Geir Evensen

Data Assimilation

Geir Evensen

Data Assimilation

The Ensemble Kalman Filter

With 63 Figures

Springer

PROF. GEIR EVENSEN

Hydro Research Centre, Bergen
PO Box 7190
N 5020 Bergen
Norway

and

Mohn-Sverdrup Center for Global Ocean Studies
and Operational Oceanography
at Nansen Environmental and Remote Sensing Center
Thormølensgt 47
N 5600 Bergen
Norway

e-mail:
Geir.Evensen@hydro.com

Library of Congress Control Number: 2006932964

ISBN-10 3-540-38300-0-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-38300-0 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media
springer.com

Cover design: Erich Kirchner
Typesetting: camera-ready by the author
Production: Christine Adolph
Printing: Krips bv, Meppel
Binding: Stürtz AG, Würzburg

Printed on acid-free paper 30/2133/ca 5 4 3 2 1 0

To Tina and Endre

Preface

The aim of this book is to introduce the formulation and solution of the data assimilation problem. The focus is mainly on methods where the model is allowed to contain errors and where the error statistics evolve through time. So-called strong constraint methods and simple methods where the error statistics are constant in time are only briefly explained, and then as special cases of more general weak constraint formulations.

There is a special focus on the Ensemble Kalman Filter and similar methods. These are methods which have become very popular, both due to their simple implementation and interpretation and their properties with nonlinear models.

The book has been written during several years of work on the development of data assimilation methods and the teaching of data assimilation methods to graduate students. It would not have been completed without the continuous interaction with students and colleagues, and I particularly want to acknowledge the support from Laurent Bertino, Kari Brusdal, François Counillon, Mette Eknes, Vibeke Haugen, Knut Arild Lisæter, Lars Jørgen Natvik, and Jan Arild Skjervheim, with whom I have worked closely for several years. Laurent Bertino and François Counillon also provided much of the material for the chapter on the TOPAZ ocean data assimilation system. Contributions from Laurent Bertino, Theresa Lloyd, Gordon Wilmot, Martin Miles, Jennifer Trittschuh-Vallès, Brice Vallès and Hans Wackernagel, on proof-reading parts of the final version of the book are also much appreciated.

It is hoped that the book will provide a comprehensive presentation of the data assimilation problem and that it will serve as a reference and textbook for students and researchers working with development and application of data assimilation methods.

Bergen, June 2006 *Geir Evensen*

Contents

List of symbols

a	De-correlation lengths in variogram models (11.2–11.4); Error in initial condition for scalar models (5.5), (5.22), (8.22) and (12.22)
$A(z)$	Vertical diffusion coefficient in Ekman model, Sect. 5.3
$A_0(z)$	First guess of vertical diffusion coefficient in Ekman model, Sect. 5.3
$\boldsymbol{a}$	Error in initial condition for vector models (5.70), (6.8–6.10), (7.2)
$\boldsymbol{A}_i$	Ensemble matrix at time t_i, Chap. 9
$\boldsymbol{A}$	Ensemble matrix, Chap. 9
$\boldsymbol{b}$	Vector of coefficients solved for in the analysis scheme (3.38)
$\boldsymbol{b}(\boldsymbol{x}, t)$	Error in boundary condition, Chap. 7
$\boldsymbol{b}_0$	Stochastic error in upper condition of Ekman model, Sect. 5.3
$\boldsymbol{b}_H$	Stochastic error in lower condition of Ekman model, Sect. 5.3
c	Constant in Fourier spectrum (11.10) and (11.14); constant multiplier used when simulating model errors (12.20)
c_i	Multiplier used when simulating model errors (12.54)
c_d	Wind-drag coefficient in Ekman model, Sect. 5.3
c_{d_0}	First guess value of wind-drag coefficient in Ekman model, Sect. 5.3
c_{rep}	Constant multiplier used when modelling model errors in representer method, Chap. 12
$C_{\psi\psi}$	Covariance of a scalar state variable ψ
$C_{c_d c_d}$	Covariance of error in wind-drag c_{d_0}
$C_{AA}(z_1, z_2)$	Covariance of error in vertical diffusion $A_0(z)$
$C_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$	Covariance of scalar field $\psi(\boldsymbol{x})$ (2.25)

$\boldsymbol{C}_{\psi\psi}$	Covariance of a discrete $\boldsymbol{\psi}$ (sometimes short for $\boldsymbol{C}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$)
$\boldsymbol{C}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$	Covariance of a vector of scalar state variables $\boldsymbol{\psi}(\boldsymbol{x})$
$C_{\epsilon\epsilon}$	Used for variance of ϵ in Chap. 3
C_{aa}	Scalar initial error covariance
C_{qq}	Model error covariance
$\boldsymbol{C}$	Matrix to be inverted in the ensemble analysis schemes, Chap. 9
$\boldsymbol{C}_{\epsilon\epsilon}$	Covariance of measurement errors $\boldsymbol{\epsilon}$
$\boldsymbol{C}^{\mathrm{e}}_{\epsilon\epsilon}$	Low-rank representation of measurement error covariance
$\boldsymbol{C}_{aa}$	Initial error covariance
$\boldsymbol{C}_{qq}$	Model error covariance
d	Measurement
$\boldsymbol{d}$	Vector of measurements
$\boldsymbol{D}$	Perturbed measurements, Chap. 9
$\boldsymbol{D}_j$	Perturbed measurements at data time j, Chap. 9
$\boldsymbol{E}$	Measurement perturbations, Chap. 9
$f(\boldsymbol{x})$	Arbitrary function, e.g. in (3.55)
$f(\psi)$	Probability density, e.g. $f(\psi)$ or $f(\boldsymbol{\psi})$ where $\boldsymbol{\psi}$ is a vector or a vector of fields
F	Distribution function (2.1)
$g(\boldsymbol{x})$	Arbitrary function, e.g. in (3.55) and (3.59)
G	Model operator for a scalar state; linear (4.1) or nonlinear (4.14)
$\boldsymbol{G}$	Model operator for a vector state; linear (4.11) or nonlinear (4.21), (9.1) and (7.1)
$h()$	Arbitrary function used at different occasions
H	Depth of bottom boundary in Ekman model, Sect. 5.3
$\boldsymbol{h}$	Innovation vector (3.51); spatial distance vector (11.1)
$i(j)$	Time index corresponding to measurement j, Fig. 7.1
$\boldsymbol{I}$	Identity matrix
J	Number of measurement times, Fig. 7.1
$\boldsymbol{k}$	Vertical unit vector $(0, 0, 1)$ in Ekman model Sect. 5.3; wave number $\boldsymbol{k} = (\kappa, \lambda)$, Chap. 11
$\boldsymbol{k}_{\mathrm{h}}$	Permeability, Chap. 16
$\boldsymbol{K}$	Kalman gain matrix (3.85)
m_j	Total number of measurements at measurement time j , Fig. 7.1
m	Sometimes used as abbreviation for m_j
$m(\psi)$	Nonlinear measurement functional in the Appendix

M	Total number of measurements over the assimilation interval
$\boldsymbol{M}$	Measurement matrix for a discrete state vector (3.76); measurement matrix operator (10.20)
n	Dimension of state vector $n = n_\psi + n_\alpha$, Chap. 9 and 10
n_α	Number of parameters, Chaps. 9 and 10
n_ψ	Dimension of model state, Chaps. 7–10
n_x	Gridsize in x-direction, Chap. 11
n_y	Gridsize in y-direction, Chap. 11
N	Sample or ensemble size
p	Error of first guess of a scalar or scalar field, Chap. 3; matrix rank, Chap. 14; probability (6.24)
p_A	Error of first guess vertical diffusion coefficient in Ekman model, Sect. 5.3
p_{c_d}	Error of first guess wind drag coefficient in Ekman model, Sect. 5.3
P	Reservoir pressure, Chap. 16
q	Stochastic error of scalar model used in Kalman filter formulations
$\boldsymbol{q}(i)$	Discrete model error at time t_i, (6.16)
$\boldsymbol{q}$	Stochastic error of vector model used in Kalman filter formulations
$\boldsymbol{Q}$	Ensemble of model noise, Sect. 11.4
r	De-correlation length in Fourier space (11.10)
r_1	De-correlation length in principal direction in Fourier space (11.11)
r_2	De-correlation length orthogonal to principal direction in Fourier space (11.11)
r_x	De-correlation length in principal direction in physical space (11.23)
r_y	De-correlation length orthogonal to principal direction in physical space (11.23)
$\boldsymbol{r}(\boldsymbol{x}, t)$	Vector of representer functions (3.39) and (5.48)
$\boldsymbol{r}$	Matrix of representers (3.80)
$\boldsymbol{R}$	Representer matrix (3.63)
R_{s}	Gas in a fluid state at reservoir conditions, Chap. 16
R_{v}	Oil in a gas state at reservoir conditions, Chap. 16
S_{w}	Water saturation, Chap. 16
S_{g}	Gas saturation, Chap. 16

S_{o}	Oil saturation, Chap. 16
$\boldsymbol{s}(\boldsymbol{x}, t)$	Vector of adjoints of representer functions (5.49)
$\boldsymbol{S}_j$	Measurement of ensemble perturbations at data time j, Chap. 9
$\boldsymbol{S}$	Measurement of ensemble perturbations, Chap. 9
t	Time variable
T	Final time of assimilation period for some examples
u	Dependent variable (5.99)
$\boldsymbol{u}(z)$	Horizontal velocity vector in Ekman model, Sect. 5.3
$\boldsymbol{u}_0(z)$	Initial condition for velocity vector in Ekman model, Sect. 5.3
$\boldsymbol{U}$	Left singular vectors from the singular value decomposition, Sect. 11.4 and (14.68)
$\boldsymbol{U}_0$	Left singular vectors from the singular value decomposition (14.19)
$\boldsymbol{U}_1$	Left singular vectors from the singular value decomposition (14.52)
$\boldsymbol{v}$	Dummy vector (5.101)
$\boldsymbol{V}$	Right singular vectors from the singular value decomposition, Sect. 11.4 and (14.68)
$\boldsymbol{V}_0$	Right singular vectors from the singular value decomposition (14.19)
$\boldsymbol{V}_1$	Right singular vectors from the singular value decomposition (14.52)
w_k	Random realization with mean equal to zero and variance equal to one (11.33)
W_{aa}	Inverse of scalar initial error covariance
$\boldsymbol{W}$	Matrix (14.63) and (14.64)
$\boldsymbol{W}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$	Functional inverse of $\boldsymbol{C}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$, e.g. (3.27)
$\boldsymbol{W}_{aa}(\boldsymbol{x}_1, \boldsymbol{x}_2)$	Functional inverse of initial error covariance
$\boldsymbol{W}_{aa}$	Inverse of initial error covariance
$\boldsymbol{W}_{\eta\eta}$	Smoothing weight (6.19)
$\boldsymbol{W}_{\epsilon\epsilon}$	Matrix inverse of the covariance $\boldsymbol{C}_{\epsilon\epsilon}$
$\boldsymbol{x}$	Independent spatial variable
x_n	x-position in grid $x_n = n\Delta x$, Chap 11
$\boldsymbol{X}_0$	Matrix (14.26) and (14.51)
$\boldsymbol{X}_1$	Matrix (14.30) and (14.55)
$\boldsymbol{X}_2$	Matrix (14.34) and (14.59)
$x,\ y,\ z$	Dependent variables in Lorenz equations (6.5–6.6)

$\boldsymbol{x}$	Dependent variable $\boldsymbol{x}^{\mathrm{T}} = (x, y, z)$ in Lorenz equations
$\boldsymbol{x}_0$	Initial condition $\boldsymbol{x}_0^{\mathrm{T}} = (x_0, y_0, z_0)$ in Lorenz equations
y_m	y-position in grid $y_m = m\Delta y$, Chap 11
$\boldsymbol{Y}$	Matrix (14.65)
$\boldsymbol{Z}$	Matrix of eigenvectors from eigenvalue decomposition
$\boldsymbol{Z}_1$	Matrix of eigenvectors from eigenvalue decomposition (14.27)
$\boldsymbol{Z}_p$	Matrix of p first eigenvectors from eigenvalue decomposition (14.15)
$\mathcal{B}$	Penalty function in measurement space, e.g. $\mathcal{B}[\boldsymbol{b}]$ in (3.66)
$\mathcal{D}$	Model domain
$\partial\mathcal{D}$	Boundary of model domain
$\mathcal{H}$	Hamiltonian, used in hybrid Monte Carlo algorithm (6.25)
$\boldsymbol{\mathcal{H}}$	Hessian operator (second derivative of model operator)
$\mathcal{J}$	Penalty function, e.g. $\mathcal{J}[\boldsymbol{\psi}]$
$\mathcal{M}$	Scalar measurement functional (3.24)
$\boldsymbol{\mathcal{M}}$	Vector of measurement functionals
$\mathcal{N}$	Normal distribution
$\boldsymbol{\mathcal{P}}$	Matrix to be inverted in representer method (3.50)
α_1, α_2	Coefficients used in Chap. 3
α_{ij}	Coefficient used in (16.1
$\boldsymbol{\alpha}(\boldsymbol{x})$	Poorly known model parameters to be estimated, Chap. 7
$\boldsymbol{\alpha}'(\boldsymbol{x})$	Errors in model parameters, Chaps. 7 and 10
β	Coefficient in Lorenz equations (6.7); constant (also β_{ini} and β_{mes}), Sect. 11.4
$\delta\psi$	Variation of ψ
ϵ	Real measurement error, Chap. 3
$\boldsymbol{\epsilon}_{\mathcal{M}}$	Representation errors in measurement operator, Sect. 5.2.3
$\boldsymbol{\epsilon}_d$	Actual measurement errors, Sect. 5.2.3
$\boldsymbol{\epsilon}$	Random or real measurement errors, Sect. 5.2.3
$\boldsymbol{\eta}$	Smoothing operators used in gradient method (6.19)
γ	Constant used in smoothing norm analysis (6.32); step length (6.22)
$\gamma(\boldsymbol{h})$	Variogram (11.1)
$\kappa_2(\boldsymbol{A})$	Condition number, Chap. 11
κ_l	Wave number in x direction, Chap. 11
λ_p	Wave number in y direction, Chap. 11
λ	Eigenvalue (13.8); scalar adjoint variable (5.37)

$\boldsymbol{\lambda}$	Vector adjoint variable
$\boldsymbol{\Lambda}$	Diagonal matrix of eigenvalues from eigenvalue decomposition
$\boldsymbol{\Lambda}_1$	Diagonal matrix of eigenvalues from eigenvalue decomposition (14.26)
$\boldsymbol{\Lambda}_p$	Diagonal matrix of eigenvalues from eigenvalue decomposition (14.14)
μ	Sample mean (2.20)
$\mu(\boldsymbol{x})$	Sample mean (2.23)
ω	Frequency variable (6.33)
ω_i	Unit variance noise process (8.10)
$\boldsymbol{\Omega}$	Error covariance of ω_i noise process (8.12)
ϕ	Scalar variable, Chap. 2
$\phi(\boldsymbol{x})$	Porosity in Chap. 16
$\phi_{l,p}$	Uniform random number (11.10)
Φ	Random scalar variable, Chap. 2
π	3.1415927
$\boldsymbol{\pi}$	Momentum variable used in hybrid Monte Carlo algorithm (6.25)
ψ	Scalar state variable (has covariance $C_{\psi\psi}$)
$\psi(\boldsymbol{x})$	Scalar state variable field (has error covariance $\boldsymbol{C}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$)
$\widehat{\psi}(\boldsymbol{k})$	Fourier transform of $\psi(\boldsymbol{x})$, Chap. 11
$\boldsymbol{\psi}$	Vector state variable, e.g. from a discretized $\psi(\boldsymbol{x})$ (has error covariance $\boldsymbol{C}_{\psi\psi}$)
$\boldsymbol{\psi}(\boldsymbol{x})$	Vector of scalar state variables (has error covariance $\boldsymbol{C}_{\psi\psi}(\boldsymbol{x}_1, \boldsymbol{x}_2)$)
Ψ	Random scalar variable, Chap. 2
Ψ_0	Best guess initial condition for dynamical scalar models, may be function of $\boldsymbol{x}$
$\boldsymbol{\psi}(\boldsymbol{x})$	Vector of fields, sometimes written just $\boldsymbol{\psi}$
$\boldsymbol{\psi}_0$	Estimate of initial condition $\boldsymbol{\Psi}_0$, Chap. 7
$\boldsymbol{\psi}_b$	Estimate of boundary condition $\boldsymbol{\Psi}_b$, Chap. 7
$\boldsymbol{\Psi}$	Combined state vector, Chap. 10
$\boldsymbol{\Psi}_0$	Best guess initial condition
$\boldsymbol{\Psi}_b$	Best guess boundary condition, Chap. 7
ρ	Correlation parameter (11.33); coefficient in Lorenz equations (6.6)
$\boldsymbol{\Sigma}$	Matrix of singular values from the singular value decomposition, Sect. 11.4 and (14.68)

$\boldsymbol{\Sigma}_0$	Matrix of singular values from the singular value decomposition (14.19)
$\boldsymbol{\Sigma}_1$	Matrix of singular values from the singular value decomposition (14.52)
σ	Standard deviation defined in Chap. 2; used as coefficient in Lorenz equations (6.5); singular values, Sect. 11.4, Chap. 13 and Chap. 14
τ	De-correlation time (12.1)
θ	Pseudo temperature variable used in simulated annealing algorithm; rotation of principal direction (11.11)
$\boldsymbol{\Theta}$	Random rotation used in SQRT analysis scheme, Chap. 13
ξ	Random number used in Metropolis algorithm
$\boldsymbol{\xi}$	Coordinate running over boundary of model domain
$\mathbf{1}_N$	$N \times N$ matrix with all elements equal to 1
$\delta()$	Dirac delta function (3.24)
$\boldsymbol{\delta}_{\psi_i}$	Vector used to extract a component of the state vector, Chap. 7
$E[]$	Expectation operator
$\mathcal{O}()$	Order of magnitude function
$\Re$	Space of real numbers, e.g. $\Re^{n\times m}$ for a real $n \times m$ matrix