1}$

Estimativa por mínimos quadrados

- Procura-se uma matriz C tal que

$$\langle \boldsymbol{\varepsilon} \rangle = \langle \mathbf{y} - \mathbf{Hx} \rangle = 0$$

- O objetivo é minimizar a função objetivo:

$$\begin{aligned}\langle \mathbf{x} - \mathbf{x}^a \rangle &= \langle \mathbf{x} - \mathbf{Ly} \rangle = \langle \mathbf{x} - \mathbf{L}(\mathbf{Hx} + \boldsymbol{\varepsilon}) \rangle = \langle \mathbf{x} - \mathbf{LHx} - \mathbf{L}\boldsymbol{\varepsilon} \rangle \\ &= \langle \mathbf{x} - \mathbf{LHx} - \mathbf{L}\boldsymbol{\varepsilon} \rangle = \langle (\mathbf{I} - \mathbf{LH})\mathbf{x} \rangle = 0\end{aligned}$$

- Ou seja, \mathbf{x}^a é não tendencioso se: $\mathbf{LH} = \mathbf{I}$

Estimativa por mínimos quadrados

- Definindo a nova matriz:

$$\mathbf{L}_0 = \left(\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1}$$

- Também satisfaz: $\mathbf{L}_0 \mathbf{H} = \mathbf{I}$
- Assim, o melhor estimador linear não tendencioso (BLUE: *Best Linear Unbiased Estimation*) é tal que $\mathbf{C} = \mathbf{R}^{-1}$.

Estimativa por mínimos quadrados

- Pretende-se estimar $\mathbf{x}^{a,1}$ a partir da estimativa $\mathbf{x}^{a,0}$ sendo a matriz de covariância dada por:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_1 \end{bmatrix}$$

- E a matriz $\mathbf{H}^T \mathbf{C} \mathbf{H}$ no cálculo de $\mathbf{x}^{a,1}$ será dada por:

$$\mathbf{P}_1^{-1} = \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{R}_0 & 0 \\ 0 & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \end{bmatrix}$$

$$\mathbf{P}_1^{-1} = \mathbf{P}_0^{-1} + (\mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{H}_1)^{-1}$$

Estimativa por mínimos quadrados

- A estimativa $\mathbf{x}^{a,1}$ não é baseada somente na medida \mathbf{y}^1 e é o melhor estimador do sistema:

$$\begin{cases} \mathbf{H}_0\mathbf{x} = \mathbf{y}_0 \\ \mathbf{H}_1\mathbf{x} = \mathbf{y}_1 \end{cases}$$

- A equação normal é: $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{Hx} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$ e $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} = \mathbf{P}^{-1}$, então a estimativa ótima é

$$\mathbf{x}^a = \mathbf{P}_1 \left(\mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{y}_0 + \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{y}_1 \right)^{-1}$$

- A nova medida, significa maior informação. Isto faz com que a matriz \mathbf{P} diminua.

Estimativa por mínimos quadrados

- Rearranjando a última expressão (considerando as medidas tomadas em $m+1$):

$$\mathbf{x}^{\alpha,m+1} = \mathbf{x}^{\alpha,m} + \mathbf{G}_m \left(\mathbf{Y}_m - \mathbf{H}_m \mathbf{x}^{\alpha,m} \right)$$

$$\mathbf{G}_m = \mathbf{P}_m \mathbf{H}_m^T \mathbf{R}_m^{-1}$$

Filtro de Kalman

- Dada a equação do modelo linear

$$\mathbf{Hx} = \mathbf{y} + \boldsymbol{\varepsilon}$$

- Onde $\boldsymbol{\varepsilon}$ é um vetor gaussiano aleatório com média zero e matriz de covariância \mathbf{R} : $\boldsymbol{\varepsilon} \sim N(0, \mathbf{R})$.
- O melhor estimador \mathbf{x}^a linear não tendencioso de \mathbf{x}^a é:

$$\mathbf{x}^a = \mathbf{P} \mathbf{H}^T \mathbf{R}_0^{-1} \mathbf{y}$$

Filtro de Kalman

- Onde a matriz \mathbf{P} é a covariância da estimativa

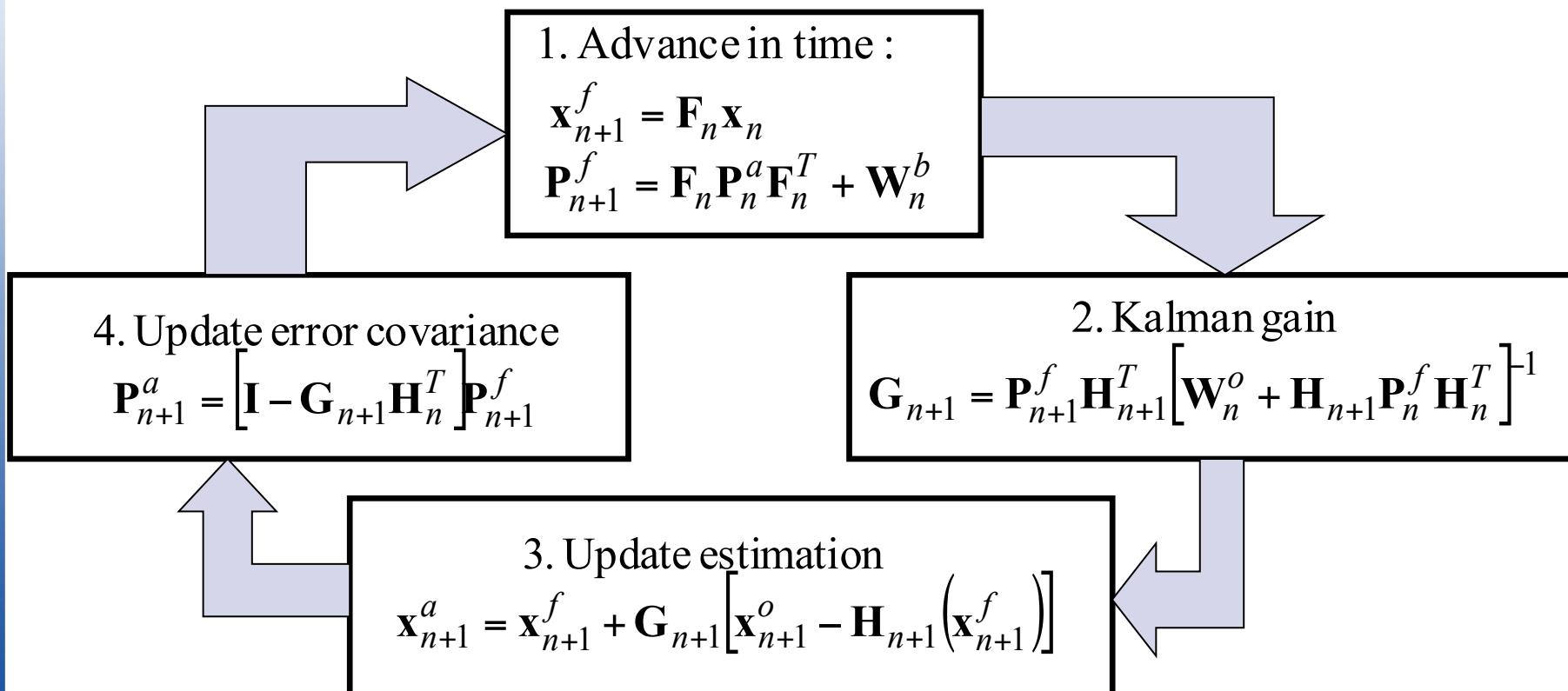
$$\mathbf{P} = \left\langle (\mathbf{x} - \mathbf{x}^a)(\mathbf{x} - \mathbf{x}^a)^T \right\rangle = (\mathbf{H}\mathbf{R}^{-1}\mathbf{H})^{-1}$$

- O filtro de Kalman calcula o melhor estimador linear não tendencioso no tempo n , dadas as medidas y_0, y_1, \dots, y_k ; e o filtro também calcula a matriz de covariância do erro.

Filtro de Kalman

$$\dot{\mathbf{x}}_{n+1} = F[\mathbf{x}_n, t_n] \approx \mathbf{F}_n + \frac{\partial F}{\partial \mathbf{x}} \Big|_{t=t_n} \mathbf{x}_n + O(\Delta t^2) \approx \mathbf{E}_n \mathbf{x}_n$$

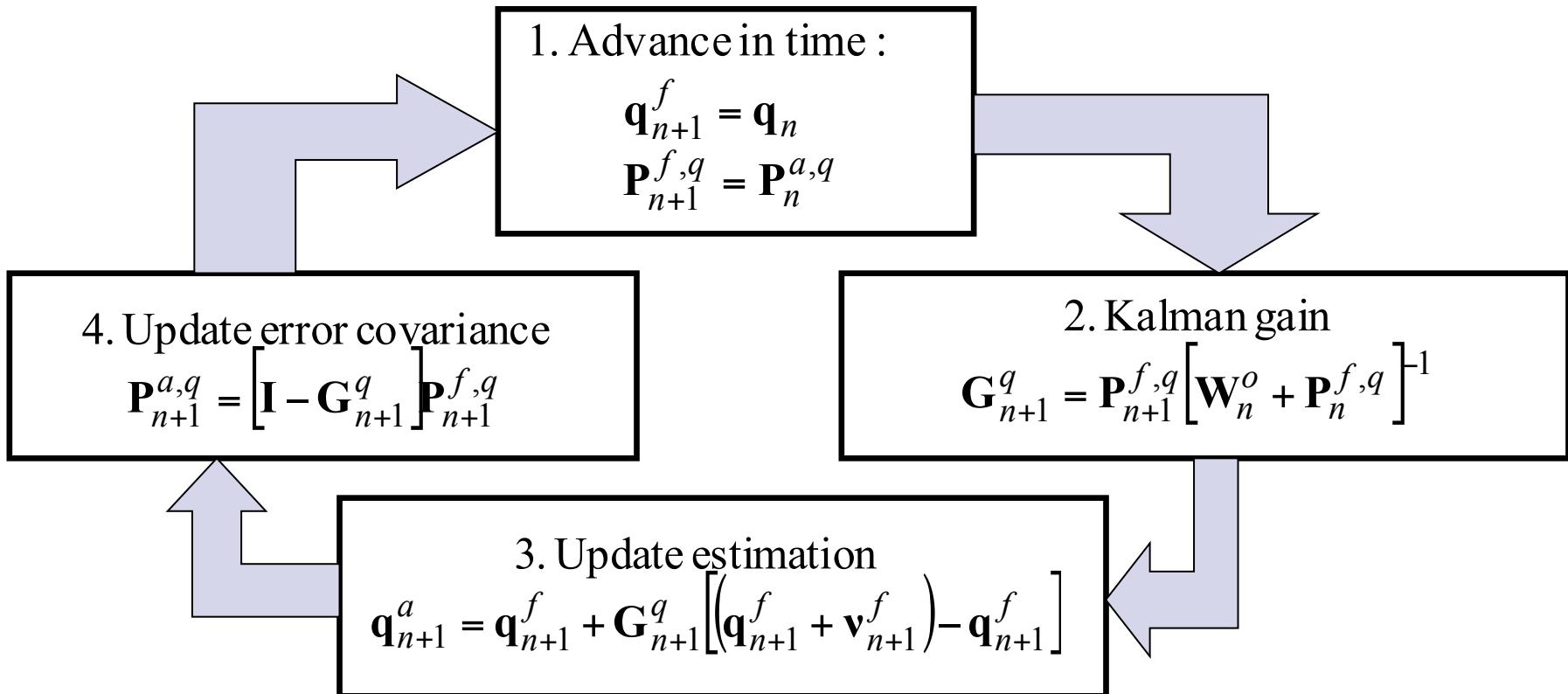
Three versions: Linear, Extended, Adaptive (next slide)



A.G. Nowosad, A. Rios Neto, H.F. de Campos Velho: Data Assimilation Using an Adaptative Kalman Filter and Laplace Transform, Hybrid Methods in Engineering, 2(3), 291-310, 2000.

Filtro de Kalman

Three versions: Linear, Extended, **Adaptive**



A.G. Nowosad, A. Rios Neto, H.F. de Campos Velho: Data Assimilation Using an Adaptive Kalman Filter and Laplace Transform, Hybrid Methods in Engineering, 2(3), 291-310, 2000.

Variational methods: 3D and 4D

Objective (cost) function (adjoint equation will not discussed here):

$$J(\mathbf{x}) = \frac{1}{2} [\mathbf{x}_n^f - \mathbf{x}]^\top \mathbf{B}^{-1} [\mathbf{x}_n^f - \mathbf{x}] + \frac{1}{2} \sum_{n=1}^{N_a} [\mathbf{x}_n^o - \mathbf{H}_n \mathbf{x}]^\top \mathbf{R}^{-1} [\mathbf{x}_n^o - \mathbf{H}_n \mathbf{x}]$$

$N_a = 1$ for 3D-Var, and $N_a > 1$ for 4D-Var

Covariance matrices: they can be estimated by Fokker-Planck equation – see:

(K Belyaev, CAS Tanajura (2002): Appl. Math Model., 26(11):1019-1027)

B: background error covariance matrix;

R: observational error covariance matrix

Assimilação de Dados

- **Problema inverso:** determinar condição inicial.
- Teoria da regularização: minimizar o funcional abaixo:

$$J_\alpha(f) = \|A(f) - \theta^\delta\|_2^2 + \alpha \Omega[f]$$

$A(f)$: Solução do problema direto

“ f ”: Condição inicial (função) candidata

θ^δ : temperatura medida em vários pontos da barra

$\Omega[f]$: Operador de regularização

α : Parâmetro de regularização

Remarks on OI, 3D-Var, Kalman filter and 4D-Var

Remark-1: OI and 3D-Var could be equivalent, dealing with optimal least square gain.

Remark-2: extended KF and 4D-Var over a given time interval, perfect model, the 4D-Var analysis at the end of the time interval is equal to the Kalman filter analysis at the same time.

Details: [http://www.ecmwf.int/newsevents/training/rcourse_notes/
DATA_ASSIMILATION/ASSIM_CONCEPTS/index.html](http://www.ecmwf.int/newsevents/training/rcourse_notes/ DATA_ASSIMILATION/ASSIM_CONCEPTS/index.html)

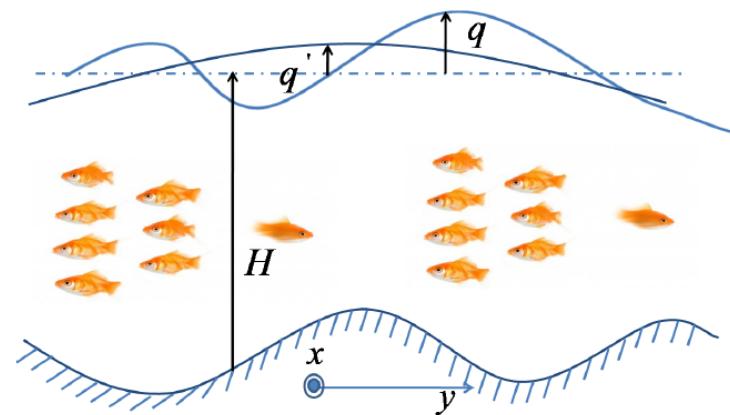
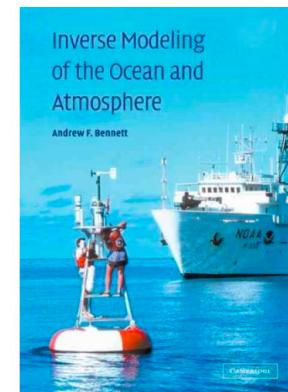
Kalman filter: application for DA

- Shallow water 2D for ocean circulation

$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$

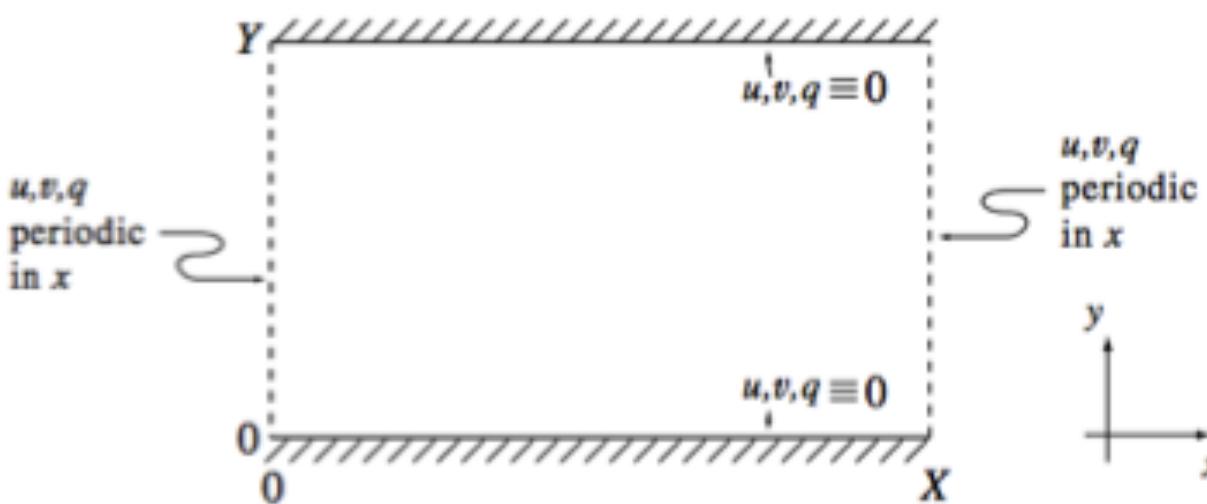


Shallow water 2D for ocean circulation

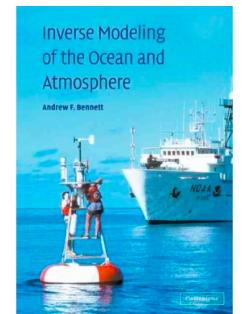
$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

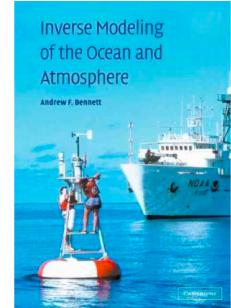
$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$



(BENNETT, 2002)



Shallow water 2D for ocean circulation



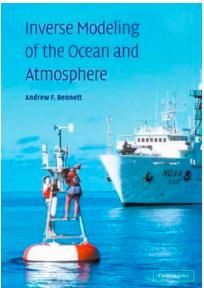
Parameter	Value	Parameter	Value
H	5000 m	r_u	$1.8 \times 10^4 \text{ s}^{-1}$
T	$1.8 \times 10^4 \text{ s}$	r_v	$1.8 \times 10^4 \text{ s}^{-1}$
g	9.806 m s^{-2}	r_q	$1.8 \times 10^4 \text{ s}^{-1}$
f	$1.0 \times 10^{-4} \text{ s}^{-1}$	ρ_a	1.275 kg m^{-3}
C_d	1.6×10^{-3}	ρ_w	$1.0 \times 10^3 \text{ kg m}^{-3}$

$$\frac{\partial u}{\partial t} - fv + g\frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g\frac{\partial q}{\partial y} + r_v v = F_v$$

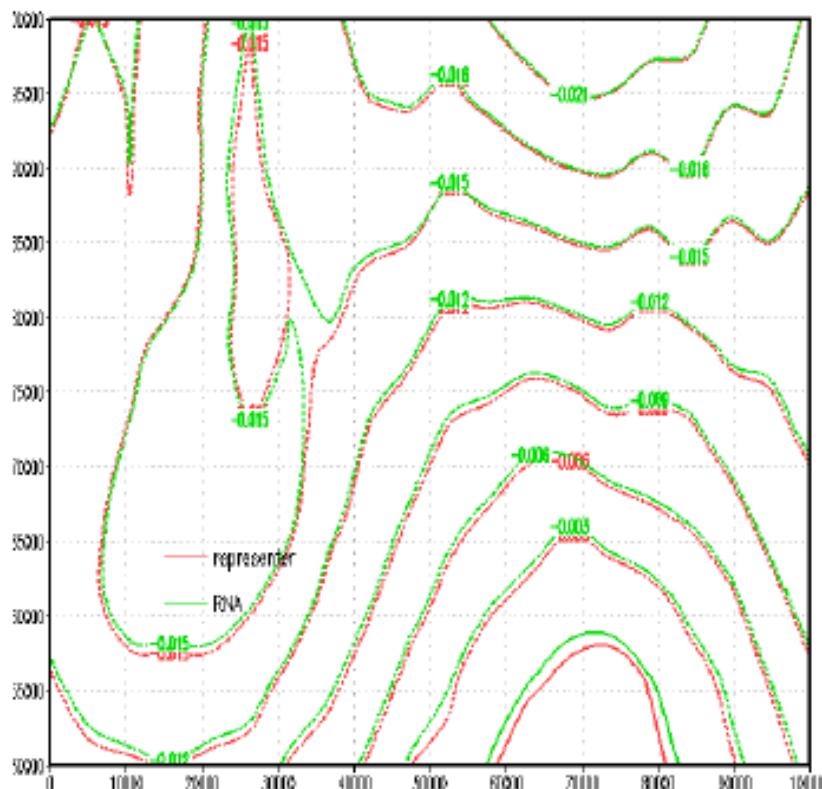
$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$

Shallow water 2D: Kalman filter x neural network

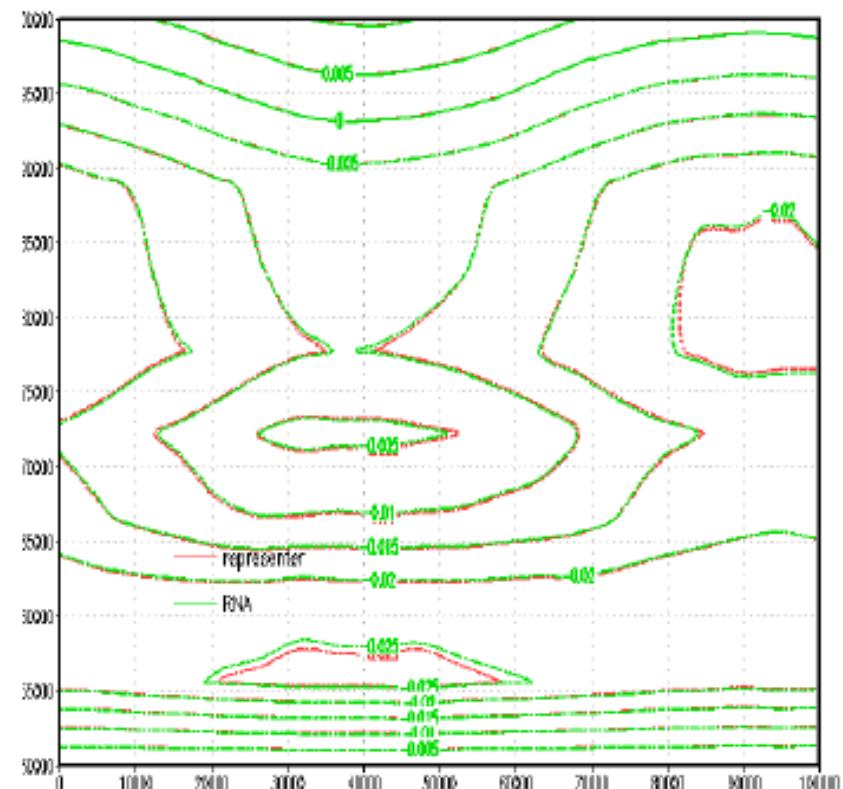


Green = Kalman filter / Red = neural network

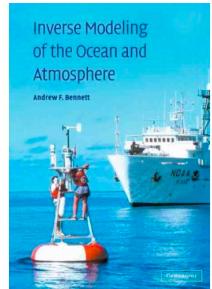
Zonal wind (u)



Meridionalwind (v)



Shallow water 2D: Kalman filter x neural network

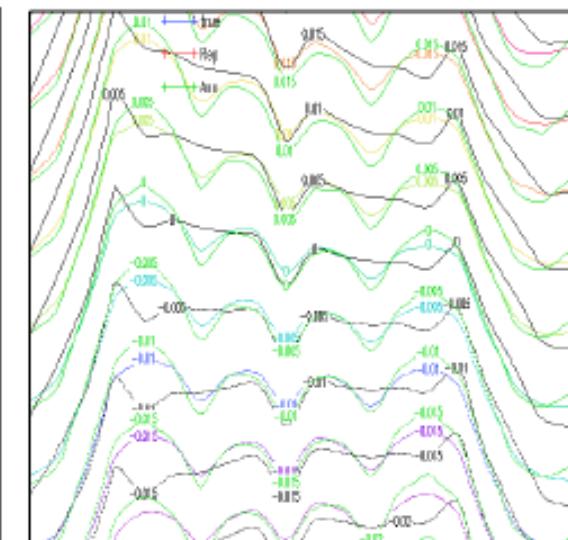
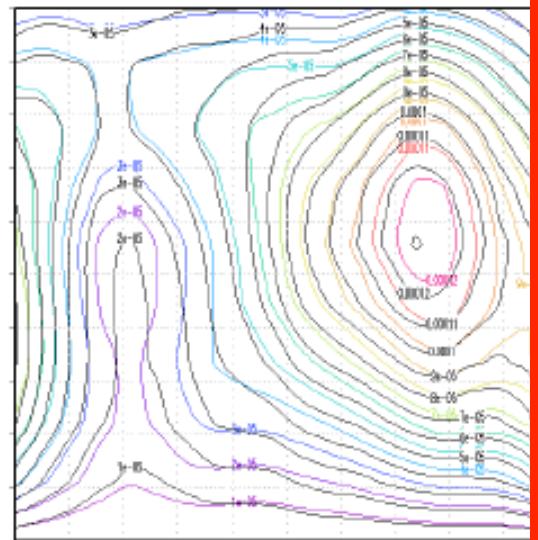
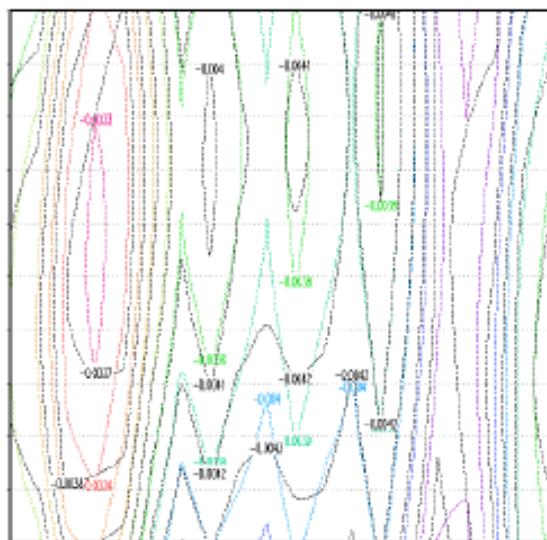


Data assimilation: Kalman filter and neural network

Zonal wind (u)

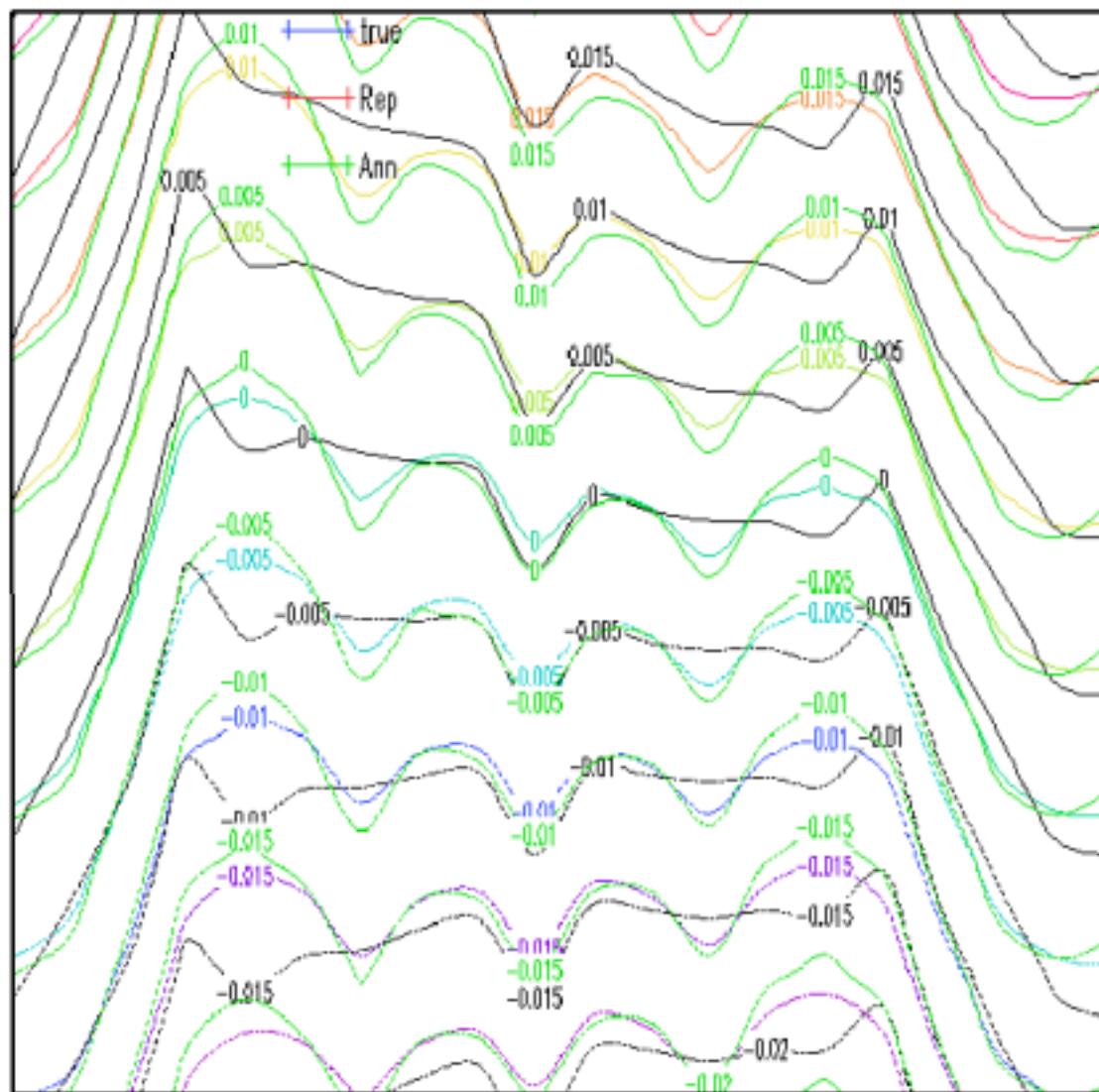
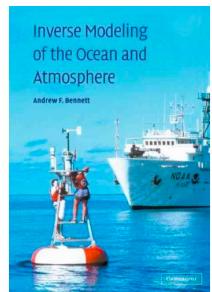
Meridionalwind (v)

height (q)



Shallow water 2D: Kalman filter x neural network

Kalman filter vs neural network: zoom for $q(x,y)$



Particle filter

- When does Kalman filter fail?
 - KF assumptions: linear (linearization) process, Gaussian statistics for modeling and measurements errors.
 - Sometimes the observations are not frequent enough.¹
 - The statistics could not be Gaussian.²
 - Therefore:
 1. Linearization will not work.
 2. Kalman solution (Gaussian statistics) can not be applied

Particle filter

- When does Kalman filter fail?
 - KF assumptions: linear (linearization) process, Gaussian statistics for modeling and measurements errors.
 - Sometimes the observations are not frequent enough.¹
 - The statistics could not be Gaussian.²
 - Therefore:
 1. Linearization will not work.
 2. Kalman solution (Gaussian statistics) can not be applied
- Solution: **4D-Var or Particle Filter!**

Methods for data assimilation

- Variational methods: 3D/4D-Var

Short Course on Data Assimilation:

Day 2 (July 12, 2022):

15:00-16:30 – Optimal interpolation and variational (3D/4D) methods
(Dr. Amos Lawless – University of Reading, UK))

State estimation: several goals

- **Prediction – Filtering – Fixed-lag Smoothing**
 - $\pi(x,z)$: conditional probability distribution
 - x : state variable
 - z : observations
 - **Prediction**: determination of $\pi(x_k, z_{1:k-1})$
 - **Filtering**: determination of $\pi(x_k, z_{1:k})$
 - **Fixed-lag Smoothing**: determination of $\pi(x_k, z_{1:k+p})$
 - where $p \geq 1$ is the fixed lag.

Bayesian filters

- **Bayesian strategy – filtering**

1. We know: $\pi(x_0, z_0) = \pi(x_0)$
2. Compute: $\pi(x_k, z_{1:k})$
 - a) Based on Bayes's theorem

$$P(A | B) = \frac{P(A \cap B)P(A)}{P(B)}$$

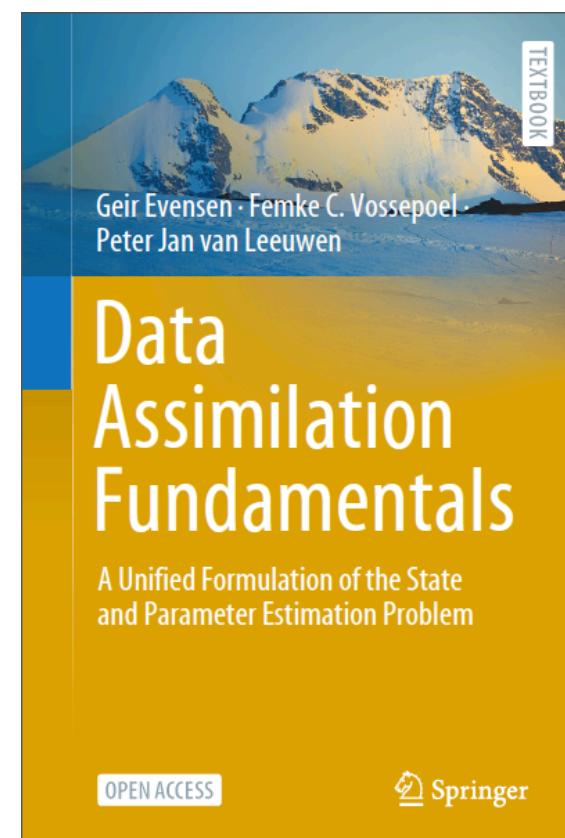
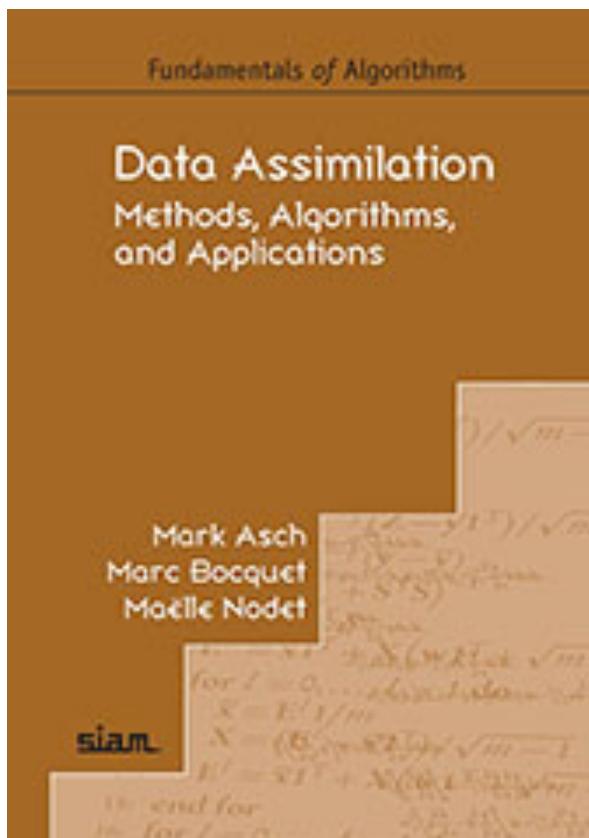
- b) Under Markovian process

$$\pi(x_k | x_{k-1}, \dots, x_1) = \pi(x_k | x_{k-1})$$

3. Bayesian filters: **prediction and update**

Bayesian filters

- Bayesian strategy – Data assimilation



Particle filter

■ Particle filter algorithm

1. Compute the initial ensemble: $\{w_{0|n-1}^{(i)}\}_{i=0}^M \sim p_{w_0}(w_0)$ (initial PDF : $N(0.,5)$)
2. Compute the weights: $q_n^{(i)} = p(y_n | w_{n|n-1}) = p_{\text{et}}[y_n - h(w_n, t_n)]$
 $p_{\text{et}}(z) = \exp(-z^2 / 4\pi)$
3. Normalize: $\hat{q}_n^{(i)} = q_n^{(i)} / \left(\sum_{i=1}^M q_n^{(i)} \right)$
4. Re-sampling (select particles): $P(w_n | Y_s) \approx \sum_{i=1}^M \hat{q}_n^{(i)} \delta(w_n - w_{n|s}^{(i)})$ $\sum_{i=1}^M \hat{q}_n^{(i)} = 1$
5. Compute new particles: $w_{n+1|n}^{(i)} = f(w_{n|n}^{(i)}, t_n) + \mu_n$

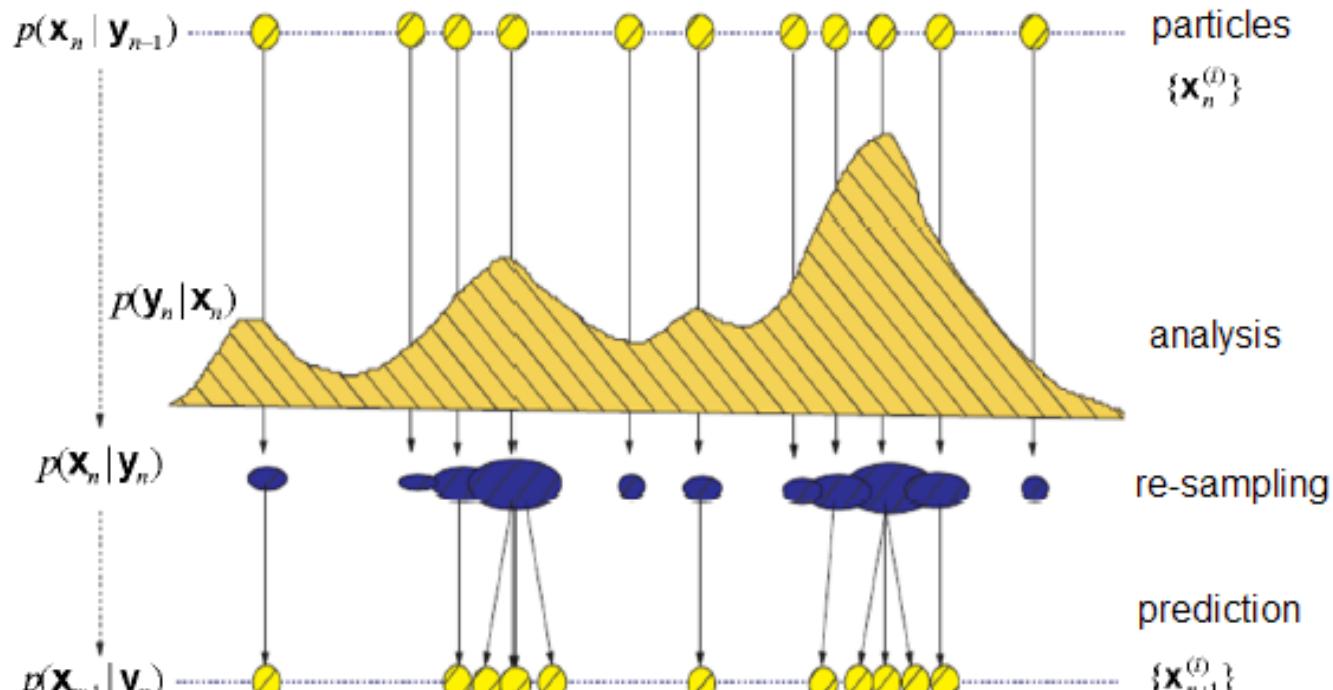
Particle filter

■ Particle filter algorithm

1. Compute the initial ensemble: $\{w_{0|n-1}^{(i)}\}_{i=0}^M \sim p_{w_0}(w_0)$ (initial PDF : $N(0.,5)$)
2. Compute the weights: $q_n^{(i)} = p(y_n | w_{n|n-1}) = p_{\text{et}}[y_n - h(w_n, t_n)]$
$$p_{\text{et}}(z) = \exp(-z^2 / 4\pi)$$
3. Normalize: $\hat{q}_n^{(i)} = q_n^{(i)} / \left(\sum_{i=1}^M q_n^{(i)} \right)$
4. Re-sampling (select particles): $P(w_n | Y_s) \approx \sum_{i=1}^M \hat{q}_n^{(i)} \delta(w_n - w_{n|s}^{(i)})$ $\sum_{i=1}^M \hat{q}_n^{(i)} = 1$
5. Compute new particles: $w_{n+1|n}^{(i)} = f(w_{n|n}^{(i)}, t_n) + \mu_n$

Particle filter

- Particle filter algorithm: resampling



(adapted from Chen (2000))

Methods for data assimilation

- Particle filter

Short Course on Data Assimilation:

Day 3 (July 13, 2022):

15:00-16:30 – On particle filters and particle flow filters and smoother:
towards fully nonlinear data assimilation

(Prof. Pete van Leeuwen – Colorado State University, USA)

Particle filter

- Kernel for the particle filter algorithm

Bayes theorem:

$$\begin{aligned} p(w_n | Y_n) &= p(w_n | y_n, Y_{n-1}) = \frac{p(y_n | w_n)p(w_n | Y_{n-1})}{p(y_n | Y_{n-1})} \\ &\propto p(y_n | w_n)p(w_n | Y_{n-1}) \end{aligned}$$

$$\underbrace{p(w_n | Y_{n-1})}_{\text{a posteriori}_{(w_n)}} \propto \underbrace{p(y_n | w_n)}_{\text{likelihood}_{(w_n)}} \underbrace{p(w_n | Y_{n-1})}_{\text{a priori}_{(w_n)}}$$



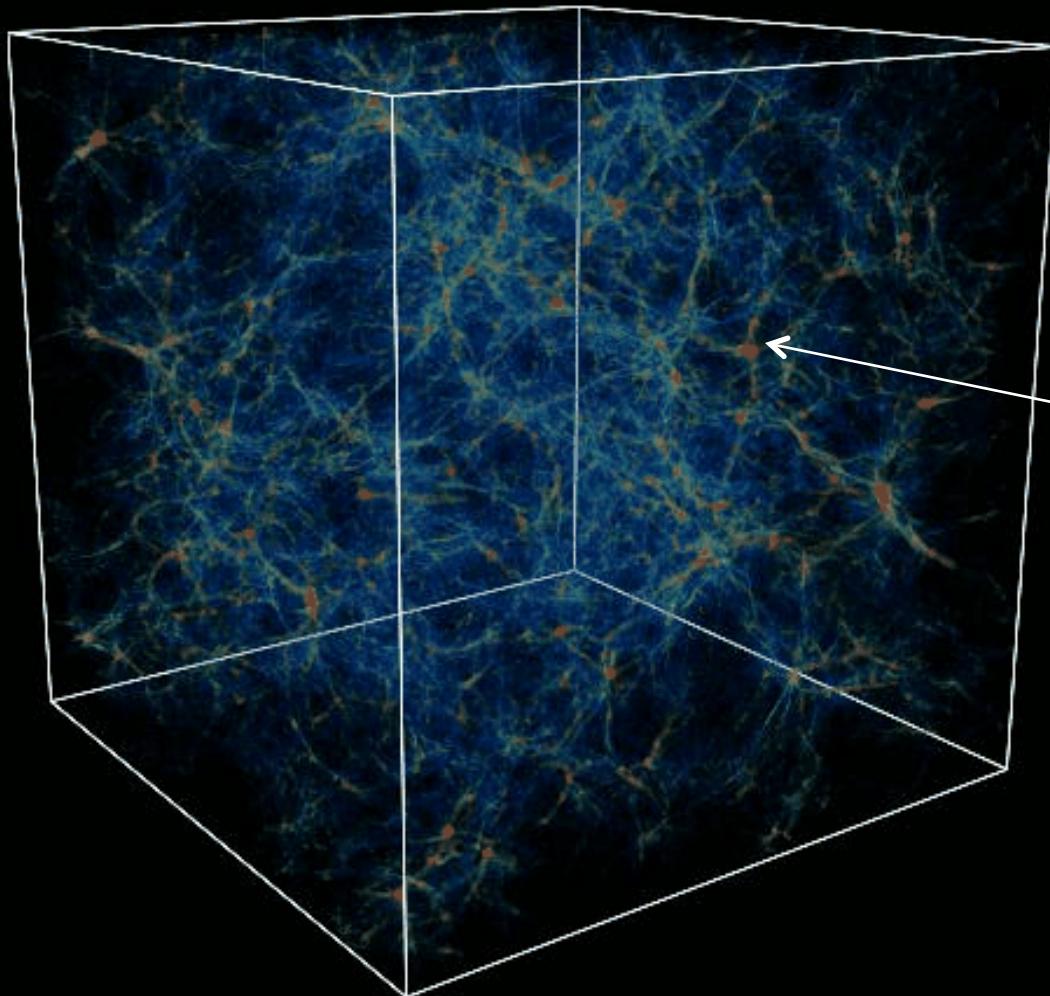
Beyond particle filter

Everything is perfect with particle filter?

Almost!

1. In the distribution space there are (at least) two attractors for stable distributions: Gaussian, and Lévy α -stable.
2. **Gaussian (Normal)**: central limit theorem.
Lévy α -stable: Lévy-Gnedenko central limit theorem

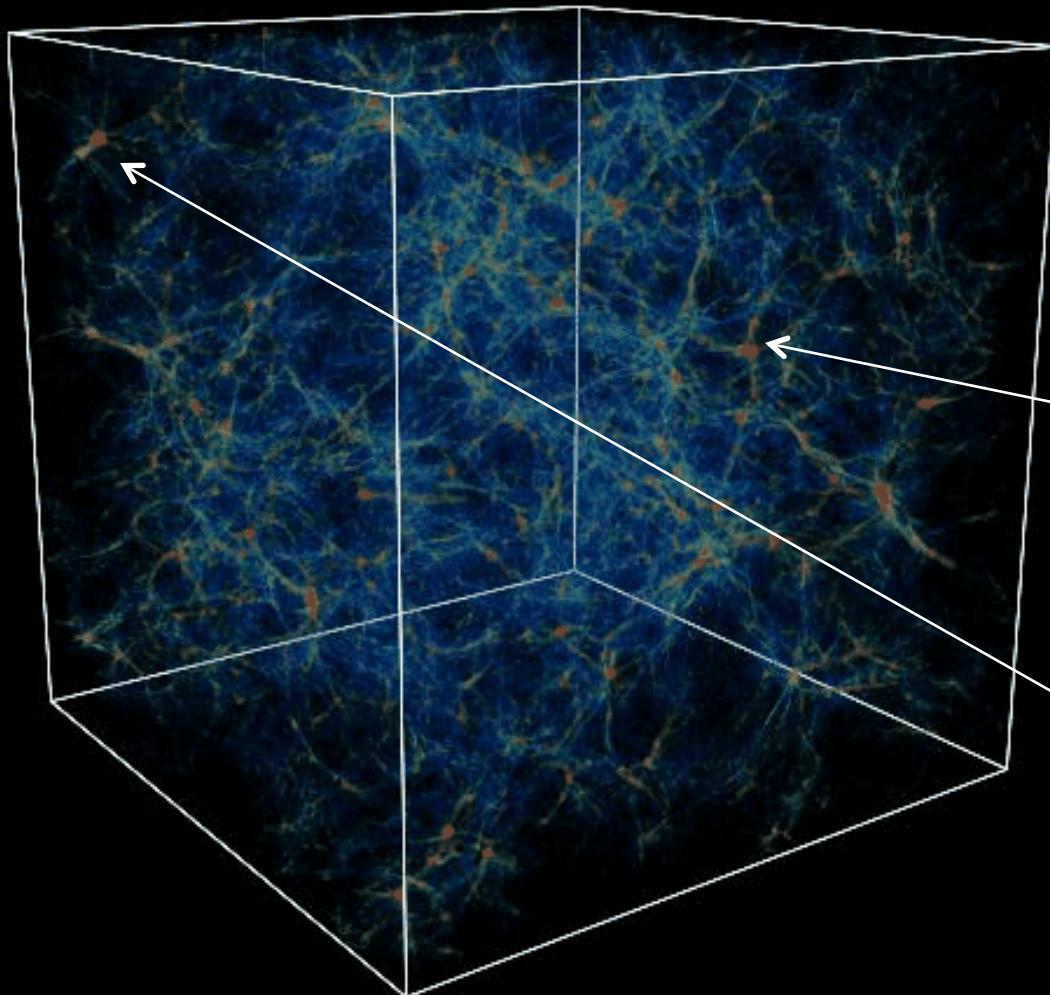
Everything is perfect with particle filter?



Where are the
attractors?

Gaussian?

Everything is perfect with particle filter?



Where are the
attractors?

Gaussian?

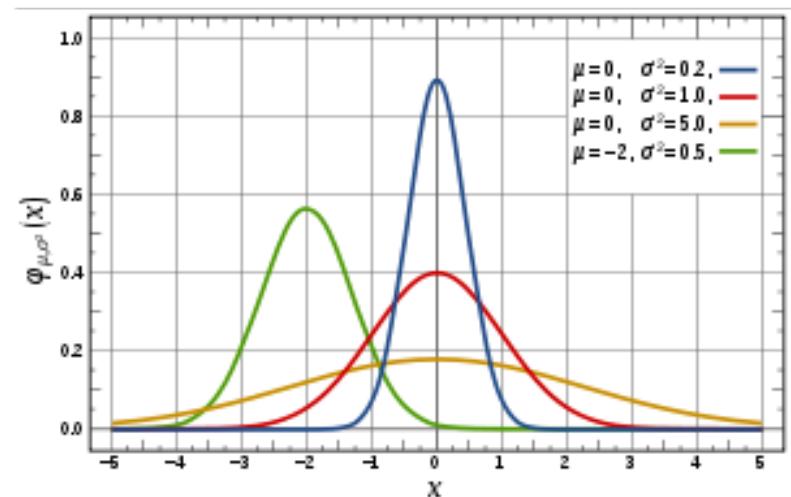
Lévy?

Beyond particle filter

Gaussian distribution

$$f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

1. mean: μ
2. variance: σ^2
3. skewness: zero
4. kurtosis: zero

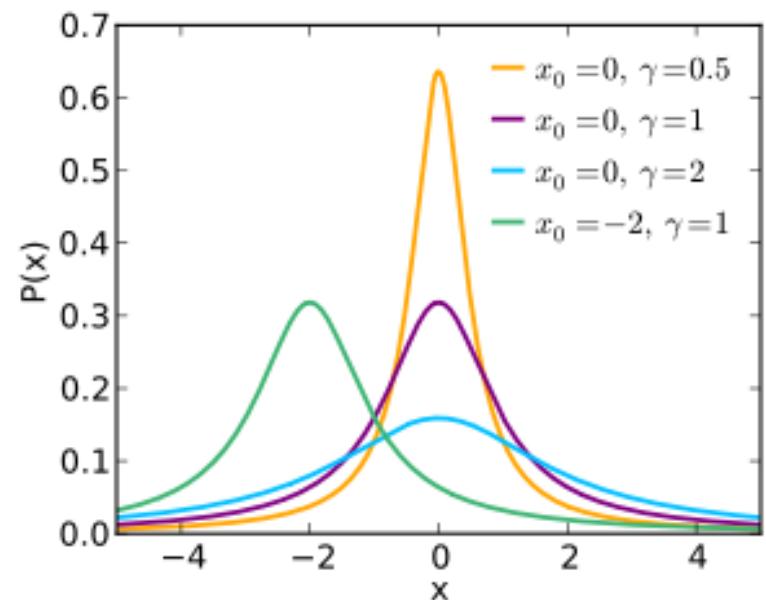


Beyond particle filter

Cauchy distribution

$$f_{\mu,\gamma}(x) = \frac{1}{\pi} \left[\frac{\gamma}{(x - \mu)^2 + \gamma^2} \right]$$

1. mean: not defined
2. variance: not defined
3. skewness: not defined
4. kurtosis: not defined

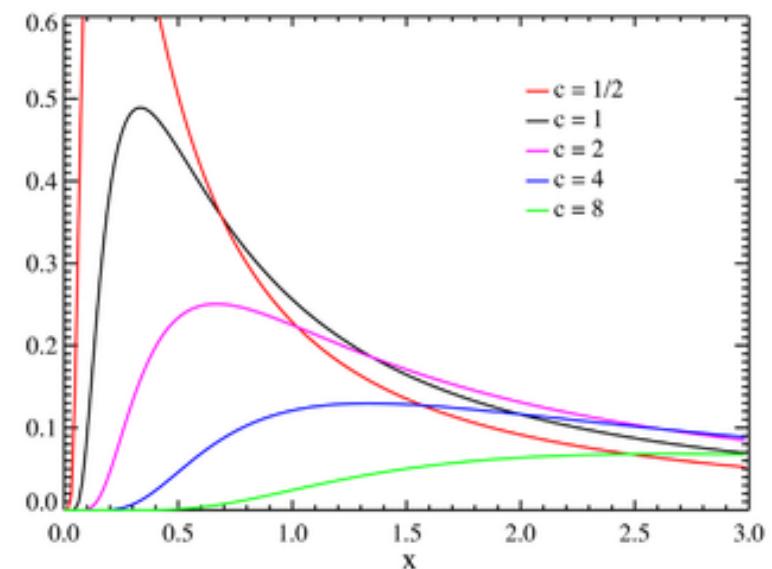


Beyond particle filter

Lévy distribution

$$f_{\mu,c}(x) = \sqrt{\frac{c}{2\pi}} \frac{e^{-c/2(x-\mu)}}{(x - \mu)^{3/2}}$$

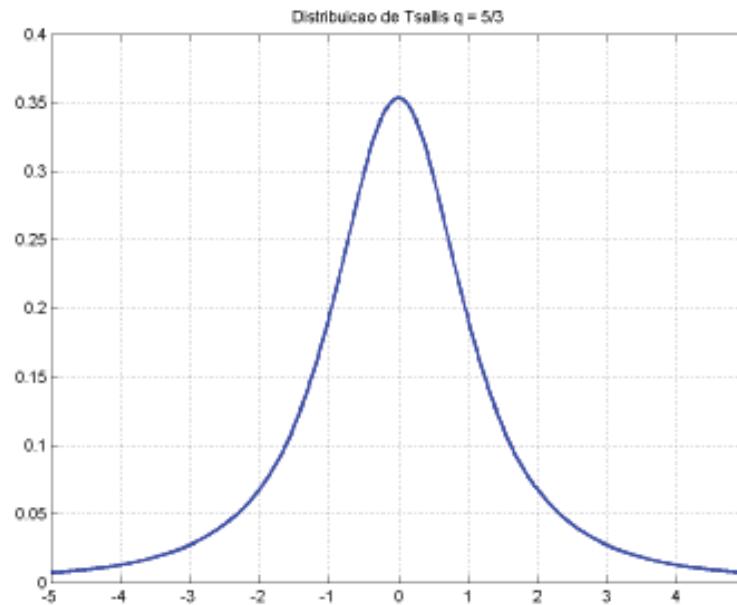
1. mean: can or not exist
2. variance: not defined
3. skewness: not defined
4. kurtosis: not defined



Beyond particle filter

Tsallis distribution: $5/3 < q < 3 \rightarrow \text{Lévy}$

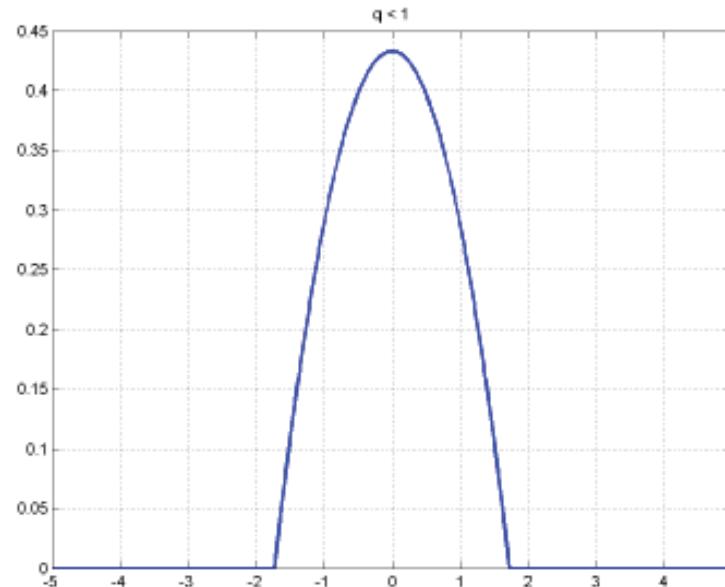
$$q > 1 \quad f(x) = \frac{1}{\sigma} \left[\frac{q-1}{\pi(3-q)} \right]^{\frac{1}{2}} \frac{\gamma\left(\frac{1}{q-1}\right)}{\gamma\left(\frac{3-q}{2(q-1)}\right)} \frac{1}{\left(1 + \frac{q-1}{3-q} \frac{x^2}{\sigma^2}\right)^{\frac{1}{q-1}}}$$



Beyond particle filter

Tsallis distribution: $q < 5/3 \rightarrow \text{Gaussian}$

$$q < 1 \quad f(x) = \frac{1}{\sigma} \left[\frac{1-q}{\pi(3-q)} \right]^{\frac{1}{2}} \frac{\gamma\left(\frac{5-3q}{2(1-q)}\right)}{\gamma\left(\frac{2-q}{1-q}\right)} \left(1 - \frac{1-q}{3-q} \frac{x^2}{\sigma^2}\right)^{\frac{1}{1-q}}$$



Beyond particle filter

- Tsallis' distribution (non-extensive formalism of thermodynamics)

$q < 1$:

$$\rho(x) = \frac{1}{\sigma} \left[\frac{q-1}{\pi(3-q)} \right]^{1/2} \frac{\Gamma(1/(q-1))}{\Gamma((3-q)/[2(q-1)])} \frac{1}{\left[1 + [(q-1)/(3-q)]x/\sigma^2 \right]^{1/(q-1)}}$$

$q = 1$:

$$\rho(x) = \frac{1}{\sigma} \left[\frac{1}{2\pi} \right]^{1/2} e^{-(x/\sigma)^2/2}$$

$q > 1$:

$$\rho(x) = \frac{1}{\sigma} \left[\frac{1-q}{\pi(3-q)} \right]^{1/2} \frac{\Gamma((5-3q)/[2(1-q)])}{\Gamma((2-q)/(1-q))} \left[1 - [(1-q)/(3-q)]x/\sigma^2 \right]^{1/(1-q)}$$

Central Limit Theorem

Our choice for using Gaussian distribution can be justified by the Central Limit theorem.

However, there is another attractor on the distribution space, for some distributions that the second statistical moment diverge.

This is the **Levy-Gnedenko's central limit theorem**.

For the case of Tsallis' distributions:

$$\rho(x) = \begin{cases} G(x) & q < 5/3 \\ L_\gamma(x) & q > 5/3 \end{cases}$$

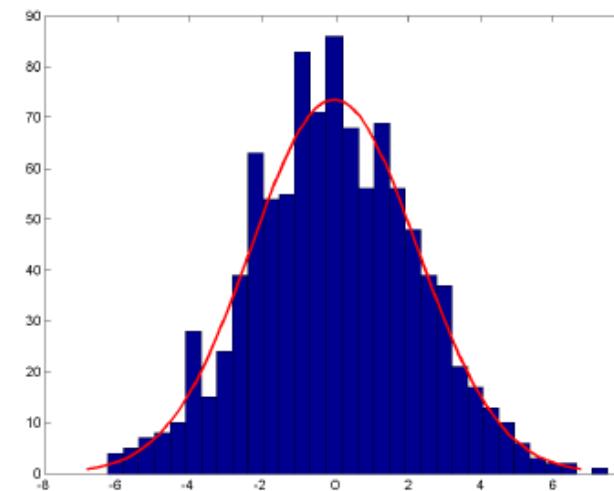
where $L_\gamma(x)$ is a Levy's distribution with index: $\gamma=(3-q)/(q-1)$.

Beyond particle filter

Everything is perfect with particle filter?

Generalized PF:

1. Likelihood function: Tsallis distribution.
2. Non-extensive parameter: estimated
3. Test: 1000 particles



Beyond particle filter

Everything is perfect with particle filter?

Generalized PF:

Prediction:

$$x_{t+1} = \frac{x_t}{2} + \frac{25x_t}{1+x_t^2} + 8\cos(1.2t) + \mu_t$$

Observation:

$$y_{t+1} = \frac{x_t^2}{20} + \nu_t .$$

Beyond particle filter

Everything is perfect with particle filter?

Generalized PF:

Table 1. Different likelihood functions and associated errors.

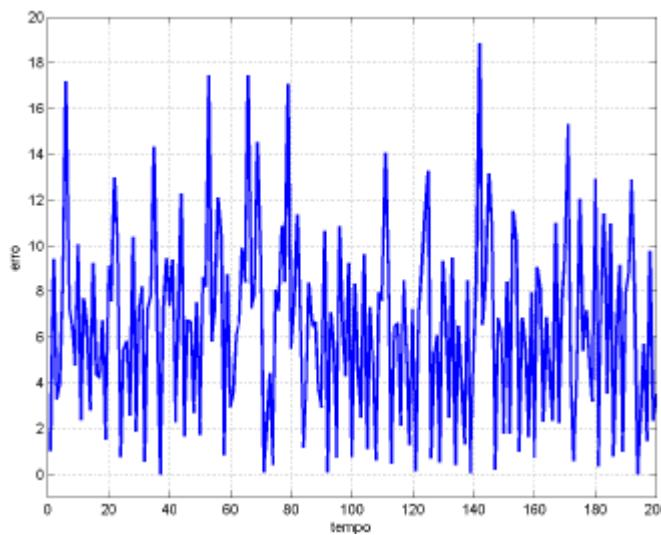
parameter q	error	distribution
1.5	5.4837	Tsallis (fixed- q)
2.9	5.3507	Tsallis (fixed- q)
*	4.5034	Tsallis (estimated- q)
2	5.4203	Cauchy
1	6.1400	Gaussian

Beyond particle filter

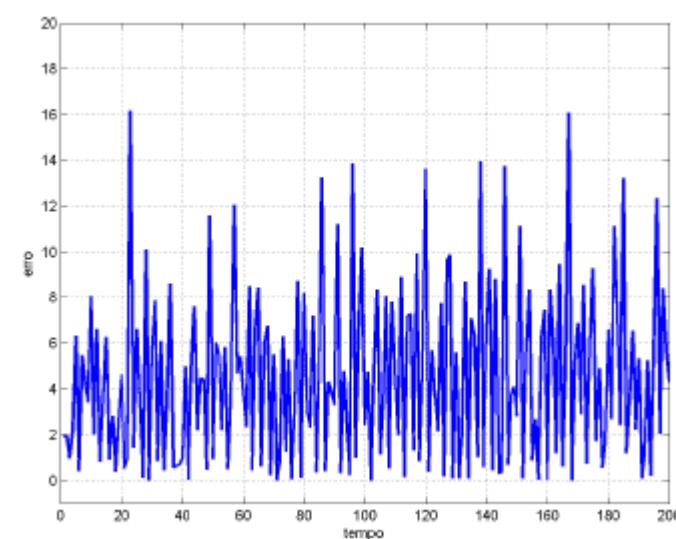
Everything is perfect with particle filter?

Generalized PF:

Gaussian



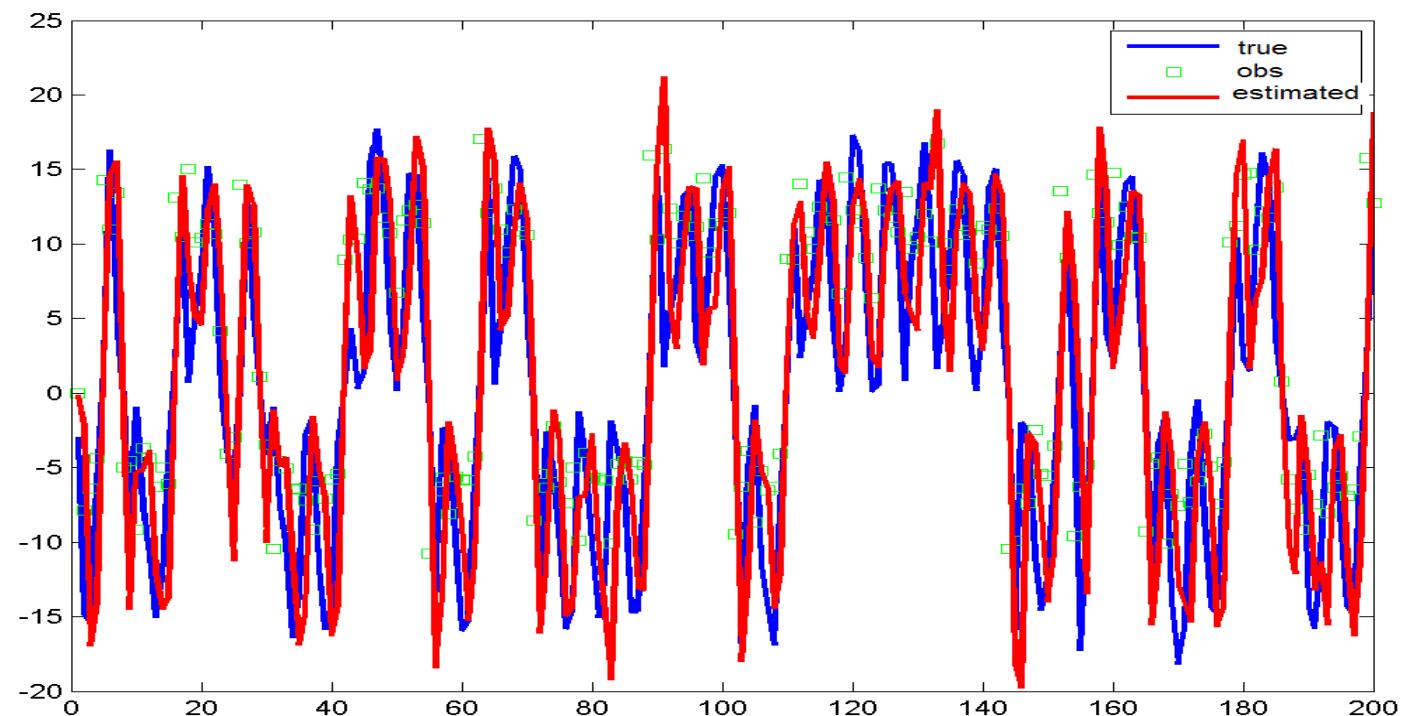
New scheme



Beyond particle filter

Everything is perfect with particle filter?

Generalized PF:

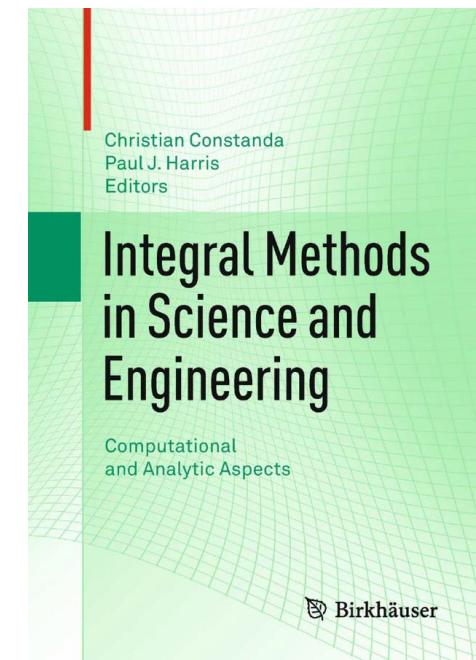


Data assimilation – NN emulating PF

- NN emulating Particle filter: Lorenz's system

Adaptive Particle Filter for Stable Distribution

H.F. de Campos Velho and H.C. Morais Furtado



Data assimilation – NN emulating PF

- NN emulating Particle filter: Lorenz's system

[Integral Methods in Science and Engineering pp 25-35](#) | [Cite as](#)

q-Calculus Formalism for Non-extensive Particle Filter

Amarisio S. Araújo, Helaine C. M. Furtado, Haroldo F. de Campos Velho



Métodos para assimilação de dados

- Relaxação newtoniana (*nudging*)

Short Course on Data Assimilation:

Day 1 (July 11, 2022):

14:30-16:00 – Nudging and backward-forward approach for data assimilation
(Prof. Didier Auroux – Université Côte d'Azur, France)



Available online at www.sciencedirect.com

ScienceDirect

[Advances in Space Research 54 \(2014\) 22–36](http://www.sciencedirect.com/science/journal/0273-1137)

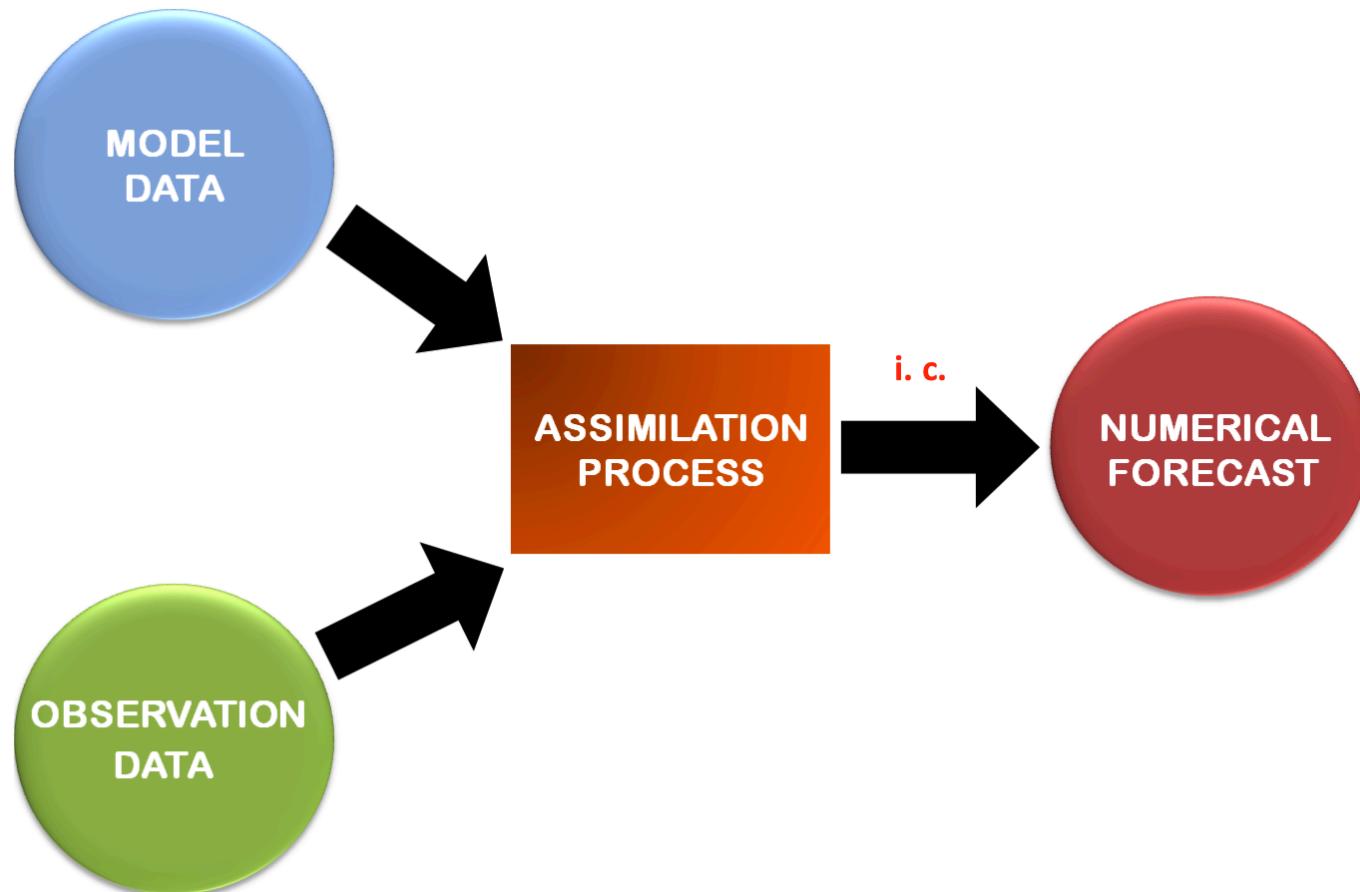
**ADVANCES IN
SPACE
RESEARCH**
(a COSPAR publication)
www.elsevier.com/locate/asr



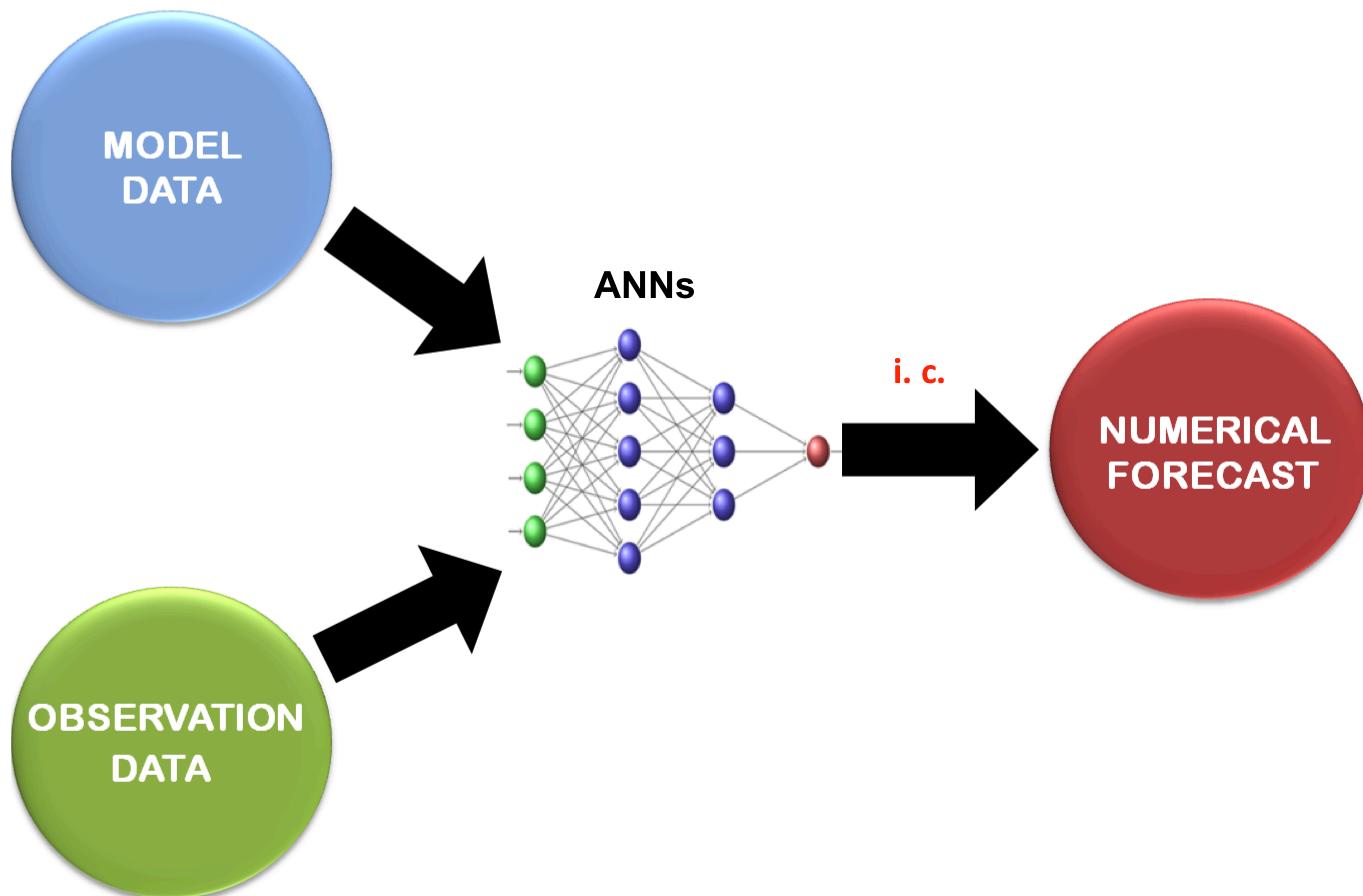
First results of operational ionospheric dynamics prediction
for the Brazilian Space Weather program

Adriano Petry ^{a,*}, Jonas Rodrigues de Souza ^{b,1}, Haroldo Fraga de Campos Velho ^{c,2},
André Grahl Pereira ^{d,3}, Graham John Bailey ^e

Data assimilation: one essential issue



Data assimilation: one essential issue



Kalman filter

Three versions: Linear, Extended, Adaptive (next slide)

1. Advance in time :

$$\begin{aligned}\mathbf{x}_{n+1}^f &= \mathbf{F}_n \mathbf{x}_n \\ \mathbf{P}_{n+1}^f &= \mathbf{F}_n \mathbf{P}_n^a \mathbf{F}_n^T + \mathbf{W}_n^b\end{aligned}$$

4. Update error covariance

$$\mathbf{P}_{n+1}^a = [\mathbf{I} - \mathbf{G}_{n+1} \mathbf{H}_n^T] \mathbf{P}_{n+1}^f$$

2. Kalman gain

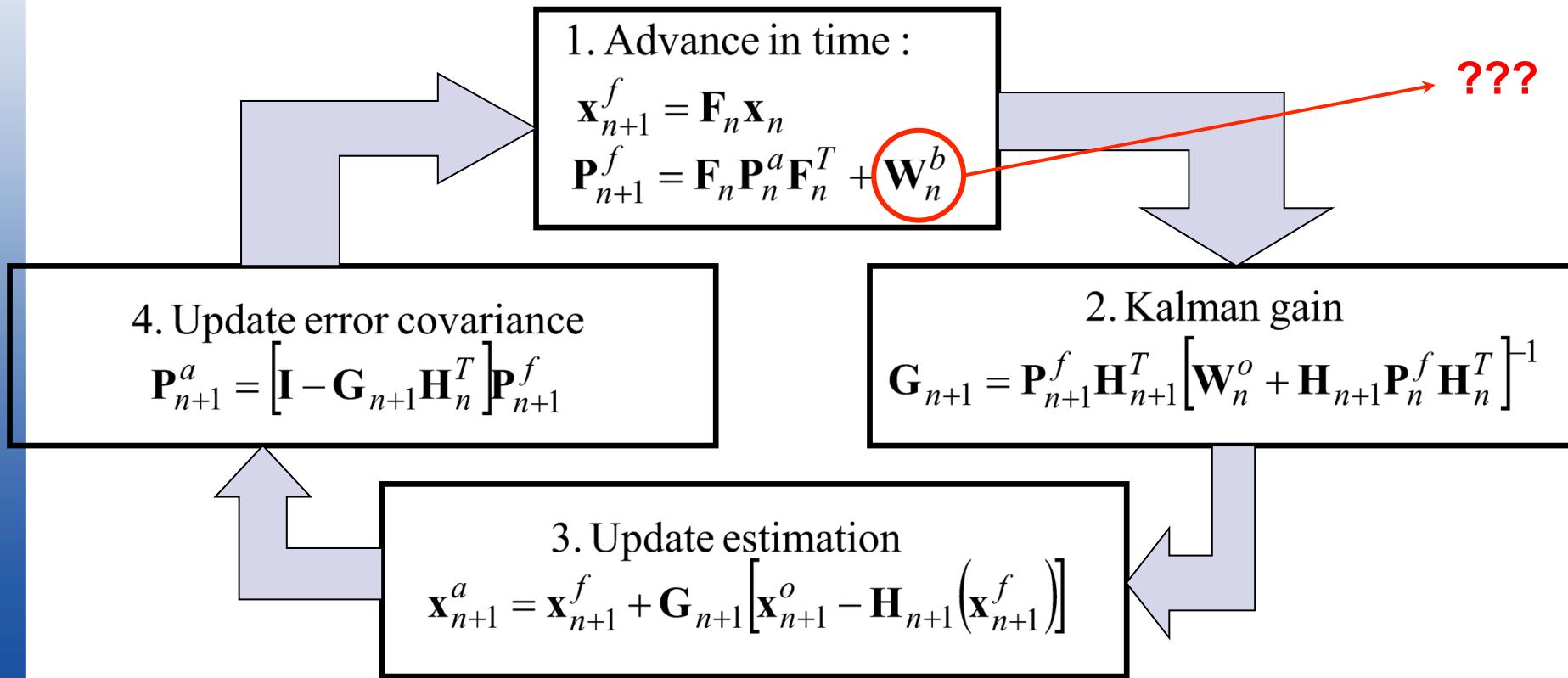
$$\mathbf{G}_{n+1} = \mathbf{P}_{n+1}^f \mathbf{H}_{n+1}^T [\mathbf{W}_n^o + \mathbf{H}_{n+1} \mathbf{P}_n^f \mathbf{H}_n^T]^{-1}$$

3. Update estimation

$$\mathbf{x}_{n+1}^a = \mathbf{x}_{n+1}^f + \mathbf{G}_{n+1} [\mathbf{x}_{n+1}^o - \mathbf{H}_{n+1} (\mathbf{x}_{n+1}^f)]$$

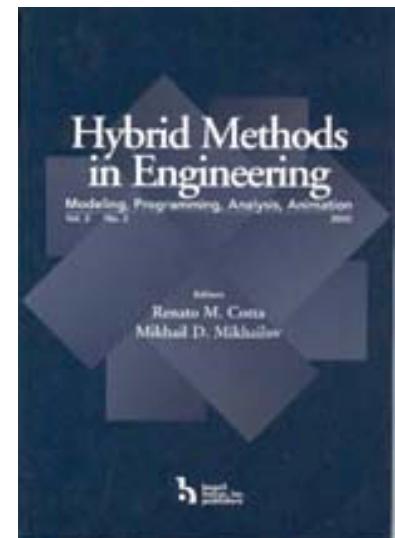
Kalman filter

Three versions: Linear, Extended, Adaptive (next slide)



Kalman filter

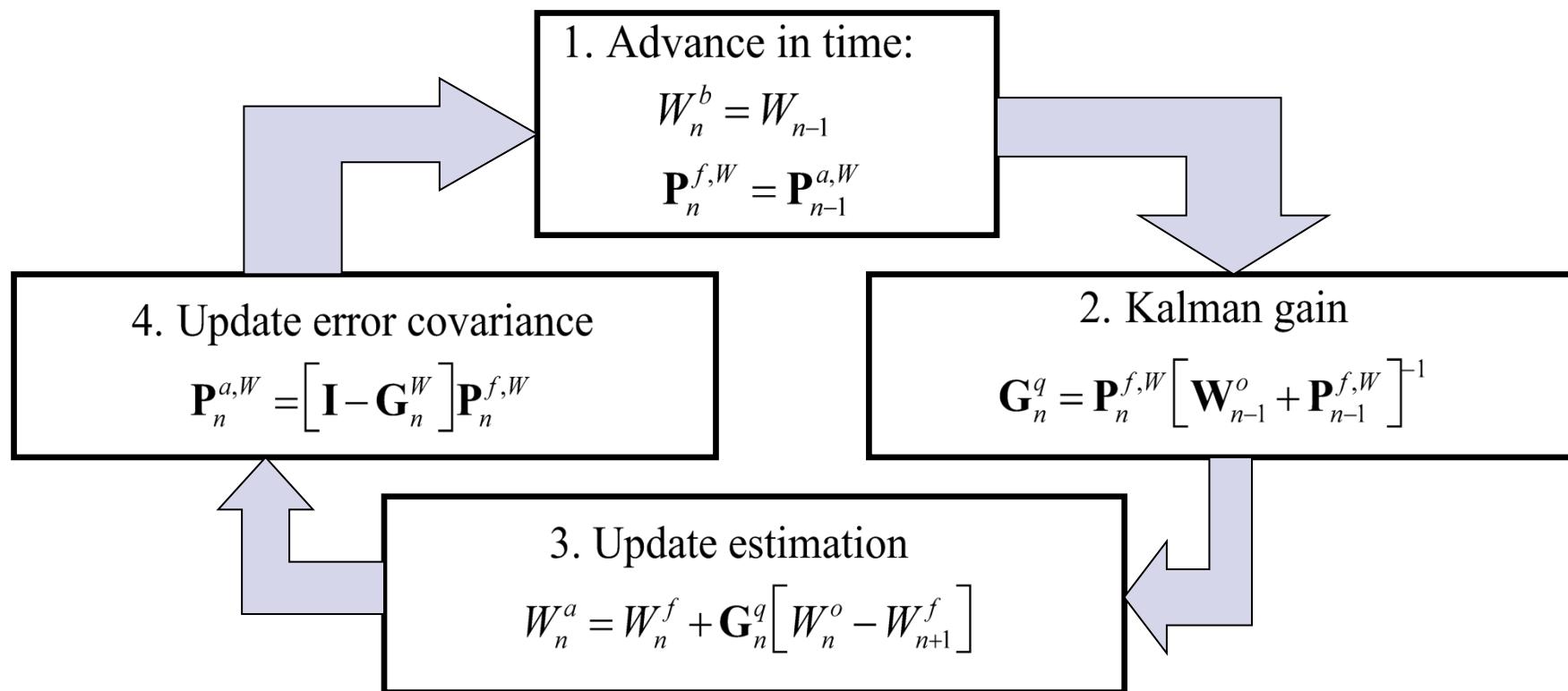
- **Identifying modeling error co-variance matrix**
- Three schemes:
 - Estimating W^b by parameterization
 - Estimating W^b by Fokker-Planck equation
 - Estimating W^b by ensemble strategy



Kalman filter

$$\mathbf{x}_{n+1} = F[\mathbf{x}_n, t_n] \approx \mathbf{F}_n + \frac{\partial F}{\partial \mathbf{x}} \Big|_{t=t_n} \mathbf{x}_n + O(\Delta t^2) \approx \mathbf{E}_n \mathbf{x}_n$$

Three versions: Linear, Extended, **Adaptive**



A.G. Nowosad, A. Rios Neto, H.F. de Campos Velho: Data Assimilation Using an Adaptive Kalman Filter and Laplace Transform, Hybrid Methods in Engineering, 2(3), 291-310, 2000.

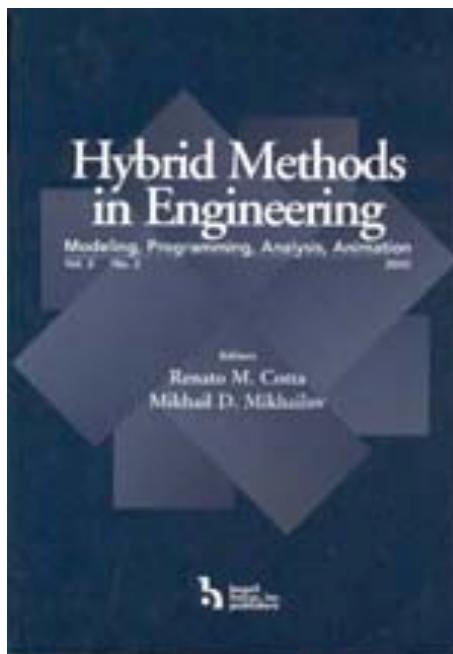
Data assimilation: adaptive Kalman filter

Data Assimilation Using an Adaptative Kalman Filter and Laplace Transform

A.G. Nowosad^a (DCM)

A. Rios Neto^b

H.F. de Campos Velho^a (LAC)



^a Instituto Nacional de Pesquisas Espaciais (INPE)
Caixa Postal 515
12201-970 – São José dos Campos (SP), BRAZIL,
E-mail: alex@met.inpe.br - haroldo@lac.inpe.br

^b Instituto de Pesquisa e Desenvolvimento (IP&D)
Universidade do Vale do Paraíba (UNIVAP)
Av. Shishima Hifumi, 2.911 - Urbanova
12245-720 - São José dos Campos (SP), BRAZIL
E-mail: atair@univap.br

Hybrid Methods in Engineering: (2000) 2(3):
291-310

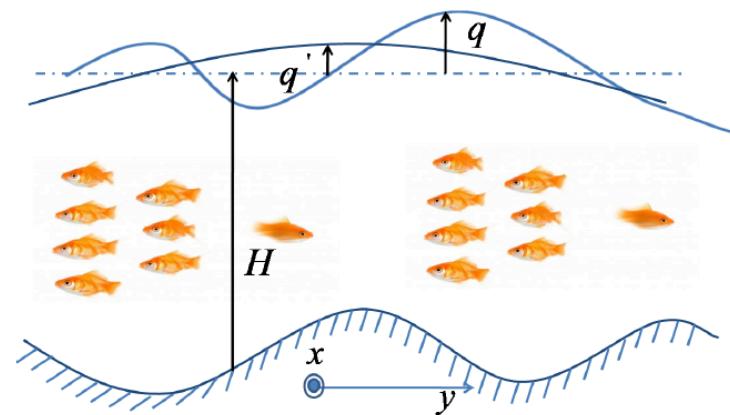
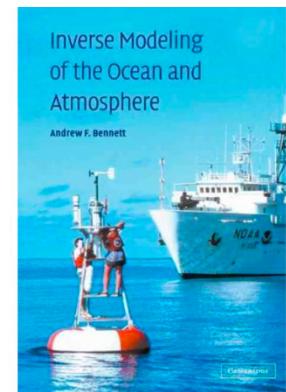
Kalman filter: application for DA

- Shallow water 2D for ocean circulation

$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$

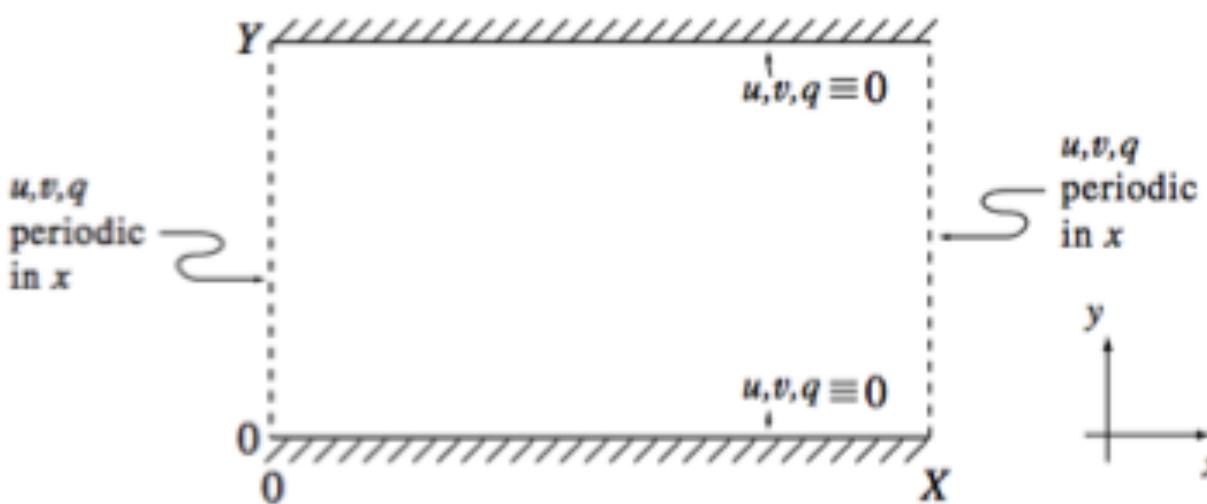


Shallow water 2D for ocean circulation

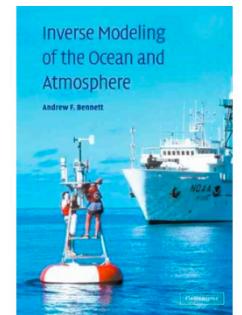
$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

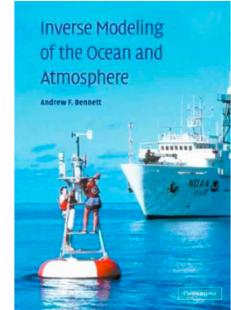
$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$



(BENNETT, 2002)



Shallow water 2D for ocean circulation



Parameter	Value	Parameter	Value
H	5000 m	r_u	$1.8 \times 10^4 \text{ s}^{-1}$
T	$1.8 \times 10^4 \text{ s}$	r_v	$1.8 \times 10^4 \text{ s}^{-1}$
g	9.806 m s^{-2}	r_q	$1.8 \times 10^4 \text{ s}^{-1}$
f	$1.0 \times 10^{-4} \text{ s}^{-1}$	ρ_a	1.275 kg m^{-3}
C_d	1.6×10^{-3}	ρ_w	$1.0 \times 10^3 \text{ kg m}^{-3}$

$$\frac{\partial u}{\partial t} - fv + g\frac{\partial q}{\partial x} + r_u u = F_u$$

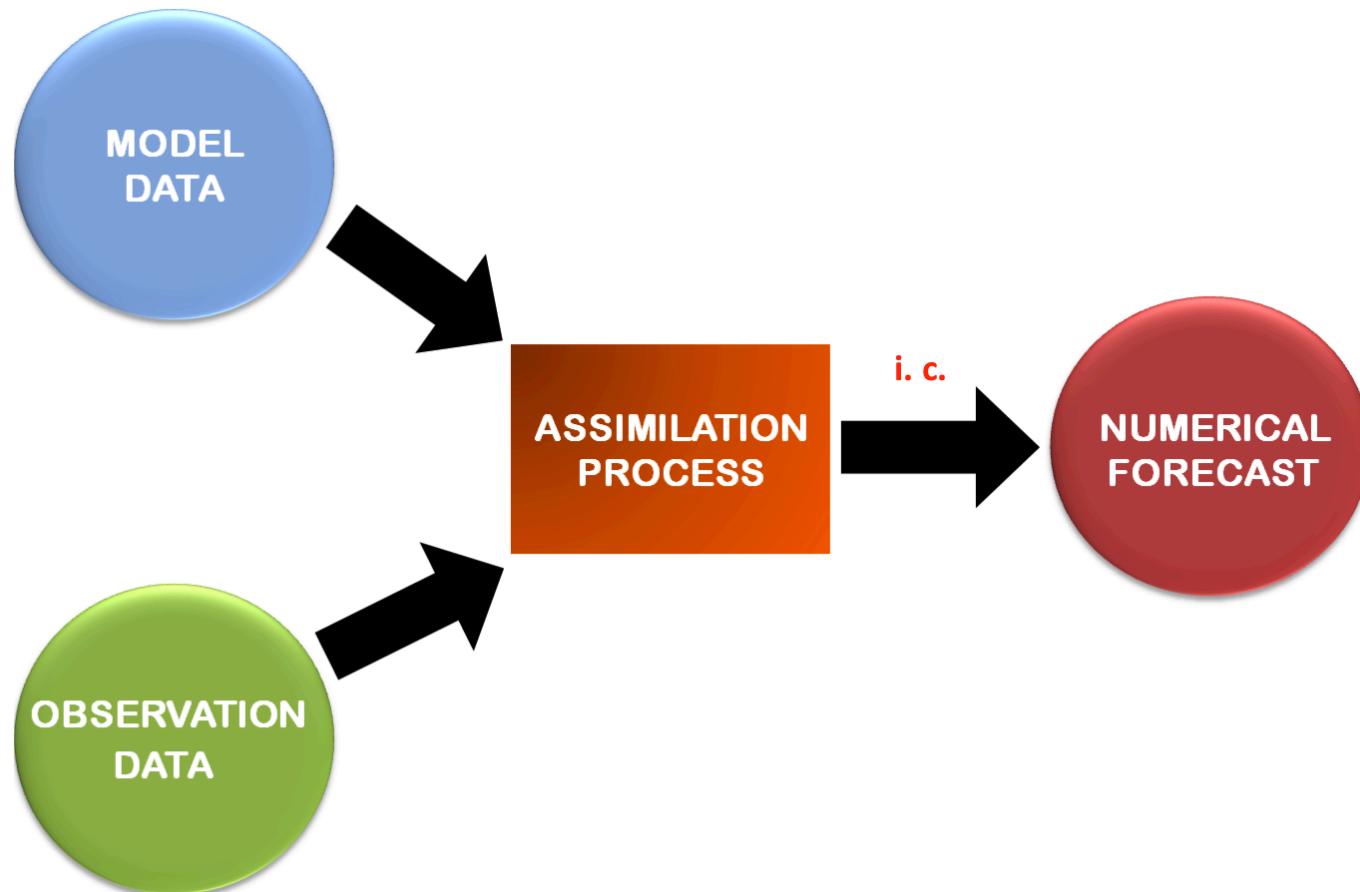
$$\frac{\partial v}{\partial t} + fu + g\frac{\partial q}{\partial y} + r_v v = F_v$$

$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$

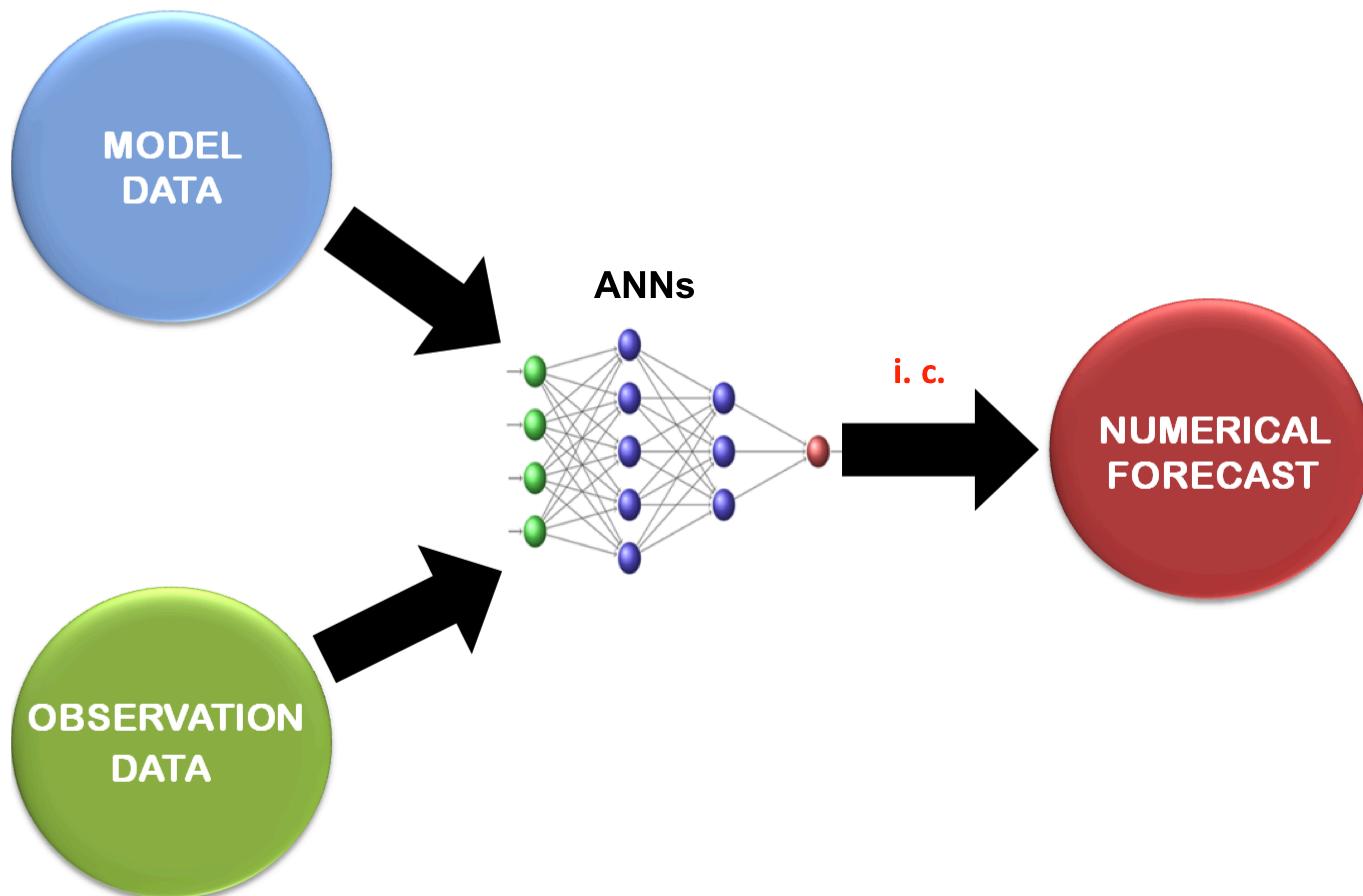
Kalman filter: application for DA

- **Reducing effort: Algorithm by KF x NN**
 - Neural network emulating KF
 - Self-configuring neural network by optimization problem
 - Optimization problem solved by meta-heuristic named Multi-Particle Collision Algorithm (MPCA)

Data assimilation: one essential issue



Data assimilation: one essential issue



Finding an OPTIMAL neural network

- Design of supervised neural network:
Optimization problem – cost function:

$$E_{trein} = \frac{1}{N} \sum_{k=1}^N (d_k - s_k)^2$$
$$E_{gen} = \frac{1}{(M-N+1)} \sum_{k=N+1}^M (d_k - s_k)^2$$
$$F_{obj} = \text{penalty} * \frac{\rho_1 * E_{trein} + \rho_2 * E_{gen}}{\rho_1 + \rho_2}$$
$$\text{penalty} = \underbrace{\left(c_1 * \left(e^{\#neuron} \right)^2 \right)}_{\text{complexity factor-1}} \times \underbrace{\left(c_2 * (\#epoch) \right)}_{\text{complexity factor-2}} + 1$$

MPCA: Multi-Particle Collision Algorithm

Available for download:

www.epacis.net/jcis/PDF_JCIS/JCIS11-art.01.pdf



Journal of Computational Interdisciplinary Sciences (2008) 1(1): 3-10

© 2008 Pan-American Association of Computational Interdisciplinary Sciences

ISSN 1983-8409

<http://epacis.org>

A new multi-particle collision algorithm for optimization in a high performance environment

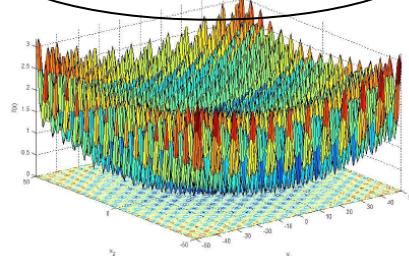
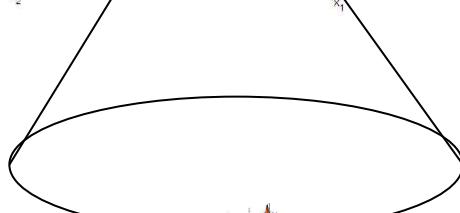
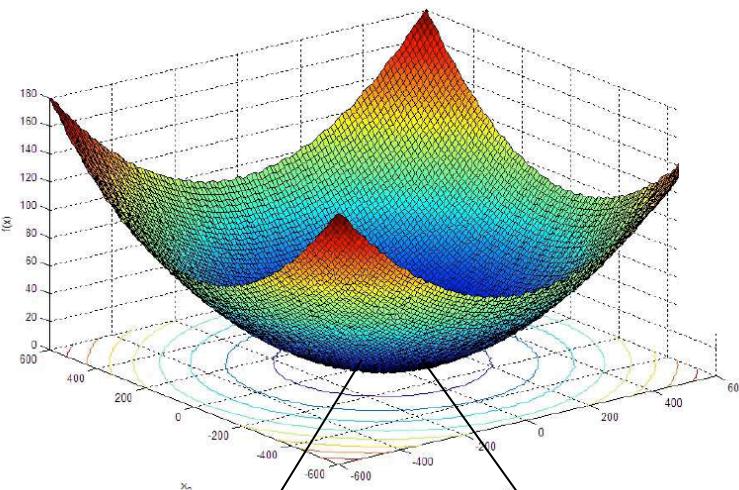
Eduardo Fávero Pacheco da Luz, José Carlos Becceneri and Haroldo Fraga de Campos Velho

Manuscript received on July 31, 2008 / accepted on October 5, 2008



PCA vs MPCA (2)

Griewank function



$$f(x_1, \dots, x_n) = 1 + \sum_{i=1}^n \frac{x_i^2}{4000} - \prod_{j=1}^n \cos\left(\frac{x_j}{\sqrt{n}}\right)$$

$$\left\| (x_1, \dots, x_n) \right\|_2^2 \leq 600$$

$$\min : (0, \dots, 0), \quad f(0, \dots, 0) = 0$$

PCA

$$(-3.14, 4.43)$$

$$f(x_1, x_2) = 7.4 \times 10^{-3}$$

MPCA

$$(-1.8 \times 10^{-8}, -3.3 \times 10^{-8})$$

$$f(x_1, x_2) = 3.3 \times 10^{-16}$$

Finding an OPTIMAL neural network

- Supervised neural network: Multi-Layer Perceptron (MLP)

MPCA solution



# hidden layers	# neurons layer-1	# neurons layer-2	# neurons layer-3	Activation function	Momentum ratio	Learning ratio
Parameters			Value			
Number of hidden layers			1 2 3			
Number of neurons for each layer			1 ... 32			
Learning ratio			0 ... 1			
Momentum			0 ... 0.9			
Activation function			Tanh Log Gauss			

Applications ...

Data assimilation by parallel NN: CPU multi-core

Cooperação INPE-LNCC

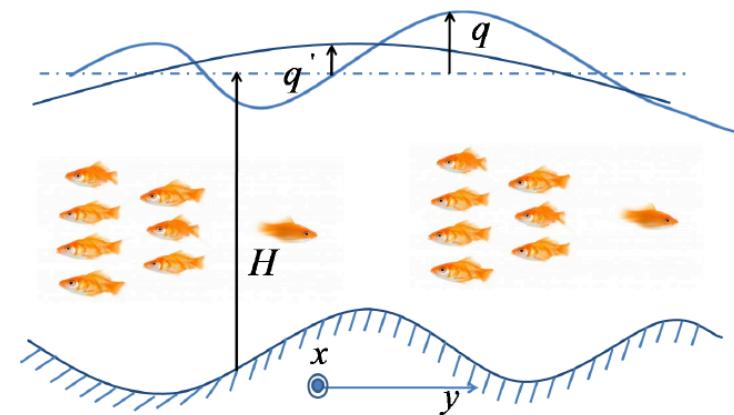
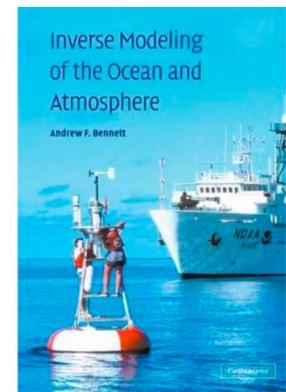
Kalman filter: application for DA

- Shallow water 2D for ocean circulation

$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$



Kalman filter: application for DA

- SW 2D for ocean circulation: KF x NN

DOI: 10.14529/jsfi220105

Data Assimilation by Neural Network for Ocean Circulation:
Parallel Implementation

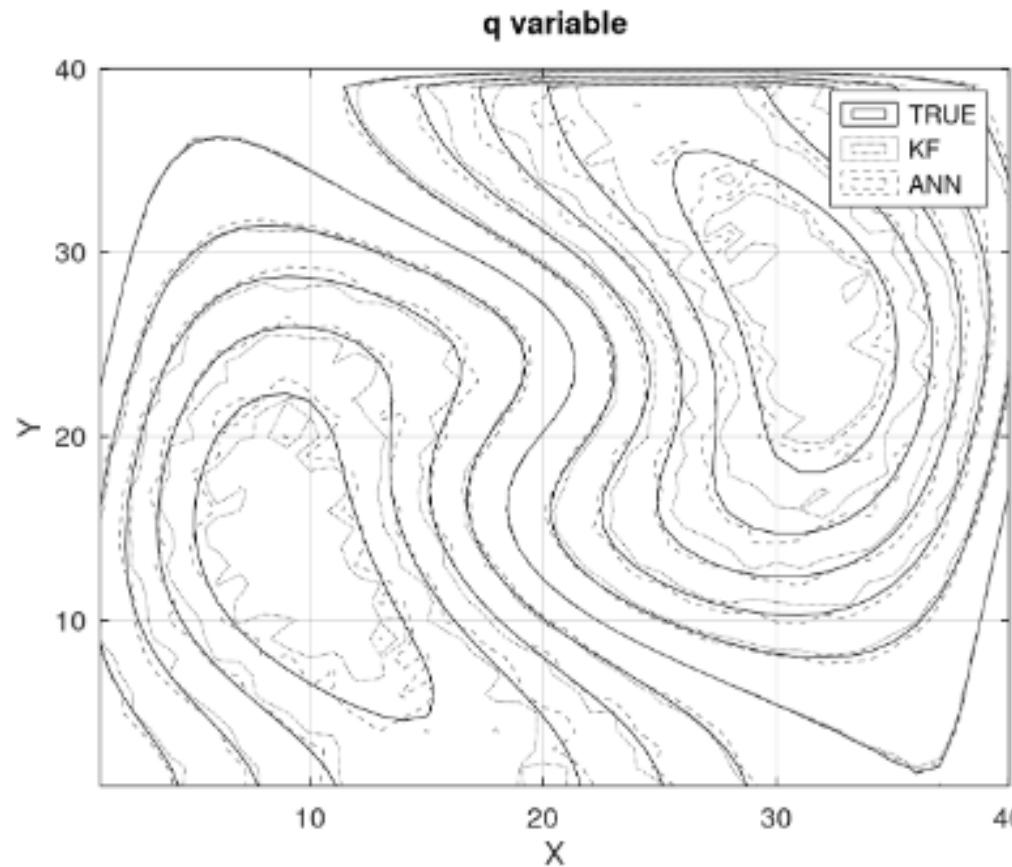
*Haroldo F. Campos Velho¹, Helaine C. M. Furtado²,
Sabrina B. M. Sambatti³, Carla O. F. Barros⁴, Maria E. S. Welter⁴,
Roberto P. Souto⁴, Diego Carvalho⁵, Douglas O. Cardoso^{6,7}*

© The Authors 2022. This paper is published with open access at SuperFri.org



Kalman filter: application for DA

- **SW 2D for ocean circulation: KF x NN**
 - Assimilating ocean high estimation



Kalman filter: application for DA

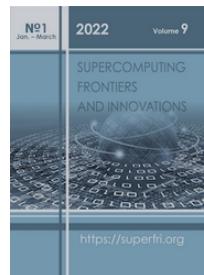
- **SW 2D for ocean circulation: KF x NN**
 - CPU-time: KF x NN



Experiment	Kalman filter	MLP-NN
Obs: 25 points	00:42:02	00:01:39
Obs: 100 points	01:19:03	00:05:01

Kalman filter: application for DA

- **SW 2D for ocean circulation: KF x NN**
 - CPU-time: KF x NN



Experiment	Kalman filter	MLP-NN
Obs: 25 points	00:42:02	00:01:39
Obs: 100 points	01:19:03	00:05:01

Applications ...

Data assimilation by TPU

Doutorado Computação-ITA: Marcelo Paiva

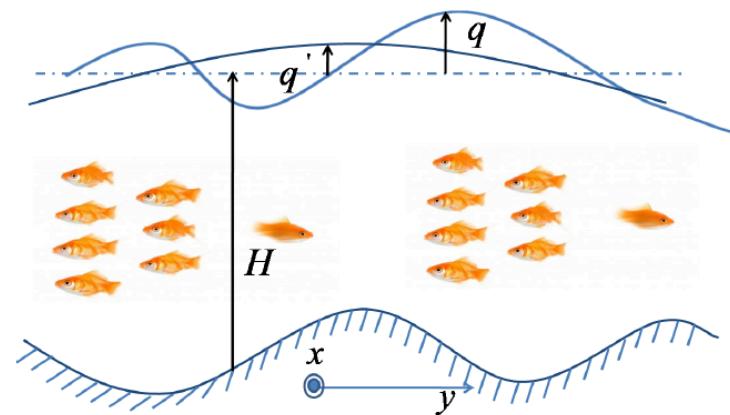
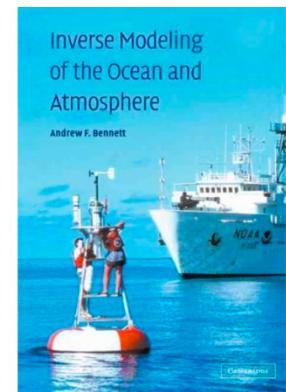
Kalman filter: application for DA

- Shallow water 2D for ocean circulation

$$\frac{\partial u}{\partial t} - fv + g \frac{\partial q}{\partial x} + r_u u = F_u$$

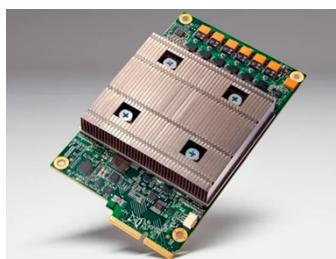
$$\frac{\partial v}{\partial t} + fu + g \frac{\partial q}{\partial y} + r_v v = F_v$$

$$\frac{\partial q}{\partial t} + H \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + r_q q = 0$$

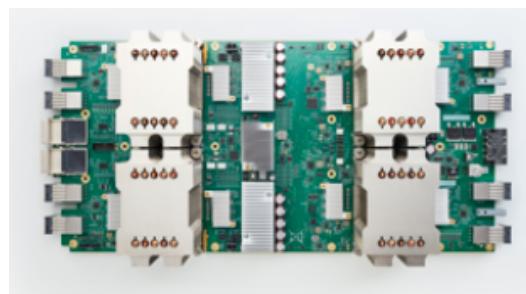


Kalman filter: application for DA

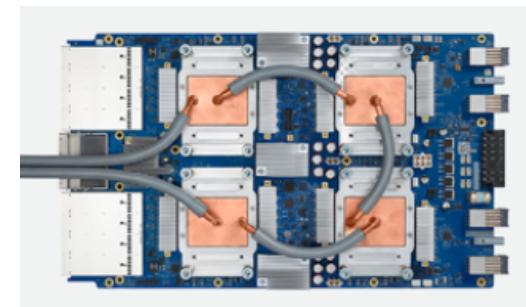
- **Reducing effort: hardware for NN**
 - Co-processor option: TPU-Google
 - Google Cloud: CoLab



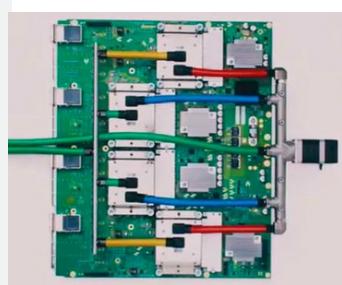
TPU-v1



TPU-v2



TPU-v3



TPU-v4

Kalman filter: application for DA

- **Reducing effort: hardware for NN**
 - TPU-v3 by Google-CoLab: 10 assimilation cycles

Resolution ($N_x \times N_y$)	CPU (seconds)	TPU (seconds)
40×40	0.53804	0.79083
100×100	0.21106	0.37790
150×150	0.32155	0.36680
200×200	0.44240	0.36723

Kalman filter: application for DA

- **Reducing effort: hardware for NN**
 - TPU-v3 by Google-CoLab: 10 assimilation cycles

Resolution ($N_x \times N_y$)	CPU (seconds)	TPU (seconds)
40×40	0.53804	0.79083
100×100	0.21106	0.37790
150×150	0.32155	0.36680
200×200	0.44240	0.36723
250×250	0.62247	0.38423
300×300	0.83743	0.35689
350×350	1.11788	0.40157
400×400	1.48095	0.38469