UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA


# NONLINEAR STATISTICS AND DYNAMICS OF ATMOSPHERIC PREDICTABILITY AND DOWNSCALING

*Rui Alexandre Pita Perdigão*

DOUTORAMENTO EM FÍSICA

2010

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA


# NONLINEAR STATISTICS AND DYNAMICS OF ATMOSPHERIC PREDICTABILITY AND DOWNSCALING

**Rui Alexandre Pita Perdigão**

DOUTORAMENTO EM FÍSICA

Tese orientada por:
Professor Doutor **Carlos Alberto Leitão Pires**
Doutor **João Paulo da Costa Campos Teixeira**

2010

*Dedicado à minha família*

# Acknowledgements

*To begin with, I would like to express my deepest appreciation for the fundamental support provided by my thesis supervisor at the Universidade de Lisboa, Prof. Doctor Carlos Alberto Leitão Pires. In fact, as an ever-present guardian angel, Prof. Carlos Pires accompanied every stage of the research, even during my lengthy stay abroad.*

*I would also like to vehemently thank my supervisor at the NATO Undersea Research Centre (NURC), Doctor João Paulo da Costa Campos Teixeira, for his hospitality and support, during the development of my research work at NURC under his supervision.*

*A word of utmost gratitude is due to Prof. Doctor Catherine Nicolis and Doctor Stéphane Vannitsem, who openly welcomed me and supervised my research work at the Royal Meteorological Institute of Belgium (RMI). Even though I was at RMI as a visitor, I was provided with excellent working conditions, namely an extremely inspiring scientific and human environment, along with "royal" infrastructures.*

*My appreciation is extensive to all those whom I had the honour to interact with on a regular basis at the various institutions where the research work unfolded.*

*At the institutional level, I begin by acknowledging the crucial financial support, throughout the whole doctoral period, from the Portuguese*

*Foundation for Science and Technology (Fundação para a Ciência e Tecnologia, FCT), through the doctoral grant SFRH/BD/21471/2005, funded by the European Commission and the Portuguese Ministry for Science, Technology and Higher Studies (Ministério da Ciência, Tecnologia e Ensino Superior). The thesis was elaborated under the auspices of the FCT-funded project PPCDT/CTE-ATM/62475/2004, designated by PREDATOR – Seasonal Predictability and Downscaling over the Euro-Atlantic Region.*

*Furthermore, I endoss my gratitude to the University of Lisbon (Universidade de Lisboa) in general and the Instituto Dom Luiz (IDL) in particular, for being my academic home and behaving as such at all times, even during the years spent performing research work abroad.*

*Even though I have met just a handful of authors referred to throughout the thesis, I hereby express my gratitude to all of them for their remarkable and inspiring documents. In fact, the edifice of my work stands on the shoulders of giants.*

*I devote my final, kindest words of appreciation to my family. To them my whole self is owed... To them my whole work is dedicated...*

# Abstract

This thesis addresses pertinent challenges underneath the estimation of the state of the system at a set of circumstances $\mathcal{B}$ given a set of conditions $\mathcal{A}$. In particular, two main problems are considered: on one hand, that of atmospheric downscaling; on the other hand, that of atmospheric predictability.

For that purpose, novel methods in nonlinear statistics and dynamics are developed and implemented in the aforementioned contexts.

As far as the atmospheric downscaling is concerned, nonlinear statistical features are assessed within the statistical response of the monthly winter precipitation to the North Atlantic Oscillation (NAO) over the North Atlantic European Region. For that purpose, two major methodologies are developed and implemented.

On one hand, a diagnostic measure is built in order to measure the asymmetric part of an estimated variable's response to its predictor, a measure undetected by linear correlation. As a practical application, that variable is chosen to be the precipitation and its predictor the NAO regime (NAO+ and NAO-). The asymmetric features are then used to define an asymmetry-based measure of non-Gaussianity.

On the other hand, an information-theoretical assessment on non-Gaussianity is performed and a corresponding measure of information-theoretical correlation − also transcending the limited scope of linear correlation − defined and applied to the aforementioned downscaling application.

As main results, the proposed estimators for asymmetry and non-Gaussianity are proven to be consistent in their domain of validity. The statistical response of monthly precipitation to NAO is seen to be asymmetric and non-Gaussian. New relevant features are brought out as a result of the application of the proposed nonlinear statistical methods.

As far as atmospheric predictability is concerned, a systematic formalism is derived for the dynamics of prediction errors under the combined influence of initial-condition and model-related errors. Its analytical results are confronted with those from numerical experiments and some generic features for the error dynamics are brought out and connected with intrinsic system properties.

Analytical and numerical results are seen to agree within the domain of validity of the analytical formulation: that of small perturbations in the short-to-intermediate time regime.

All in all, the proposed formulation for assessing the error dynamics is seen to allow for an evaluation of the dynamics of prediction errors without the need to actually run the numerical model under analysis. Moreover, it reveals new generic model-independent features of the error dynamics, under the combined influence of initial condition and model-related errors, not only for low-order but also to systems exhibiting a higher order of complexity.

# Resumo

A presente tese analisa desafios pertinentes subjacentes à estimação do estado de um sistema num conjunto de circunstâncias $\mathcal{B}$ dado um conjunto de condições $\mathcal{A}$. Em particular, dois problemas centrais são considerados: por um lado, o do "downscaling" atmosférico; por outro, o da predictabilidade atmosférica.

Para tal, novos métodos em estatística e dinâmica não lineares são desenvolvidos e implementados nos contextos acima referidos.

No que diz respeito ao "downscaling" atmosférico, são analisadas características estatísticas não-lineares da resposta estatística da precipitação de Inverno à Oscilação do Atlântico Norte (NAO) sobre a região Euro-Atlântica. Para tal, são desenvolvidas e implementadas duas metodologias fundamentais:

Por um lado, é elaborada uma medida de diagnóstico por forma a medir a parte assimétrica da resposta de uma variável estimada ao seu predictor, uma medida não detectada pela correlação linear. Como aplicação prática, a variável escolhida é a precipitação e o seu predictor o regime NAO (NAO+ ou NAO-). As características assimétricas são então utilizadas para definir uma medida de não-Gaussianidade baseada na assimetria.

Por outro lado, a não-Gaussianidade é abordada na perspectiva da teoria da informação, sendo definida uma medida de correlação baseada no conceito de informação mútua, transcendendo a correlação linear.

Esta medida é então aplicada ao problema de "downscaling" atrás referido.

Como resultados principais, é provada a consistência, no respectivo domínio de validade, de estimadores propostos para a assimetria e não-Gaussianidade. Para além disso, é verificado que a resposta estatística da precipitação mensal à NAO é assimétrica e não-Gaussiana. São reveladas novas características relevantes dessa resposta estatística em resultado da aplicação dos métodos estatísticos não-lineares propostos.

No que diz respeito à predictabilidade atmosférica, é elaborado um formalismo sistemático para a dinâmica dos erros de previsão sob a influência combinada de erros nas condições iniciais e no próprio modelo. Os seus resultados analíticos são confrontados com os de experiências numéricas e são reveladas algumas características genéricas, bem como a sua relação com propriedades intrínsecas do sistema.

É verificada a concordância entre os resultados analíticos e numéricos no domínio de validade da formulação analítica: o das pequenas perturbações no regime de prazos curtos a intermédios.

No geral, verifica-se que a formulação proposta para analisar a dinâmica do erro permite uma avaliação da dinâmica dos erros de previsão sem a necessidade de executar o modelo numérico em análise. Para além disso, revela novas características independentes do modelo em uso, sob a influência combinada de erros nas condições iniciais e no modelo, não só para sistemas de baixa ordem mas também para sistemas com mais elevado grau de complexidade.

# Keywords

Nonlinear Statistics; Atmospheric Downscaling;
Non-Gaussianity; Information Theory.
Nonlinear Dynamics; Atmospheric Predictability;
Error Dynamics; Chaos.

# Palavras-chave

Estatística não-linear; "Downscaling" Atmosférico;
Não-Gaussianidade; Teoria da Informação.
Dinâmica não-linear; Predictabilidade Atmosférica;
Dinâmica do Erro; Caos.

# Contents

# List of Figures

# Nomenclature

### Roman Symbols

| | |
|---|---|
| $\mathbf{A}$ | Self-adjoint operator defined semi-positive |
| $\mathbf{B}$ | Positive-defined operator |
| $c_+$ | Positive side correlation, i.e. correlation between two variables $X$ and $Y$ conditioned to $X$ being above its median $M_X$ |
| $c_-$ | Negative side correlation, i.e. correlation between two variables $X$ and $Y$ conditioned to $X$ being above its median $M_X$ |
| $c_g$ | Gaussian Correlation |
| $c_{\text{inf}}$ | Information Correlation |
| $c_M$ | Interset Covariance or Central Correlation |
| $\mathbf{C}$ | Covariance operator |
| $c(X, Y)$ | Pearson Correlation between variables $X$ and $Y$ |
| $\mathbf{d}$ | Integral term due to model error or drift |
| $\mathbf{D}$ | Dissipative term in the QG model |
| $D$ | Relative Entropy or Kullback-Leibler Distance |

| | |
|---|---|
| $\mathbf{E}$ | Ekman term in the QG model |
| exp | Exponential function, $\exp(a) = e^a$ |
| $E(.)$ | Expectancy |
| $E_{Z_+}$ | Conditional Expectancy of generic variable Z given data above the median |
| $E_{Z_-}$ | Conditional Expectancy of generic variable Z given data below the median |
| $\mathbf{F}$ | Smooth function representing a dynamical system |
| $f$ | generic nonlinear function |
| $\mathbf{f}$ | Model formulation, smooth nonlinear function with continuous derivatives over the whole domain |
| $FH(h)$ | Dependency of the drag coefficient on height $h$ |
| $g$ | generic nonlinear function |
| $\mathcal{G}$ | Gaussian Anamorphosis |
| $h$ | Height |
| $H(.)$ | Statistical Entropy |
| $h(.)$ | Differential Entropy |
| $H_i$ | Hermite orthogonal polynomials of order $i$ |
| $\mathcal{H}$ | Hilbert space |
| $\mathsf{H_o}$ | Null hypothesis |
| $\mathbf{I}$ | Identity matrix |

$I_g$ Mutual Information within the bivariate Gaussian distribution with correlation $c$

$I_{ng}$ Mutual Information term associated to non-Gaussianity

$I_{ng(EF)}$ Estimation of the Mutual Information term associated to non-Gaussianity from the Edgeworth expansion method

$I_{ng(EI)}$ Mutual Information term associated to non-Gaussianity obtained through an estimated integral

$I_{ng(MI)}$ Estimation of the Mutual Information term associated to non-Gaussianity by Gaussian quadrature

$\Im$ Imaginary part of the term onto which it operates

$\mathcal{I}$ Interval within $\mathbb{R}$

$I$ Mutual Information

$\mathbf{J}$ Jacobian of the model relative to the state vector

$\mathsf{J}$ Negentropy

$\mathcal{J}_c$ Asymmetry measure

$\mathbf{J}_l$ $l$-th order time derivative of $\mathbf{J}$

$\mathbf{J}_s$ Jacobian of the 2D field over the sphere

$\mathsf{J}(V, W)$ Joint Negentropy of V and W

$\mathbf{K_J}$ Operator defined by $\mathbf{M}^{-1}\mathbf{J}\mathbf{M}$

$\delta_{ij}^{\mathrm{kr}}$ Kronecker Delta

$k^{(p,q)}$ Cumulants of order $p$ in $V$ and of order $q$ in $W$

$kur$ Kurtosis

$l$ Positive adimensional quantity increasing with the truncation order (except when used as a generic subscript)

ln Natural logarithm, i.e. $\log_e(.)$

$LS(\lambda, \phi)$ Fraction of land within a grid box n the QG model

$\mathbf{M}$ Fundamental (resolvent) matrix

$M$ Median

$n$ Generic non-negative integer, $n \in \mathbb{N}_0$

$N(\cdot)$ Norm

$N_{df}$ Number of temporal degrees of freedom

$n_{\text{eq}}$ Number of i.i.d. variables

$NL$ Number of levels considered in the spectral mode

$NT$ Truncation of a spectral model

$\mathbf{O}$ Generalised self-adjoint Oseledec operator

$\mathbf{O}_l$ $l$-th order time derivative of $\mathbf{O}$ at forecast time $\tau = 0$

$\mathcal{O}(i)$ i-th order term

$\mathcal{P}$ Padé Approximant

$P_{neg}$ Integral of the EDG-PDF in the domain of negative values of the estimated truncated density

$P_{pos}$ Integral of the EDG-PDF in the domain of positive values of the estimated truncated density

$p$ Probability function

$\mathbf{p}$ Tendency error

$p_T$ Pressure on top of the troposphere

$p_X(x)$ Probability that $X = x$

$P(X = x)$ Probability that $X = x$

$\mathbf{q}$ Potential Vorticity

$Q$ Generic Quantile

$\Re$ Real part of the term onto which it operates

$\mathbb{R}$ Real set

$r_E$ Average radius of the Earth

$rhs$ Right hand side

$R_j$ Rossby radii of deformation in the QG model

$\mathbf{S}$ Forcing term in the QG model

$\text{sgn}(\cdot)$ Signal of a function

$t$ Time

$T_i$ Coefficient of the $\tau^i$ term in the Taylor expansion in $\tau$

$t_+$ Test Positive Side Correlation

| | |
|---|---|
| $t_-$ | Test Negative Side Correlation |
| $t_c$ | Crossover time, for which the magnitudes of the contributions of initial and model errors match each other |
| $t_M$ | Test Central Correlation |
| $t_m$ | Time at which the mean square errors attain their minimum value |
| $\mathsf{T}$ | Transpose (as superscript) |
| $\mathbf{u}$ | Error in the state vector |
| $\mathbf{u}_0$ | Error in the initial conditions, i.e. of the state vector at time $t = 0$ |
| $\mathbf{u}_I$ | Integral term due to error in the initial conditions |
| $V$ | Standard variable assumed as arithmetic average of $n_{\text{eq}}$ independent and identically distributed (i.i.d.) variables. |
| $var$ | Variance |
| $W$ | Standard variable assumed as arithmetic average of $n_{\text{eq}}$ independent and identically distributed (i.i.d.) variables. |
| $w_i$ | Weighing term representing the metric defined on an orthogonal basis |
| $\mathbf{x}$ | State vector of a dynamical system |
| $X$ | Generic variable with values $x \in \mathbb{R}$ |
| $\mathbf{X}$ | Generic $n$-dimensional variable |
| $\mathbf{X}_G$ | Gaussian random variable with the same covariance matrix as that of $\mathbf{X}$ |
| $\mathbf{x}_N$ | Reference (Nature) state vector |

$Y$ Generic variable with values $y \in \mathbb{R}$

$Y_r$ Standardised residue of the linear prediction of $Y$ from $X$

$Z$ Generic variable with values $z \in \mathbb{R}$

$\mathbf{Z}$ Generic operator

**Greek Symbols**

$\epsilon$ Isotropically distributed errors in the initial conditions

$\epsilon_{\Delta t}$ One-step integration error given the time-step $\Delta t$

$\epsilon_s$ Error in the direction of the stable modes

$\epsilon_u$ Error in the direction of the unstable modes

$\eta_{\alpha t''}$ Eigenvalues of $\mathbf{O}$ (real-valued)

$\Gamma$ Actual phase trajectory

$\Gamma_R$ Reference phase trajectory

$\lambda_{\alpha t''}$ Finite Lyapunov exponent between times $t''$ and $\tau$

$\Lambda_{\alpha,t''}$ Set of finite Lyapunov vectors optimised between times $t''$ and $\tau$, $\alpha$ being the index referring to each vector

$\lambda_{\mathrm{I}}$ Sensitivity to the initial conditions

$\lambda_{\mathbf{x}_0}$ Lyapunov exponents

$\mu$ Control parameters of a dynamical system

$\mu_N$ Reference control parameters of a dynamical system

| | |
|---|---|
| $\nu$ | Truncated fitting polynomial vanishing if the joint PDF is Gaussian |
| $\Omega$ | Alphabet of a random variable |
| $\mathbf{\Phi}$ | Jacobian of the model ($\mathbf{f}$) relative to the control parameters |
| $\mathbf{\Phi}_l$ | $l$-th order time derivative of $\mathbf{\Phi}$ |
| $\Phi$ | Cumulative standard Gaussian distribution function |
| $\phi_\alpha$ | Orthonormalisation vector |
| $\varphi$ | Standard Gaussian PDF |
| $\varphi_V(v)$ | Marginal Probability Distribution Function of V |
| $\mathbf{\Pi}$ | Projection operator over a sub-space of the basis functions $\phi_\alpha$ |
| $\mathbf{\Psi}$ | Streamfunction |
| $\rho$ | Probability Density Function |
| $\rho_{V,W}(v,w)$ | Joint Probability Distribution Function of $V$ and $W$ |
| $\sigma$ | Standard Deviation |
| $\sigma_d$ | Rate of decay of stable modes |
| $\sigma_{Z_+}$ | Conditional Standard Deviation of the generic variable $A$ given data above the median of Z |
| $\sigma_{Z_-}$ | Conditional Standard Deviation of the generic variable $A$ given data below the median of Z |
| $\tau$ | Forecasting time |

$\tau_{DIF}$ Diffusive time scale in the QG model

$\tau_R$ Radiative time scale in the QG model

$\xi_p$ Potential Vorticity in the pressure coordinate system

**Other Symbols**

$*$ Complex conjugate (as superscript)

$[\![\cdots]\!]$ Phase space average over the attractor

$\delta\mathbf{f}$ Deviation of the model formulation relative to the true dynamics it indends to represent

$\delta\mu$ Error in the control parameters

$\Delta p$ Pressure difference between levels in the QG model

$\delta\mathbf{\Psi}$ Perturbation in the Streamfunction

$\delta\mathbf{q}$ Perturbation in the Potential Vorticity

$\Delta t$ Time-step

$\delta$SST Sea surface temperature anomaly with respect to the climatology

$\langle\cdot,\cdot\rangle$ Inner product

$\nabla$ 2D gradient operator over the sphere

$\overline{\cdots}$ Average over an ensemble of initial conditions

$\rightarrow$ Tends towards

$\sphericalangle(\mathbf{x},\mathbf{y})$ Spherical angle between $\mathbf{x}$ and $\mathbf{y}$

## NOMENCLATURE

### Acronyms

2D Bidimensional

4D-Var Four-dimensional variational data assimilation

ATL Central Atlantic

BAL Balearic Islands

BT Breeding time

DJF December to February

ECMWF European Centre for Medium-Range Weather Forecasts

EDG Edgeworth expansion

EDG-PDF Probability Distribution Function estimated through an Edgeworth expansion

EUS Eastern United States

FCT Fundação para a Ciência e Tecnologia

FS Fraction of statistically significant area

FS-MCG Fraction of statistically significant area for the MCG method

FS-MCR Fraction of statistically significant area for the MCR method

GA Gaussian Anamorphosis

GRE Greenland

GS Greenland-Scandinavian dipole

ICA Independent Component Analysis

IDL Instituto Dom Luiz

i.i.d. Independent and identically distributed

KLD Kullback-Leibler Distance

L3 Three-level

MCG Monte Carlo technique through generation of Gaussian noise

MCR Monte Carlo technique through random reordering of working series

MI Mutual Information

NAE North Atlantic European Region

NAO+ Positive phase of the North Atlantic Oscillation

NAO- Negative phase of the North Atlantic Oscillation

NAO North Atlantic Oscillation

NATO North Atlantic Treaty Organisation

NCEP/NCAR National Centers for Environmental Prediction − National Center for Atmospheric Research

NURC NATO Undersea Research Centre

PDF Probability Density Function

QG Quasi-Geostrophic

RDG Atlantic anticyclonic ridge

RMI Royal Meteorological Institute of Belgium

RUS Russia

SCO Northwest Scotland

SLP Sea level pressure

SR Significant Regions

SST Sea surface temperature

Y/N Yes or no

# Chapter 1

# Introduction

A fundamental challenge driving scientific work consists on finding a solution for a problem given a set of available factors. For that purpose, adequate methodologies have to be developed and implemented that somehow allow for the factors to weigh in, providing input that will help drive the problem towards a solution.

In physical sciences, given the state of a system under a certain set of circumstances (say, $\mathcal{A}$), quantitative methods can be elaborated and used for its estimation at another set of circumstances, say $\mathcal{B}$. By *circumstances* it is meant, for instance, a series of equations, inequalities, probabilities of events, all expressed in terms of the set of variables characterising the state of the system.

When $\mathcal{B}$ positions the system at a future state relative to $\mathcal{A}$, its estimation is commonly denoted as *prediction*. The aforementioned quantitative methods take the form of a model representation also denoted as *prediction system*.

There are, however, major obstacles in achieving reasonable skill in such estimation. On one hand, there are uncertainties in the input factors, stemming from the limited accuracy of their measurement and assessment. In highly unstable systems, even the slightest inaccuracy way below the resolution limits of the most advanced measurement equipment at hand can ultimately contribute to an utter

loss of predictability. This rather surprising fact is a core signature of *chaos*, in the sense of sensitive dependence to prior conditions (Lorenz, 1963).

On the other hand, the prediction system itself does not encompass the full extent of the physics involved in the process. In fact, there are physical relationships and additional factors not taken into account, such as sub-grid scale processes and errors in the model formulation itself. This, too, can seriously undermine the predictive ability.

These issues are of utmost relevance for the atmospheric sciences. In fact, atmospheric prediction systems are known to suffer from fundamental uncertainties associated with their sensitivity to the initial conditions and with the inaccuracy in the model representation. The idea that error in weather forecasting comes from both these sources of uncertainty can be traced back to as early as Bjerknes et al. (1911).

Initial condition uncertainties and their role in affecting the predictability have been widely studied for a long time, e.g. as early as Thompson (1957), Lorenz (1963). This error source was addressed both in the context of dynamical systems predictability (e.g. Lorenz, 1963; Smith et al., 1999) and ensemble forecasting (e.g. Nicolis, 1992; Toth and Kalnay, 1993; Nicolis et al., 1995, Buizza and Palmer, 1998).

More recently model errors have been addressed both at the level of numerical forecast models (e.g. Dalcher and Kalnay, 1987; Reynolds et al., 1994) and their generic features (Vannitsem and Toth, 2002; Nicolis, 2003, 2004). Moreover, their connections with such error sources in numerical truncation (e.g. Teixeira et al., 2007) sub-grid scale parameterisations (e.g. Teixeira and Reynolds, 2008; Buizza et al., 1999) and boundary conditions (Nicolis, 2007) have also been assessed.

Ordinarily, initial condition and model errors are studied separately. However, by solely focusing on one of these error types only limited improvement in prediction can be attained (Lorenz, 2005). Moreover, the combination of these factors is known to place severe limits on our ability to predict future states of even thoroughly-understood systems. It is thus of interest to study the dynamics in the presence of both error types in a systematic manner. This was one of the main objectives of the present thesis.

For this purpose, methods in nonlinear dynamics were developed, tested and implemented that allowed for generic and relevant case-by-case features to be brought out in systems relevant to the physical sciences in general and the atmospheric sciences in particular. In this context, the theoretical developments and their applications to illustrative systems are presented in chapter 3, followed by an implementation in a Quasi-Geostrophic (QG) model (Marshall and Molteni, 1993) in chapter 4.

Chapter 3 begins with an overview on dynamical systems and predictability, followed by the general framework on the dynamics of model error in a generic, infinite-dimensional functional space. The formulation is then simplified to finite-dimensional systems, envisioning the practical applications to be presented afterwards. Some generic, model-independent features are brought out, such as the role of the mechanisms of error transfer between a particular initial direction to components along other directions, the existence of a minimum in the error evolution and the relative importance of the IC errors and model errors in the global evolution. These theoretical results are then tested and illustrated in low-order systems exhibiting bistable, saddle-point and chaotic behaviour.

In chapter 4 the generic analytical approach to the error dynamics introduced in 3 is applied to the QG model and validated with numerical experiments. While

some generic features are identified that come in agreement with those seen in lower-order systems, further properties of physical relevance, stemming from the higher-order of complexity of the system, are also unveiled.

A rather different approach to the error dynamics in a QG model from that of the preceding chapters is then presented in chapter 5. There, a simple model of error evolution is presented by considering the relative error growth rates in the stable and unstable directions, with the addition, in the latter case, of a nonlinear saturation term that allows for the error evolution to be modelled on the long term until saturation.

Another major objective brings us back to challenges underneath the estimation of the state of the system at $\mathcal{B}$ given $\mathcal{A}$, though in a different context: that of *atmospheric downscaling*, in which local or regional-scale atmospheric features are inferred from large-scale atmospheric predictor variables.

In this context, while a dynamical approach would be a natural favourite from a physical point of view, there is no conclusive evidence that dynamics should be favoured over statistics. In fact, by comparing these approaches overall similar skills can be obtained, with statistical methods coming out on top when it comes to such practical aspects as computational effort (Murphy, 1999; Kidson and Thompson, 1998). Anyway, regardless of the methodology one must never lose sight of the underlying physical phenomenology.

The scientific literature exhibits a wealth of relevant work on atmospheric downscaling, most notably on extracting such features as regional circulation, temperature and precipitation regimes from large-scale patterns in the climatic system such as the North Atlantic Oscillation (NAO) (e.g. Trigo et al., 2002, 2004; Pires and Perdigão, 2007) or the El Niño / Southern Oscillation (ENSO) (e.g. Ropelewski and Halpert, 1986; Pozo-Vázquez et al., 2001), just to list a handful

of examples. It is thus of considerable relevance to provide the community with statistical methods that significantly increase the quality of the estimation, along with relevant examples of application.

In the present thesis, nonlinear statistical features are assessed within the statistical response of the monthly winter (December to February) precipitation to the North-Atlantic Oscillation (NAO) over the North Atlantic European Region (NAE). For that purpose, two major methodologies are developed and implemented:

On one hand, a diagnostic measure is built in order to measure the asymmetric part of an estimated variable's response to its predictor, a measure undetected by linear correlation. As a practical application, that variable is chosen to be the precipitation and a certain predictor: here, the NAO index, splitting the NAO+ and NAO- regimes. The asymmetry is based on the different anti-symmetric responses to NAO+ and NAO-. The asymmetric features are then used to define an asymmetry-based measure of non-Gaussianity.

On the other hand, an information-theoretical assessment on non-Gaussianity is also performed and a corresponding measure of information-theoretical correlation − also transcending the limited scope of linear correlation − defined and applied to the aforementioned downscaling application.

This key objective encompassing nonlinear statistics and downscaling is thoroughly addressed in chapter 2, as it chronologically preceded the work on nonlinear dynamics and predictability (chapters 3 − 5), from which it is technically independent.

While each chapter has its own conclusions, the thesis is concluded with a closure chapter with final, concluding remarks and future prospects (chapter 6).

# Chapter 2

# Nonlinear statistical downscaling

## 2.1 Introduction

The characterisation of large-scale extratropical atmospheric variability is a matter of open debate. A paradigmatic view is to regard the low-frequency projection of the atmospheric attractor as a superposition of dynamical regimes (Dole, 1983).

It has been suggested and argued that atmospheric large scale variability at the monthly time scale over the North Atlantic European Region (NAE) is characterised by transition or permanence among four dynamical winter regimes: the positive phase of the North Atlantic Oscillation (NAO+), its negative phase (NAO-), the Greenland-Scandinavian dipole (GS) and the Atlantic anticyclonic ridge (RDG) (Cassou et al., 2004). These regimes, obtained by cluster analysis, are not exactly organised symmetrically around the climatology, in terms of symmetry relative to axis of uncorrelated variables. In fact, they are asymmetric, as it normally occurs in chaotic nonlinearly forced systems (Palmer, 1999). Moreover, composites of anomaly surface forcings [e.g. sea surface temperature (SST)], computed for opposite regimes, show some degree of asymmetry (Robinson et al., 2003; Wu and Hsieh 2004). In terms of conditional expectan-

cies: $E(\delta\text{SST}|\text{NAO+}) \neq -E(\delta\text{SST}|\text{NAO-})$, where $\delta$SST is the SST anomaly with respect to the climatology. This has been shown in particular for the tripole Atlantic and Pacific SST forcings of the NAO+ and the NAO- quasi anti-symmetric regimes (Cassou et al., 2004). The link between average surface climatic conditions (e.g. large-scale precipitation and surface temperature) and middle-tropospheric regimes can be decomposed in terms of a monotonic linear influence and nonlinear terms responsible for possible asymmetric responses. A linear response of the mean square error (MSE) to a forcing is a necessary, though not sufficient condition of joint Gaussianity between the response and forcing. Therefore, a nonlinear MSE-response implies non-Gaussianity. As a consequence, the joint probability density functions (PDFs) of large-scale indexes and climatic variables can express some degree of asymmetry and non-Gaussianity. Some studies corroborate this fact, e.g. the nonlinear, asymmetric response of the surface temperature over Europe for symmetric quantiles of the North Atlantic Oscillation (NAO) index (Pozo-Vásquez et al., 2001; Trigo and Palutikof, 1999). Another example is the asymmetry of dry and wet Self-Organizing Maps (Cavazos, 2000) and their different correlations with Arctic Oscillation and NAO indices. Another study hypothesises the asymmetric response of the Indian Ocean precipitation to the NAO (Sapiano and Arkin personal communication, 2005).

Motivated by this issue, we hereby bring a contribution to inferring the degree of non-Gaussianity and asymmetry within the statistical response of the monthly winter [December−February (DJF)] precipitation to the NAO over the NAE.

Both the mean influence of NAO and its trend on the monthly winter precipitation over the NAE are well documented (Hurrell et al., 2004). That influence is essentially due to: a) the different tracking of synoptic storms in the presence of NAO+ or NAO- regimes (Rogers, 2002; Hurrell, 1995); b) the enhancement,

for particular regimes, of local systems associated to particular geographical and orographic conditions, for instance for Greenland and Iceland (Serreze et al., 1997).

The linear component of that influence can be assessed through the one point linear correlation map between the monthly precipitation and the NAO monthly index, which is defined as the Lisbon, Portugal, minus Stykkisholmur, Iceland, normalised average monthly sea level pressure (SLP) anomaly or other correlated quantities (Osborn et al., 1999). In practice one begins by computing a quantity $\mathsf{A}$ given by the normalised monthly anomaly at Lisbon minus the normalised monthly anomaly in Iceland. Then the NAO index is obtained by normalising $\mathsf{A}$.

The aforementioned correlation map shows a dipolar structure with extreme positive correlations near 0.6, at south of Iceland, around 60°N latitude, and extreme negative values near $-0.6$ at the North Atlantic basin around 40°N of latitude.

A diagnostic measure is hereby built in order to measure the asymmetric part of the precipitation response to NAO, undetected by the linear correlation. For that purpose, *side* or *asymmetric correlations* between NAO index and monthly precipitation are computed. This essentially consists of evaluating conditional correlations for both the positive and negative NAO regimes, thus revealing possible asymmetric or non-Gaussian precipitation responses to NAO. Asymmetry is only a particular aspect of something more general: non-Gaussianity.

The Edgeworth expansion method is used to evaluate bivariate non-Gaussian probability densities (PDFs). Then, relevant diagnostics are obtained from Information Theory (Shannon, 1948), such as Negentropy and Mutual Information as well as its Gaussian and non-Gaussian counterparts (Kraskov et al., 2004). This method assumes a weak non-Gaussianity scenario. The non-Gaussianity degree

of the joint PDF (NAO, Monthly Precipitation at the point basis) is evaluated through Edgeworth Expansions (Edgeworth, 1905; Comon, 1994), based on the Hermite Polynomials (Abramowitz and Stegun, 1972) and higher-order statistical moments.

PDF evaluation is also possible numerically based in the Maximum Entropy Principle (Pires and Perdigão, 2007; Jaynes, 1982), through the Maximum Likelihood Method (Sivia, 1996) and the Kernels estimation (Silverman, 1986).

Apart from the present document, Information Theory is used in other applications such as predictability studies (DelSole, 2004), forecast evaluation (Roulston and Smith, 2002), Independent Component Analysis (ICA) of climatological data (Aires et al., 2002; Perdigão, 2004), computation of Mutual Information among climatic data (Marwan and Kurths, 2002; Perdigão, 2004).

In the upcoming sections a theoretical overview is presented on the asymmetric correlation (section 2.2) and information-theoretical concepts (section 2.3) relevant for the problem. Methodologies for the estimation of information-theoretical measures of non-Gaussianity are presented in section 2.4. Data and processing methodologies are then presented (section 2.5), followed by results and their analysis (section 2.6). The chapter is then concluded with a final discussion (section 2.7).

The new developments presented in this chapter have been included in the research paper by Pires and Perdigão (2007), published in the Monthly Weather Review of the American Meteorological Society.

## 2.2 Asymmetric correlation approach

### 2.2.1 General correlation measures

Let us begin by recalling the widely used Pearson Product-Moment Correlation Coefficient, also known as Pearson Correlation (Papoulis, 1991). It is a measure of the degree of linear relationship between two variables $X$ and $Y$, given by

$$c(X,Y) = \frac{E[X - E(X)]E[Y - E(Y)]}{\sqrt{var(X)\, var(Y)}} \tag{2.1}$$

where $var(X) = E[(X - E(X))^2]$.

While practical in most applications, this measure does not account for the overall statistical dependence between variables – nonlinear relationships are left unaccounted for. In fact, it is possible to have such dependence even when $c(X,Y) = 0$.

In order to illustrate this fact, consider a uniformly distributed random variable with values $x \in [0, 2\pi[$. Consider now two nonlinear functions, $f(x) = \sin(x)$ and $g(x) = \cos(x)$. These are clearly dependent from each other; in fact, $g(x) = f(|x + \pi/2|)$ for all $x$ in $[0, 2\pi[$. However, it can easily be seen that $c[f(x); g(x)] = 0$, $\forall x \in \mathbb{R}$. One is thus led to the conclusion that null Pearson correlations do not imply statistical independence.

Anyway, one must note that, while not a *sufficient* condition for independence, a null Pearson correlation is a *necessary* condition for that purpose. In fact, $c(X,Y) \neq 0$ implies statistical dependence between the intervening variables.

Even though $c(X,Y) \neq 0$ does not take into account nonlinear relationships between the intervening variables, it is possible to peer into such relationships by considering the Pearson correlation between general nonlinear functions $f(X)$ and $g(Y)$: $c(f(X); g(Y))$. When $f(X) = E(Y|X = x)$ and $g(Y) = E(X|Y = y)$

these relationships are maximised in absolute value. This is a way of using the linear, Pearson correlation to define nonlinear correlations.

A particular example of nonlinear correlation measure is the Spearman or rank correlation (Wilks, 1995), where the nonlinear functions are simply the sampling ranks of $X$ and $Y$ respectively. These ranks measure the degree of monotonic association between both variables.

Another nonlinear correlation measure can be defined by taking, as nonlinear functions $f$ and $g$, the standard Gaussian transformation or *Gaussian Anamorphosis* of the original variables (Wackernagel, 2003). It is hereby denoted as *Gaussian Correlation* and defined as:

$$c_g = c(\mathcal{G}_X; \mathcal{G}_Y) \tag{2.2}$$

where $\mathcal{G}_X$ (equivalently for $Y$) is the *Gaussian Anamorphosis* of $X$, given by:

$$\mathcal{G}_X(x) = \Phi^{-1}\left[\int_{-\infty}^{x} \rho_X(u)\, du\right], \tag{2.3}$$

where $\rho_X(u)$ is the probability density function of $X$ and $\Phi^{-1}$ is the inverse of the cumulative standard Gaussian distribution function.

This transformation is a common data analysis procedure in geostatistical kriging and climatic data analysis (Biau et al., 1999) that ensures that the marginal distributions are standardised Gaussian.

Since the correlation in (2.2) relates two Gaussian variables, $\mathcal{G}_X$ and $\mathcal{G}_Y$, it is an optimal measure of their dependence, as Gaussian variables only exhibit linear statistical dependence. One must, however, note that while the marginal distributions are standard Gaussian, their joint distribution is not necessarily Gaussian. In the particular case for which $X$ and $Y$ are Gaussian, $c_g(X, Y) = c(X, Y)$.

Both rank and Gaussian correlations are nonlinear correlations invariant for the class of monotonous homeomorphisms on $X$ and $Y$ individually (though not necessarily homeomorphisms mixing these variables). Consequently, an advantage of both these measures over the Pearson correlation is the fact that, unlike the latter, the formers are not artificially inflated by the coincidence of large outlier values (Jolliffe and Stephenson, 2003).

### 2.2.2 Asymmetric Gaussian correlations

The correlation measures presented in the previous section take into account the statistical sensitivity of variables to one another in their domain as a whole. However, their statistical relationships within a particular subdomain are not accountable by such general measures. In order to overcome this limitation, we refer to the concept of *Asymmetric Correlation* (Pires and Perdigão, 2007). It consists on the conditional correlation between two standard variables $X, Y$ for a certain interval $\mathcal{I}_X$ of $X$, represented by $c(X, Y|X) \in \mathcal{I}_X$. The asymmetric correlation measure is particularly relevant when the sensitivity of one variable $Y$ to another $X$ is not necessarily the same over every subdomain of $X$.

A simple partition of $X$ can be considered into two complementary intervals, separated by a quantile $Q_X$ of $X$. Within this context, a particular case can be taken for the $Q_X = M_X$, the latter being the median of $X$. In this case, the two partitions are equally-sized.

The correlations between two variables $X$ and $Y$ conditioned to $X$ being below and above its median $M_X$ are then given, respectively, by:

$$\begin{aligned} c_+ &= c(X, Y|X > M_X) \\ c_- &= c(X, Y|X < M_X) \end{aligned} \tag{2.4}$$

and the conditional standard deviations, for each half of the data:

$$\begin{aligned} \sigma_{Z_+} &= [var(Z|X > M_X)]^{1/2} \\ \sigma_{Z_-} &= [var(Z|X < M_X)]^{1/2}, \\ Z &= X \text{ or } Y, \end{aligned} \tag{2.5}$$

where *var* denotes variance.

By considering $X$ and $Y$ to be centred and to have unit-variance, their expectation values within each of $X$'s considered subdomains relate to each other in the following way:

$$\begin{aligned} E(Z|X < M_X) &= -E(Z|X > M_X), \\ Z &= X \text{ or } Y \end{aligned} \tag{2.6}$$

For simplicity of notation, let $E(Z|X < M_X) \equiv E_{Z_-}$ and $E(Z|X > M_X) \equiv E_{Z_+}$. Under the same conditions – $X$ and $Y$ centred and with unit-variance – the following relation is also held:

$$\begin{aligned} 1 &= (E_{Z_+})^2 + \frac{1}{2}\left(\sigma^2_{Z_-} + \sigma^2_{Z_+}\right), \\ Z &= X \text{ or } Y \end{aligned} \tag{2.7}$$

A nonlinear asymmetric relationship between $X$ and $Y$ can be assessed through differences between the asymmetric correlations $c_+$ and $c_-$ in (2.4). For standard variables the correlation equals the covariance, which can be decomposed into *intra-* and *interset* covariances.

By taking the aforementioned partition of $X$ into two halves, the correlation becomes:

$$c(X,Y) = c_M + \frac{c_+ \sigma_{X_+} \sigma_{Y_+}}{2} + \frac{c_- \sigma_{X_-} \sigma_{Y_-}}{2}, \tag{2.8}$$

where

$$c_M = E_{Y_+} E_{X_+} = E_{Y_-} E_{X_-} \tag{2.9}$$

works as the interset covariance of $(X,Y)$, which is proportional to the difference between the asymmetric conditional $Y$ means. Given the constraints (2.6) and

(2.7) it is also clear that $|c_M| \leqslant 1$ and that $|c_M|$ tends to increase for low conditional $Y$ variance. The other two terms of (2.8) are also less than one in absolute value.

### 2.2.3 Measuring non-Gaussianity through asymmetric correlations

The preceding subsections have provided us with the tools to derive a correlation-based diagnostic measure of non-Gaussianity. For that purpose, we begin by comparing the true values of $c_M$, $c_+$ and $c_-$ with those that would be obtained if $X$ and $Y$ were jointly Gaussian. In this case, the following simplifications to the aforementioned quantities hold:

$$\begin{aligned} E_{X_+} &= -E_{X_-} = \sqrt{\alpha} \\ E_{Y_+} &= -E_{Y_-} = \operatorname{sgn}(c)\sqrt{\alpha} \\ \sigma^2_{X_+} &= \sigma^2_{X_-} = 1-\alpha \\ \sigma^2_{Y_+} &= \sigma^2_{Y_-} = 1-\alpha\, c^2 \\ c_M &= \alpha\, c\,, \end{aligned} \tag{2.10}$$

where $\alpha = \frac{2}{\pi}$ and $c \equiv c(X,Y)$.

The asymmetric correlations $c_-$ and $c_+$ are thus equal to each other and given by a nonlinear increasing function of the correlation $c$ (proof in Appendix A):

$$c_- = c_+ = c\sqrt{\frac{1-\alpha}{1-\alpha c^2}}\,. \tag{2.11}$$

By taking into account (2.11) and (2.10) it can be seen that the absolute values of the asymmetric correlations, $|c_+|$ and $|c_-|$, are lower than the absolute value of the correlation, $|c|$. Consequently, under Gaussian conditions the global correlation $c$ is always greated in absolute value than each of the asymmetric correlations, the contributions of which to $c$ in (2.8) are equal to $\frac{1-\alpha}{2}\,c$.

We are now in position to define statistical tests of bi-Gaussianity, hereafter denoted as *test central correlation* ($t_M$), *test positive side correlation* ($t_+$) and *test negative side correlation* ($t_-$). These are directly related to the correlation in the Gaussian case, through:

$$\begin{aligned} t_M(X,Y) &= \frac{c_M}{\alpha} \\ t_+(X,Y) &= \frac{c_+\,\sigma_{X_+}\,\sigma_{Y_+}}{1-\alpha} \\ t_-(X,Y) &= \frac{c_-\,\sigma_{X_-}\,\sigma_{Y_-}}{1-\alpha} \end{aligned} \tag{2.12}$$

Therefore, $c$ is a weighted form of $t_m$, $t_+$ and $t_-$:

$$c = \alpha\, t_m + \frac{1-\alpha}{2}\, t_+ + \frac{1-\alpha}{2}\, t_- \tag{2.13}$$

Differences between the correlation $c$ and the tests in (2.12) are measures of the distribution's asymmetry and thus a diagnostic for non-Gaussianity.

While the existence of such differences ensures non-Gaussianity, their nullity is not a sufficient condition for joint Gaussianity. It is, nevertheless, a necessary one.

In order to get a measure of asymmetry that is independent from the correlation $c$, we consider the pair of uncorrelated variables $X$ and $Y_r$, where $Y_r$ is the standardised residue of the linear prediction of $Y$ from $X$,

$$Y_r = \frac{Y - c\,X}{\sqrt{1-c^2}} \tag{2.14}$$

The correlation (null) between $X$ and $Y_r$ can be rewritten as the following combination of expectancies:

$$0 = c(X,Y_r) = \frac{1}{2}\Big[E(X\,Y_r|X<M_X) + E(X\,Y_r|X>M_X)\Big]. \tag{2.15}$$

Both terms in (2.15) are symmetric, vanish under Gaussian conditions and

lead, after a few algebraic calculations, to a measure $\mathcal{J}_c$ of asymmetry, given by:

$$\begin{aligned}\mathcal{J}_c(X,Y) &= E(X\,Y_r|X > M_X) \\ &= \left[\frac{(1-\alpha)(t_+ - t_-) - c(\sigma^2_{X_+} - \sigma^2_{X_-})}{2\sqrt{1-c^2}}\right]. \end{aligned} \tag{2.16}$$

By subtracting $M_X$ from $X$ and taking the absolute value of the product $XY_r$ in (2.16), it can be inferred that $\mathcal{J}_c$ is proportional to the nonlinear correlation $c(|X - M_X|, Y_r)$, which has, under some conditions, a monotonic relationship with the nonlinear correlation $c(|X - X_M|^2, Y_r)$.

## 2.3 Information-theoretical approach

### 2.3.1 Statistical Entropy

Consider a system with a multitude of a priori available states and an observer facing the challenge of identifying the actual state of the system. Imagine now that the observer is entitled to ask "yes" or "no" questions until that actual state is unequivocal to him. For optimal efficiency and economy, those questions must be non-redundant. In fact, the purpose is to find the state of the system with the least number of questions.

As a simple example, consider a fair coin toss. Two outcomes are possible: head or tail. How many "yes" or "no" (Y/N) questions are needed for someone to find out which of those outcomes have taken place at a particular toss? The answer: Just one. In fact, one can simply ask: "Is it head?" or, equivalently, "is it tail?". In any of those cases alone, a Y/N answer will immediately determine what the actual outcome was.

By regarding the number of available states as possible outcomes of a random variable, the number of Y/N questions needed to identify the actual state will give a measure of the uncertainty of the random variable. Moreover, it will give

a measure of the extent of the minimum string of Y/N answers that has to be considered in order to inequivocally describe the state of the underlying system.

One such measure is Statistical Entropy, Information Entropy, or Entropy for short in the context of Information Theory (Shannon, 1948), of which it is a central concept. In fact, it allows for information to be quantified and digitally encoded, with obvious applications to signal processing, such as data compression and communication.

Mathematically speaking, the Statistical Entropy $H(X)$, of a discrete random variable $X$ with alphabet $\Omega$ and probability function $p_X(x)$ [i.e., $P(X = x)$], can be defined by (Shannon, 1948):

$$H(X) = -\sum_{x \in \Omega} p_X(x) \log_2 p_X(x) \tag{2.17}$$

expressed in *bit*.

For instance, the Statistical Entropy of a fair coin toss is 1 bit, corresponding to the one Y/N question needed to find its actual outcome.

If, instead, we use the natural-based logarithm, ln, the Entropy is given in *nat*.

The convention that $0 \ln 0 = 0$ is assumed by continuity, since $p_X(x) \ln p_X(x) \to 0$ as $p_X(x) \to 0$ . Hence, adding terms of zero probability does not change the Entropy.

Should $\mathbf{X}$ be an $n$-dimensional random variable, (2.17) also refers to the joint Entropy of its components $X_1, \cdots, X_n$, which can be rewritten as

$$H(\mathbf{X}) = -\sum_{x_1} \cdots \sum_{x_n} p_X(x_1, \cdots, x_n) \log_2[p_X(x_1, \cdots, x_n)] \tag{2.18}$$

where $p_X(x_i) = P(X_i = x_i)$, or, equivalently, as

$$H(X_1, \cdots, X_n) = \sum_{i=1}^{n} H\left(X_i | X_{i-1}, \cdots, X_1\right). \tag{2.19}$$

This formula allows us to write the total Entropy as a sum of conditional Entropies, given by:

$$H(X_i|X_j) = -\sum_{x_j,x_i} p(x_j, x_i) \log_2 p(x_i|x_j) \tag{2.20}$$

If the conditional probability models are available, then we can each term in the sum can be sequentially computed.

## 2.3.2 Differential Entropy

### 2.3.2.1 Definition

The Differential Entropy $h(X)$ of a continuous random variable $X$ with values $x$ and probability density function $\rho(x)$ is defined as (Cover and Thomas, 1991):

$$h(X) = -\int_S \rho(x) \ln \rho(x)\, dx\,, \tag{2.21}$$

where S is the support set of $\rho(x)$, i.e., the set where $\rho(x) > 0$. As written in the previous equation (with natural logarithm), the Differential Entropy comes in *nat*.

If, on the above definition, $X$ is replaced by an $n$-dimensional random variable $\mathbf{X}$ with components $X_i$, $i = 1, \cdots, n$, (2.21) shall define the joint Differential Entropy of those components, i.e., of the scalar random variables $X_1, \cdots, X_n$. Equivalently, we can say that the Differential Entropy of a set $X_1, \cdots, X_n$ of random variables with probability density function $\rho(x_1, \cdots, x_n)$ is defined as:

$$h(X_1, \cdots, X_N) = -\int \rho(x_1, \cdots, x_n) \ln \rho(x_1, \cdots, x_n)\, dx_1\, dx_n. \tag{2.22}$$

### 2.3.2.2 Relation to Discrete Statistical Entropy

Let us consider a continuous random variable $X$ with probability density function $\rho(X)$. Next, let us divide the range of $X$ into bins (classes) of length $\delta$,

not without assuming that the probability density function is continuous within each bin. Then, by the mean value theorem, there exists a value $x_k$ within each bin such that:

$$\rho(x_k)\delta = \int_{k\delta}^{(k+1)\delta} \rho(x)\, dx. \tag{2.23}$$

Let us consider now the quantised random variable $X_\delta$, defined by:

$$X_\delta = x_k \text{ , if } k\delta \leqslant X < (k+1)\delta. \tag{2.24}$$

Then, the probability that $X_\delta = x_k$ is given by:

$$p_k = \int_{k\delta}^{(k+1)\delta} \rho(x_k)\, \delta. \tag{2.25}$$

The Entropy of the quantised version is:

$$\begin{aligned} H(X_\delta) &= -\sum_{k=-\infty}^{\infty} p_k \ln p_k \\ H(X_\delta) &= -\sum_{k=-\infty}^{\infty} \rho(x_k)\, \delta\, \ln[\rho(x_k)\, \delta] \\ H(X_\delta) &= -\sum_{k=-\infty}^{\infty} \delta\, \rho(x_k) \ln \rho(x_k) - \sum_{k=-\infty}^{\infty} \rho(x_k)\, \delta\, \ln \delta \\ H(X_\delta) &= -\sum_{k=-\infty}^{\infty} \delta\, \rho(x_k) \ln \rho(x_k) - \ln \delta \end{aligned} \tag{2.26}$$

since $\sum \rho(x_k)\, \delta = \int \rho(x) dx = 1$. If $\rho(x) \ln \rho(x)$ is Riemann integrable, then the first term approaches the integral of $-\rho(x) \ln \rho(x)$ by definition of Riemann integrability. This leads to the following theorem:

*If the probability density function $\rho(x)$ of the continuous random variable $X$ is Riemann integrable, then*:

$$H(X_\delta) + \ln \delta \to h(X) \text{ as } \delta \to 0. \tag{2.27}$$

Therefore, the Entropy of an $n$-bit quantisation of a continuous random variable $X$ is approximately $h(X) + n$.

The relation between Entropy and Differential Entropy thus depends on the scale of the discretisation. Even so, by fixing the discretisations, the Discrete and Differential Entropy variations shall be nearly identical, i.e. $\Delta H(X_\delta) \approx \Delta h(X)$.

### 2.3.3 Negentropy

We are now able to define an information-theoretical measure of non-Gaussianity, consisting on a normalised version of Differential Entropy: *Negentropy* (Hyvärinen et al., 2001).

Let $\mathbf{X}$ be an $n$-dimensional continuous random variable as defined in previous subsections. Negentropy is defined by:

$$\mathsf{J}(\mathbf{X}) = h(\mathbf{X}_G) - h(\mathbf{X}), \tag{2.28}$$

where $\mathbf{X}_G$ is a Gaussian random variable with the same mean and covariance matrix as that of $\mathbf{X}$.

The Differential Entropy of a multivariate Gaussian distribution such as that of $\mathbf{X}_G$ is provided by the following theorem:

*Let $\mathbf{X} = X_1, \cdots, X_n$ have a multivariate Gaussian distribution with mean $\mathfrak{m}$ and covariance matrix $\mathbf{K}$, denoted by $\mathcal{N}_n(\mathfrak{m}, \mathbf{K})$. Its Entropy is then given by:*

$$h(X_1, \cdots, X_n) = h[\mathcal{N}_n(\mathfrak{m}, \mathbf{K})] = \frac{1}{2} \ln \left[ (2\pi e)^n |\mathbf{K}| \right], \tag{2.29}$$

*where $|\mathbf{K}|$ denotes the determinant of $\mathbf{K}$ and is the inner volume enclosed by the ellipsoid with principal axes given by the eigenvectors of the covariance matrix.*

A fundamental result from Information Theory says that the Entropy of a Gaussian random variable is the highest as compared with Entropies of random variables of equal variance. Therefore, Negentropy is always non-negative: zero

only for Gaussian random variables, and growing as the probability distribution of the variable deviates further from Gaussianity. It is thus a measure of non-Gaussianity of a distribution.

The relationship between Differential and discrete Statistical Entropies presented in the previous subsection allows for Negentropy to be equivalently defined in the discrete case as well, yielding approximately the same result regardless of the discretisation scale $\delta$. Negentropy is thus scale-invariant.

An alternative way to define Negentropy can be considered by invoking the concept of *Relative Entropy*, presented in the next subsection.

### 2.3.4 Relative Entropy

The *Relative Entropy* or *Kullback-Leibler Distance* (KLD) between two probability functions represented by $p_X(x)$ and $\widetilde{p}_X(x)$ can be defined as:

$$D\left[p_X(x)\|\widetilde{p}_X(x)\right] = \sum_{x\in\Omega} p_X(x) \ln \frac{p_X(x)}{\widetilde{p}_X(x)}, \tag{2.30}$$

as far as discrete random variables are concerned. Here it is assumed, by convention based on continuity arguments, that $0 \ln \frac{p_X(x)}{\widetilde{p}_X(x)}$ and $p_X(x) \ln \frac{p_X(x)}{0} = \infty$ (Cover and Thomas, 1991).

In the continuous case, the Kullback-Leibler Distance between two probability density functions $\rho_X(x)$ and $\widetilde{\rho}_X(x)$ can be defined as:

$$D\left[\rho_X(x)\|\widetilde{\rho}_X(x)\right] = \int_{x\in\Omega} \rho_X(X) \ln \frac{\rho_X(x)}{\widetilde{\rho}_X(x)} dx. \tag{2.31}$$

By continuity arguments, we set $0 \ln \frac{0}{0} = 0$.

A necessary condition for $D\left[\rho_X(x)\|\widetilde{\rho}_X(x)\right]$ to be finite is that the support set of $\rho_X(x)$ is contained in the support set of $\widetilde{\rho}_X(x)$.

It can be proven [as in Cover and Thomas (1991)] that

$$D\left[p_X(x)\|\widetilde{p}_X(x)\right] \geqslant 0, \forall p, \widetilde{p} \tag{2.32}$$

with the equality holding *iff* $p_X(x) = \widetilde{p}_X(x)$, for all random variables $X$. A similar relation holds for the continuous case.

The Relative Entropy is a scale-invariant measure of the distance between two distributions and can be interpreted as a measure of the inefficiency of assuming that the probability function of a certain random variable $X$ is $\widetilde{p}_X(x)$ when the true one is $p_X(x)$ – or, equivalently for a continuous $X$, that its probability density function is $\widetilde{\rho}_X(x)$ when the true one is $\rho_X(x)$.

Note, however, that Kullback-Leibler Distance is not a formal distance between distributions, since it neither is symmetric nor satisfies the triangle inequality. Still, it is always non-negative, as a consequence from the strict convexity of the logarithm and the application of the classic Jensen's inequality (Hyvärinen et al., 2001).

A corresponding chain rule can be written as follows:

$$D\left[p_{X,Y}(x,y)\|\widetilde{p}_{X,Y}(x,y)\right] = D\left[p_X(x)\|\widetilde{p}_X(x)\right] + D\left[P(Y|X)\|\widetilde{P}(Y|X)\right], \quad (2.33)$$

where $p_{X,Y}(x,y)$ is the joint probability of random variables $X$ and $Y$ (i.e., $P(X = x, Y = y)$), $P(Y|X)$ is the conditional probability of $Y$ given $X$, and $P(X|Y)$ that of $X$ given $Y$. An analogous meaning applies to $\widetilde{p}$ and $\widetilde{P}$, with the difference that they refer to a different probability distribution.

Another way of looking at Relative Entropy is by considering, in the definition (2.31), the particular case in which $\widetilde{\rho}_X(x) = \rho_{X_G}(x)$ is the Gaussian probability distribution function with the same first and second-order statistics as those of $\rho_X$. In this case, (2.31) becomes:

$$D\left[\rho_X(x)\|\rho_{X_G}(x)\right] = \int_{x\in\Omega} \rho_X(X) \ln \frac{\rho_X(x)}{\rho_{X_G}(x)} dx. \quad (2.34)$$

By interpreting this equation and bearing in mind the previous section, it can be

seen that we are in the presence of an alternative − yet equivalent − definition for Negentropy, as invoked in Pires and Perdigão (2007) and references therein.

### 2.3.5 Mutual Information

Consider two discrete random variables $X$ and $Y$ with a joint probability function $p_{X,Y}(x,y)$ and marginal probability functions $p_X(x)$ and $p_Y(y)$. The *Mutual Information* $I(X,Y)$ between those variables is the Relative Entropy between the joint distribution and the product distribution $p_X(x)\,p_Y(y)$, i.e.,

$$\begin{aligned} I(X,Y) &= D\left[p_{X,Y}(x,y) \| p_X(x) p_Y(y)\right] \\ &= \sum_x \sum_y p_{X,Y}(x,y) \ln \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \end{aligned} \tag{2.35}$$

Analogously, by considering $X$ and $Y$ as continuous random variables with marginal probability density functions $\rho_X(x)$ and $\rho_Y(y)$ and joint probability density function $\rho_{X,Y}(x,y)$, the Mutual Information can be defined as:

$$\begin{aligned} I(X,Y) &= D\left[\rho_{X,Y}(x,y) \| \rho_X(x) \rho_Y(y)\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_{X,Y}(x,y) \ln \frac{\rho_{X,Y}(x,y)}{\rho_X(x) \rho_Y(y)}. \end{aligned} \tag{2.36}$$

As a Kullback-Leibler distance, the Mutual Information is always non-negative, will nullity happening *iff* the considered random variables are independent from each other. In fact, as it had previously been seen, saying that $X$ and $Y$ are independent is equivalent to the statement that $p_{X,Y}(x,y) = p_X(x) p_Y(y)$. This, applied into (2.35), leads to $I(X,Y) = 0$, and analogously in the continuous case.

This is rather intuitive, since if the variables are independent, one cannot tell us anything about the other. The Mutual Information is thus a practical way of diagnosing a degree of dependence between random variables, and, in particular, of checking whether they are independent. Apart from that, it can equivalently be seen as a measure of the amount of information that one random variable

contains about another, i.e., as the reduction in the uncertainty of one random variable due to the knowledge of the other.

By relating the definition of Mutual Information $I(X,Y)$ with that of Statistical Entropy we are led to the following expression:

$$I(X,Y) = H(X) - H(X|Y) \tag{2.37}$$

By symmetry, it also follows that

$$I(X,Y) = H(Y) - H(Y|X) \tag{2.38}$$

Hence, $X$ says as much about $Y$ as $Y$ about $X$.

Since:

$$H(X,Y) = H(X) + H(Y|X), \tag{2.39}$$

we have:

$$I(X,Y) = H(X) + H(Y) - H(X,Y). \tag{2.40}$$

It is interesting to note that

$$I(X,X) = H(X) - H(X,X) = H(X). \tag{2.41}$$

Therefore, the Mutual Information between a random variable and itself is the Statistical Entropy of the random variable, which leads to the reason why Entropy is sometimes regarded as *self-information*.

A general definition of Mutual Information between $n$ scalar random variables $X_i$ (or components of a $n$-dimensional random variable $\mathbf{X}$) can thus be stated as follows:

$$I(X_1, \cdots, X_n) = \sum_{i=1}^{n} H(X_i) - H(\mathbf{X}) \tag{2.42}$$

The Mutual Information (MI) vanishes *iff* $X$ and $Y$ are statistically independent from each other or equivalently *iff* all nonlinear correlations are zero

for smooth PDFs, thus making MI a stronger measure of independence than the Pearson correlation is. Furthermore, MI is invariant for any $X$ and $Y$ single homeomorphisms. In particular, it is invariant when $X$ and $Y$ are replaced by the corresponding Gaussian Anamorphosis, say:

$$I(X,Y) = I(\mathcal{G}_X, \mathcal{G}_Y) \geqslant 0. \tag{2.43}$$

Given the fact that, while a null correlation does not necessarily imply independence, a null Mutual Information actually does, the latter is used in the present work as a way to define a new concept of relationship between random variables that will allow us to diagnose non-Gaussianity within a distribution without the practical difficulties of having to determine Negentropy. That is the purpose of the next subsection.

### 2.3.6 Information Correlation

We hereby present a measure of statistical dependence sharing the intuitiveness of the Pearson Correlation and the robustness of Mutual Information.

By imposing the knowledge of the correlation $c$, a positive lower bound for Mutual Information, $I_g$ (hereby denoted as Gaussian Mutual Information) can be found by solving a constrained variational problem of MI minimisation (Kraskov et al., 2004), thus leading to the decomposition:

$$I(X,Y) = I_g(X,Y) + I_{ng}(X,Y) \leqslant I_g(X,Y), \tag{2.44}$$

where:

$$I_g(X,Y) = -\frac{1}{2}\ln(1-c^2) \tag{2.45}$$

is the Mutual Information within a bivariate Gaussian distribution with the same mean and covariance matrix as that of $(X,Y)$, with correlation $c$ (Cover and

Thomas, 1991). As for $I_{ng}(X,Y)$, it denotes the contribution of non-Gaussianity to MI.

The upper bound (2.45) can be generalised by replacing the correlation $c$ by any nonlinear $(X,Y)$ correlation. The Gaussian upper bound (2.45) of MI is also used in speech analysis (Abdallah and Plumbey, 2003).

Mutual Information can be compared with the Gaussian correlation by defining a distance between $X$ and $Y$, hereby denoted as *Information Correlation* and given by (Pires and Perdigão, 2007):

$$c_{\text{inf}}(X,Y) = \sqrt{(1-\exp[-2I(X,Y)])} \geqslant |c(X,Y|. \tag{2.46}$$

The equality holds iff the joint distribution of $(X,Y)$ is Gaussian or, equivalently, *iff* $I_{ng}$ is null. As it happens with MI, $c_{\text{inf}}$ vanishes in the case of statistical independence. By applying the chain rule of KLD (2.33) or resorting to the derivation in Appendix B, the contribution of non-Gaussianity to MI, $I_{ng}$, can be decomposed as:

$$\begin{aligned} I_{ng}(X,Y) &= \mathsf{J}(X,Y) - \mathsf{J}(X) - \mathsf{J}(Y) \\ &= \mathsf{J}(X,Y_r) - \mathsf{J}(X) - \mathsf{J}(Y), \end{aligned} \tag{2.47}$$

where $\mathsf{J}(.)$ is Negentropy (2.28).

In (2.47), the Negentropy $\mathsf{J}(X,Y)$ is invariant under a two-dimensional linear homeomorphism of $(X,Y)$. Therefore, without loss of generality, it is equal to the Negentropy between the uncorrelated variables $X$ and the prediction residue $Y_r$ (2.14). If $X$ and $Y$ are previously subjected to Gaussian Anamorphosis, the marginal Negentropies $\mathsf{J}(X)$ and $\mathsf{J}(Y)$ vanish.

## 2.4 Estimation of information-theoretical measures of non-Gaussianity

The numerical estimation of information-theoretical measures such as Mutual Information or Negentropy is rather difficult and has no unbiased estimators (Paninski, 2003, 2004). Still, several approaches do exist for that assessment, such as the plug-in, bin-adaptive networks (Kraskov et al., 2004), and Maximum Entropy (ME) (Abramov, 2006; Pires and Perdigão, 2007). While efficient, these numerical methods are somewhat computationally costly.

An alternative method is hereby presented, based on an analytical approach presented in Pires and Perdigão (2007). It essentially consists on using expansions in series of Chebyshev-Hermite polynomials for estimating the joint PDF of $(X, Y)$ or any transformed pair such as $(\mathcal{G}_X, \mathcal{G}_Y)$ in terms of its statistical cumulants, then to determine its Mutual Information and Negentropy.

By taking rotated standardised variables $V = X$ and $W = Y_r$ (2.14) from $(X, Y)$, the proposed formulation will be considerably simplified, without loss of generality. In fact, the variable transformation leads to much simpler to handle uncorrelated variables, while retaining generality due to being bijective.

Several approaches exist to approximate PDFs in series of Chebyshev-Hermite polynomials, namely the Gram-Charlier, Gauss-Hermite and Edgeworth Expansions (Edgeworth, 1905).

One must, however, note that these expansions hold for weak non-Gaussianity scenarios. Anyway, this fact poses no problem to the applications contemplated in the present thesis − and, by that matter, to a wide variety of studies in weak non-Gaussian scenarios from other fields such as Fluid Mechanics (e.g. Tabeling et al., 1996) and Astrophysics (e.g. Blinnikov and Moessner, 1998).

Blinnikov and Moessner (1998) showed that while the Gram-Charlier exhibited poor convergence properties and the Gauss-Hermite had no intrinsic measure of accuracy, the Edgeworth expansion provided the best results in PDF estimation. Moreover, as noted in Comon (1994), while in the Gram-Charlier expansion it is not possible to peer into the relative significance of the various terms in the expansion, in the Edgeworth expansion the terms are ordered according to their decreasing significance.

In the present work the expansion of choice will thus be the Edgeworth Expansion, generalised and applied to a bivariate probability density function, as presented in the following subsection.

### 2.4.1 Edgeworth Expansion of a Bivariate Probability Density Function

Let us consider the joint probability distribution function $\rho_{V,W}(v,w)$ of two standard uncorrelated variables $V, W$, each of which assumed as arithmetic averages of $n_{\text{eq}}$ independent and identically distributed (i.i.d.) variables. The joint PDF $\rho_{V,W}(v,w)$ can be approximated by:

$$\rho_{V,W}(v,w) = \varphi(v)\varphi(w)[1 + \nu_{V,W}(v,w) + \mathcal{O}(n_{\text{eq}}^{-l})], \tag{2.48}$$

where $\varphi()$ is the standard Gaussian PDF, $\nu_{V,W}(v,w)$ is a truncated fitting polynomial vanishing if the joint PDF is Gaussian and $l$ is a positive adimensional quantity increasing with the truncation order. The greater $n_{\text{eq}}$, the smaller the truncation error of the expression and the closer to Gaussianity the joint $(V,W)$ PDF, due to the central limit theorem (Papoulis, 1991).

The function $\nu_{V,W}(v,w)$ is expanded in terms of Hermite orthogonal polynomials $H_i$ of order $i$ and joint cumulants $k^{(p,q)}$ of order $p$ in $V$ and of order $q$ in $W$,

which are appropriate joint polynomial expectancies of $V$ and $W$. In the present case, where we assume $V$ and $W$ to be averages of $n_{\text{eq}}$ i.i.d. random variables, the cumulants $k^{(p,q)}$ are scaled to the order of $n_{\text{eq}}^{-(p+q)/2+1}$ (Comon, 1994).

For $l = 3/2$, the function $\nu_{V,W}(v, w)$ is given by:

$$\nu_{V,W}(v,w) = \frac{1}{3!}\nu^{(a)}_{V,W}(v,w) + \frac{1}{4!}\nu^{(b)}_{V,W}(v,w) + \frac{10}{6!}\nu^{(c)}_{V,W}(v,w) \tag{2.49a}$$

where

$$\begin{aligned}
\nu^{(a)}_{V,W}(v,w) &= k^{(3,0)}H_3(v) + 3k^{(2,1)}H_2(v)H_1(w) \\
&\quad + 3k^{(1,2)}H_1(v)H_2(w) + k^{(0,3)}H_3(w) \\
\nu^{(b)}_{V,W}(v,w) &= k^{(4,0)}H_4(v) + 4k^{(3,1)}H_3(v)H_1(w) + 6k^{(2,2)}H_2(v)H_2(w) \\
&\quad + 4k^{(1,3)}H_1(v)H_3(w) + k^{(0,4)}H_4(w) \\
\nu^{(c)}_{V,W}(v,w) &= k^{(3,0)}k^{(3,0)}H_6(v) + 6k^{(3,0)}k^{(2,1)}H_5(v)H_1(w) \\
&\quad + \left[9k^{(3,0)}k^{(1,2)} + 6k^{(2,1)}k^{(2,1)}\right]H_4(v)H_2(w) \\
&\quad + \left[2k^{(3,0)}k^{(0,3)} + 18k^{(2,1)}k^{(1,2)}\right]H_3(v)H_3(w) \\
&\quad + \left[9k^{(0,3)}k^{(2,1)} + 6k^{(1,2)}k^{(1,2)}\right]H_2(v)H_4(w) \\
&\quad + 6k^{(1,2)}k^{(0,3)}H_1(v)H_5(w) + k^{(0,3)}k^{(0,3)}H_6(w)
\end{aligned} \tag{2.49b}$$

The terms $H_p(v)$, $H_q(w)$ are Hermite polynomials, given by the following recurring relationship (Abramowitz and Stegun, 1972):

$$H_{n+1}(r) = rH_n(r) - \frac{\partial}{\partial r}H_n(r), \quad H_0(r) = 1 \tag{2.50}$$

and satisfying the orthogonal relationships:

$$\int_{-\infty}^{\infty} \varphi(r)H_p(r)H_q(r)dr = p!\,\delta_{pq},$$

where (2.51)

$$\varphi(r) = \frac{1}{\sqrt{2\pi}}\exp\left(\frac{-r^2}{2}\right).$$

The Hermite polynomials used in (2.49) thus become:

$$\begin{aligned} H_1(r) &= r \\ H_2(r) &= r^2 - 1 \\ H_3(r) &= r^3 - 3r \\ H_4(r) &= r^4 - 6r^2 + 3 \\ H_5(r) &= r^5 - 10r^3 + 15r \\ H_6(r) &= r^6 - 15r^4 + 45r^2 - 15. \end{aligned} \tag{2.52}$$

The coefficients of the Hermite Polynomials $H_p(v)$ and $H_q(w)$ in (2.49) are expressed as products of the $(p+q)$-order cumulants $k^{(p,q)}$ of the $(V, W)$ distribution (Kenney and Keeping, 1951).

In its general form, the joint statistical cumulant of the generic random variables $Z_1, \cdots, Z_n$ is given by (Kendall and Stuart, 1969):

$$k(Z_1, \cdots, Z_n) = \sum_{\mathsf{P}} (|\mathsf{P}| - 1)! \, (-1)^{|\mathsf{P}|-1} \prod_{\mathsf{B} \in \mathsf{P}} E\left[\prod_{i \in \mathsf{B}} Z_i\right] \tag{2.53}$$

where $\mathsf{P}$ runs through the list of all partitions of $\{1, \cdots, n\}$, and $\mathsf{B}$ runs through the list of all blocks of the partition $\mathsf{P}$.

By considering the case in which $Z_i = V$, with $i = 1, \cdots, p$ and $Z_j = W$, with $j = 1, \cdots, q$, such that $p + q = n$, we introduce the notation $k^{(p,q)}(V, W)$ or simply $k^{(p,q)}$ to denote the joint cumulants of order $p$ in $V$ and of order $q$ in $W$:

$$k^{(p,q)} \equiv k(V_1, \cdots, V_p, W_1, \cdots, W_q) \tag{2.54}$$

When dealing with uncorrelated variables $V$ and $W$ of zero mean and unit variance, as it is the case in our present application, the cumulants used for the chosen truncation assume rather simple expressions:

$$k^{(p,q)} = E(V^p W^q), \;\; p, q \geqslant 0 \;\; \text{and} \;\; p + q \leqslant 3\,; \tag{2.55a}$$

$$k^{(4,0)} = E(V^4) - 3 \quad \text{and} \quad k^{(0,4)} = E(W^4) - 3\,; \tag{2.55b}$$

$$k^{(1,3)} = E(VW^3),\ k^{(3,1)} = E(V^3W)\,; \tag{2.55c}$$

$$k^{(2,2)} = E(V^2W^2) - 1\,. \tag{2.55d}$$

Under Gaussian conditions, all cumulants of order $(p+q)$ equal to or greater than 3 will vanish. Cumulants with $p \neq 0$ and $q \neq 0$ can easily be expressed in terms of nonlinear correlations between $V$ and $W$.

Nonzero cumulants of order $p+q$ higher than or equal to three reveal non-Gaussianity. In particular, if $X$ is rendered Gaussian the self $V$-cumulants $k^{(p,0)}$, $p \geqslant 3$ vanish. The Edgeworth PDF (hereafter EDG-PDF) expansion converges in $\mathcal{L}^2$, needing a large truncation or $l$ in cases of high non-Gaussianity. It has the drawback that errors in tail regions of the distribution may be comparable to the PDF itself and even present negative values. In order to verify the positivity and normalisation of the EDG-PDF, we numerically compute the integrals $P_{pos}$ and $P_{neg}$ of the EDG-PDF, respectively in the domain of positive and negative values of the estimated truncated density (2.48). The integrals are estimated by bivariate Gaussian quadrature, mapping the open interval $]-\infty,\infty[$ into $]-1,1[$ through the transformation $x \rightarrow f(x) = x/(1+|x|)$. The use of 50 weighting quadrature points has been sufficient for convergence of integrals with an accuracy of $\sim 10^{-4}$. A satisfactory condition for (2.48) to be a density is: $|P_{neg}| \ll P_{pos} \sim 1$, which holds if cumulants in (2.49) are sufficiently small. A non-general rule for approaching the referred condition is to perform Gaussian anamorphosis of the original variables. A reduced Edgeworth truncation is still valid considering a "tilted" variable whose modal region is nearer to the neighborhood where we wish to approximate, using the saddle approximation technique (Daniels, 1954).

That way, the PDF $\rho_{X,Y}(x,y)$ of the unrotated variables can be retrieved through the following expression:

$$\rho_{X,Y}(x,y) = \frac{1}{\sqrt{1-c^2}}\,\rho_{V,W}\left(x, \frac{y-c\,x}{\sqrt{1-c^2}}\right). \tag{2.56}$$

### 2.4.2 Application to non-Gaussianity estimation

In order to compute the non-Gaussian MI (2.47) of $(X,Y)$ one determines the joint Negentropy $\mathsf{J}(X,Y_r)$. For simplicity of notation, let $V = X$ and $W = Y_r$. Then $\mathsf{J}(X,Y_r) = \mathsf{J}(V,W)$, which is simply the KLD between the joint PDF $\rho_{V,W}(v,w)$ and the product of the marginal PDFs $\varphi_V(v)\varphi_W(w)$, reducing to:

$$\begin{aligned}\mathsf{J}(V,W) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(v)\varphi(w)[1+\nu_{V,W}(v,w)+\mathcal{O}\left(n_{\text{eq}}^{-l}\right)] \\ \cdot \ln[1+\nu_{V,W}(v,w)+\mathcal{O}\left(n_{\text{eq}}^{-l}\right)]\,dv\,dw.\end{aligned} \tag{2.57}$$

The marginal negentropies $\mathsf{J}(X)$ and $\mathsf{J}(Y)$ are estimated through the equivalent equation to (2.57) for single variables. Two ways of computing (2.57) are followed. First, it is numerically computed in the domain: $(u,w) : 1+\nu_{V,W}(v,w) \geqslant 0$, in the same way as $P_{neg}$ and $P_{pos}$. The non-Gaussian MI obtained through the estimated integral is denoted as $I_{ng(EI)}$. Then, we consider the Taylor expansion:

$$\begin{aligned}\left[1+\nu+\mathcal{O}\left(n_{\text{eq}}^{-l}\right)\right]\ln\left[1+\nu+\mathcal{O}\left(n_{\text{eq}}^{-l}\right)\right] \\ \sim \nu+\frac{\nu^2}{2}+\mathcal{O}\left(\nu^3\right)+\mathcal{O}\left(n_{\text{eq}}^{-l}\right),\end{aligned} \tag{2.58}$$

where the logarithm expansion error $\mathcal{O}\left(\nu^3\right)$ is $\mathcal{O}\left(n_{\text{eq}}^{-3/2}\right)$ since $\nu \sim n_{\text{eq}}^{-1/2}$. By noting that the function $\nu_{V,W}(v,w)$ is orthogonal to the product of Gaussian PDFs $\varphi(u)\varphi(w)$, the Negentropy becomes the following positive quantity:

$$\begin{aligned}\mathsf{J}(V,W) = \frac{1}{2}\int_{-\infty}^{\infty}\varphi(v)\varphi(w)\nu_{V,W}^2(v,w)\,dv\,dw \\ +\mathcal{O}\left(n_{\text{eq}}^{-3/2}\right)+\mathcal{O}\left(n_{\text{eq}}^{-l}\right).\end{aligned} \tag{2.59}$$

By using the norms and orthogonality properties of the Hermite polynomials, (2.59) is expanded in terms of powers of cumulants. For $l = 3/2$ we have:

$$\mathsf{J}(V,W) = \mathsf{J}(X,Y_r) = \frac{1}{12}\mathsf{J}^{(a)} + \frac{1}{48}\mathsf{J}^{(b)} + \frac{1}{72}\mathsf{J}^{(c)} + \mathcal{O}\left(n_{\text{eq}}^{-3/2}\right), \tag{2.60a}$$

where

$$\begin{aligned}
\mathsf{J}^{(a)} &= 3\left[k^{(2,1)}\right]^2 + 3\left[k^{(1,2)}\right]^2 + \left[k^{(0,3)}\right]^2 + \left[k^{(0,3)}\right]^2 \\
\mathsf{J}^{(b)} &= \left[k^{(4,0)}\right]^2 + 4\left[k^{(3,1)}\right]^2 + 6\left[k^{(2,2)}\right]^2 + 4\left[k^{(1,3)}\right]^2 + \left[k^{(0,4)}\right]^2 \\
\mathsf{J}^{(c)} &= 5\left[k^{(3,0)}\right]^4 + 30\left[k^{(3,0)}k^{(2,1)}\right]^2 + \frac{2}{3}\left[9k^{(3,0)}k^{(1,2)} + 6k^{(2,1)}k^{(2,1)}\right]^2 \\
&\quad + \frac{2}{3}\left[9k^{(0,3)}k^{(2,1)} + 6k^{(1,2)}k^{(1,2)}\right]^2 + \frac{1}{2}\left[2k^{(3,0)}k^{(0,3)} + 18k^{(2,1)}k^{(1,2)}\right]^2 \\
&\quad + 30\left[k^{(1,2)}k^{(0,3)}\right]^2 + 5\left[k^{(0,3)}\right]^4.
\end{aligned} \tag{2.60b}$$

The above estimation is valid for standard uncorrelated variables $V, W$. Equation (2.60) is a sum of the quadratic positive contributions from cumulants of order higher than two, while also resulting from a truncated expansion of the logarithm function.

On one hand, a simplification has been performed on the truncation order with respect to what had been obtained by Comon (1994), where a truncation of up to $\mathcal{O}(n_{\text{eq}}^{-2})$ had been considered. On the other hand, while simplifying on the order of truncation, a generalisation has been performed as to obtain the Edgeworth Expansion of the joint Negentropy, whereas Comon (1994) had studied the one-dimensional case.

The Non-Gaussiam MI obtained through (2.60) is denoted by $I_{ng(EF)}$. By having an explicit estimation of the PDF and taking into account the integral properties of Hermite Polynomials, we can derive an analytic expression for the conditional expectancy of $Y$ given $X$:

$$E(Y|X=x) = c\,x + \sqrt{1-c^2}\,E(W|V=x), \tag{2.61}$$

which highlights the effects of non-Gaussianity. The first *rhs* term of (2.61) is the linear prediction of $Y$ from $X$, also represented by $Y(\text{lin})$, whereas the second *rhs* is the additive correction due to non-Gaussianity, thus yielding the full nonlinear prediction $Y(\text{nolin})$. Given the Edgeworth Expansions of both the joint and the marginal probability distributions, the conditional expectancy of the variable $W$ given the variable $V$ can be approximated by:

$$E(W|V=x) = \frac{1/2\, k^{(2,1)} H_2(x) + 1/6 k^{(3,1)} H_3(x)}{1+\gamma_V(x)} + \mathcal{O}\left(n_{\text{eq}}^{-3/2}\right), \tag{2.62}$$

where $H_i(\cdot)$ are single Hermite polynomials with the standard Gaussian kernel (2.50), (2.51). The correctional polynomial $\gamma_V(x)$ at truncation $l = 3/2$ of the $X$ marginal EDG-PDF is given by:

$$\gamma_V(x) = \frac{1}{3!} k^{(3,0)} H_3(x) + \frac{1}{4!} k^{(4,0)} H_4(x) + \frac{10}{6!} k^{(3,0)} k^{(3,0)} H_6(x) + \mathcal{O}\left(n_{\text{eq}}^{-3/2}\right). \tag{2.63}$$

The cumulants $k^{(3,0)}$ and $k^{(4,0)}$ are, respectively, the skewness and the kurtosis (relative to that of the normal distribution) of the probability distribution of $X = V$ [see (2.55)]. Equations (2.61) and (2.62) provide an easy nonlinear downscaling relationship of $Y$ from $X$.

## 2.5 Dataset and processing

### 2.5.1 Data

We use National Centers for Environmental Prediction − National Center for Atmospheric Research (NCEP/NCAR) reanalysis CD-ROMs (Kistler et al., 2001) to extract December to February (DJF) monthly data of sea-level pressure (SLP) and corrected precipitation, from 1951 to 2003 over the NAE domain (30°N−70°N, 80°W−40°E), with grid size of 2.5° in latitude and longitude. We consider $X$ to be the standardised (zero average and unit variance) NAO index

given by the first Principal Component (PC) of the detrended SLP monthly data over the NAE in DJF, over the above-mentioned period. This PC-based NAO index ($X$) is related to more traditional indexes based on SLP differences (Osborn et al., 1999). The PDF of the PC-based NAO index is slightly platykurtic. In fact, its kurtosis $kur(X)$ is equal to $-0.7$, with $kur(X) = E(X^4)/\sigma^4(X) - 3$ (where $\sigma$ denotes the standard deviation). Moreover, it is bimodal with the presence of two regimes: NAO+ and NAO-. The grid-point standardised monthly detrended precipitation is assigned to the variable $Y$. The skewness and kurtosis of the distribution of monthly DJF precipitation over the Euro-Atlantic area are shown, respectively, in Figs. 2.1a and 2.1b. Given the definition of skewness and kurtosis, higher absolute values indicate larger deviation of the distribution from Gaussianity. When the distribution is Gaussian both skewness and kurtosis are null.

It can thus be seen that the distribution of monthly DJF precipitation over oceanic areas is closer to the Gaussian than the one over land is. It is positively skewed and some locations exhibit outliers with large positive anomalies, with high positive kurtosis values, especially over Greenland [$kur(Y) \approx 5$] Canadian Arctic [$kur(Y) \approx 5$] and some deserted areas over North Africa [$kur(Y) \geqslant 10$]. Those extreme precipitation values inflate joint $(X, Y)$ cumulants, thus rendering inapplicable the Edgeworth formalism of estimating mutual information and non-Gaussianity. In order to obtain, in general, smaller cumulants, we apply Gaussian Anamorphosis both to $X$ and $Y$. For that purpose, in practice we start by sorting data within $X$ in ascending order. The $k$-th value of the Gaussian variable $X_{\mathcal{G}}$ of $X$ in ascending order is given by:

$$X_{\mathcal{G}}(k) = \mathcal{G}_X[x(k)] = \Phi^{-1}\left(\frac{k-0.5}{N}\right), \; k = 1, \cdots, N, \tag{2.64}$$

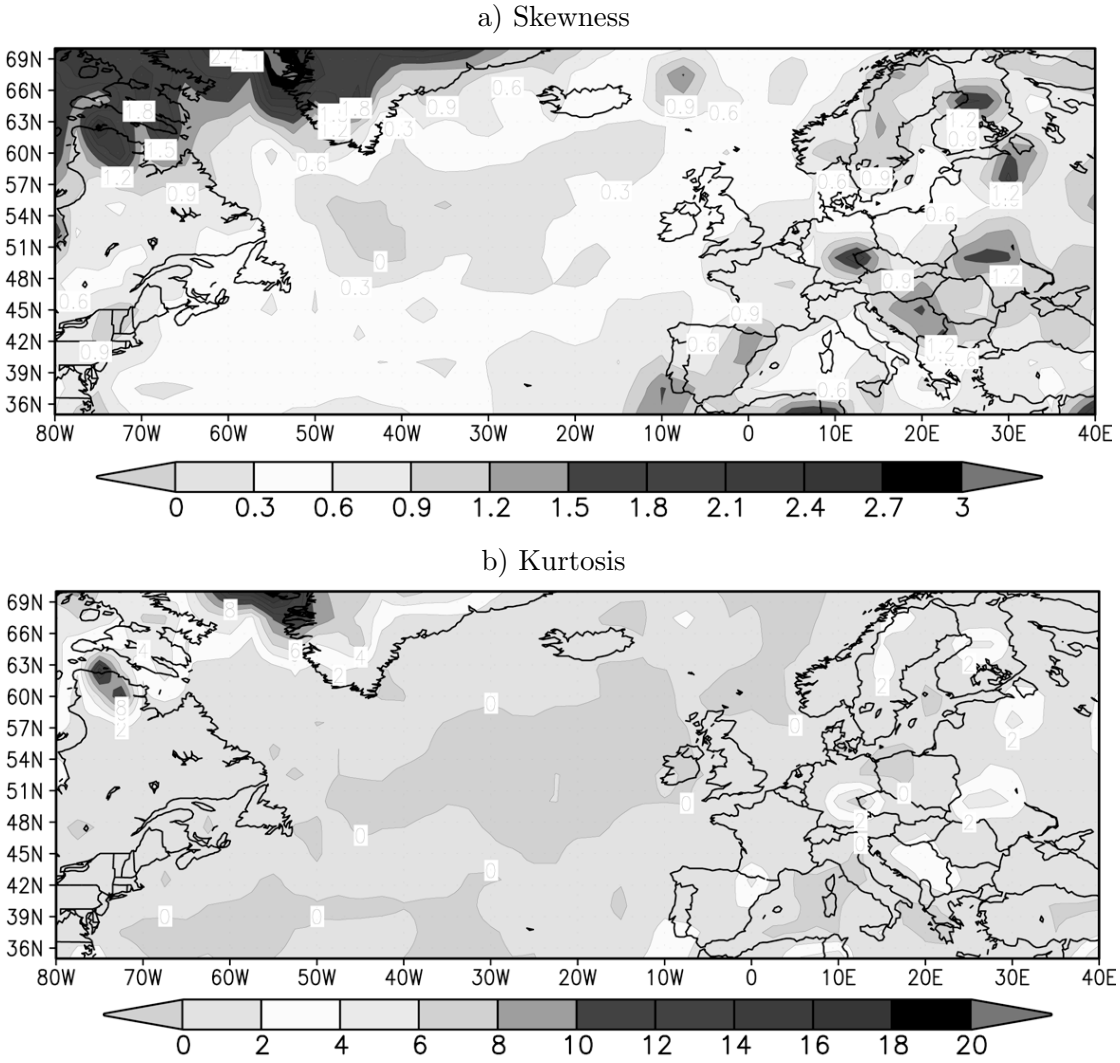



Figure 2.1: Skewness (a) and Kurtosis (b) of the distribution of monthly DJF precipitation over the Euro-Atlantic area. Higher absolute values mean larger deviation from Gaussianity.

where $N = 53 \times 3$ is the total number of months in the sample, and $\Phi^{-1}(\cdot)$ is defined as in (2.3). The same procedure applies to $Y$. The transformation (2.64) assumes that data uniformly cover the true probability distribution and consequently suffer in practice from sampling errors. In order to illustrate the relevance of both non-Gaussianity and asymmetry of the Precipitation response to NAO, we choose six grid-points: 1 − Central Atlantic (ATL; 37.5°N, 25°W); 2 − Northwest Scotland (SCO; 60°N, 12.5°W); 3 − Balearic Islands (BAL; 37.5°N, 2.5°E); 4 − Greenland (GRE; 62.5°N, 45°W); 5 − Eastern United States (EUS; 40°N, 60°W); and 6 − Russia (RUS; 62.5°N, 22.5°E).

### 2.5.2 Statistical Tests

In order to check whether the test side correlations (2.12) and the non-Gaussianity measures are significantly different from zero, we apply the Monte Carlo technique in two versions: the generation of Gaussian white noise (MCG) and the random reordering of working series (MCR). In MCG we generate an ensemble of 10000 bivariate uncorrelated time series of standard Gaussian white noise with $N_{df}$ temporal degrees of freedom. The estimated $N_{df}$ of the pair (monthly NAO index $X$, monthly precipitation $Y$) uses the 1- and 2-month-lag $X, Y$ auto correlations over the DJF period (Livezey and Chen, 1983),

$$N_{df} = N \left[ 1 + 2 \sum_{i=1}^{2} \rho_X(i \text{ months}) \rho_Y(i \text{ months}) \right]^{-1} . \tag{2.65}$$

The $N_{df}$ is fairly uniform over NAE and close to its spatial average $N_{df} \sim 0.95N = 151$, both for Gaussianised and original variables. In the MCR version we consider 300 different proxy NAO series by randomly permuting the 53 analysed years while keeping the DJF monthly sequence, thus keeping in the proxy

Table 2.1: Rejection intervals of the null hypothesis at $\alpha$ level of significance, using the MCG, MCR (Gaussianised data), and MCR (original data) tests from the top to the bottom of each cell, respectively (see text for details).

| $\alpha$ | $\lvert c\rvert$ | $\lvert t_M\rvert$ | $\lvert t_-\rvert\lvert t_+\rvert$ | $\lvert k^{(2,1)}\rvert$ | $\lvert k^{(3,1)}\rvert$ | $\lvert \mathcal{J}_c\rvert$ | $I_g$ | $I_{ng(EF)}$ | $c_{\inf}$ |
|---|---|---|---|---|---|---|---|---|---|
| 90% | $> 0.13$ | $> 0.16$ | $> 0.31$ | $> 0.18$ | $> 0.31$ | $> 0.041$ | $> 0.009$ | $> 0.041$ | $> 0.31$ |
| | 0.13 | 0.16 | 0.31 | 0.18 | 0.28 | 0.040 | 0.009 | 0.027 | 0.28 |
| | 0.13 | 0.16 | 0.31 | 0.18 | 0.28 | 0.040 | 0.009 | – | 0.29 |
| 95% | $> 0.16$ | $> 0.19$ | $> 0.37$ | $> 0.22$ | $> 0.38$ | $> 0.048$ | $> 0.013$ | $> 0.044$ | $> 0.34$ |
| | 0.16 | 0.19 | 0.36 | 0.22 | 0.33 | 0.047 | 0.013 | 0.033 | 0.31 |
| | 0.16 | 0.17 | 0.37 | 0.18 | 0.25 | 0.043 | 0.013 | – | 0.32 |
| 99% | $> 0.21$ | $> 0.26$ | $> 0.49$ | $> 0.29$ | $> 0.54$ | $> 0.063$ | $> 0.022$ | $> 0.078$ | $> 0.43$ |
| | 0.21 | 0.25 | 0.47 | 0.28 | 0.43 | 0.062 | 0.022 | 0.052 | 0.38 |
| | 0.21 | 0.26 | 0.48 | 0.23 | 0.33 | 0.057 | 0.022 | – | 0.39 |

timeseries the 1- and 2-month lag correlations or the serial correlation in the available data. Then, the statistical tests of the randomised NAO and precipitation are computed in each of 833 NAE grid points and then collected altogether in an ensemble of $833 \times 300$ realisations.

In both MCG and MCR cases, the values of the statistical tests are sorted in such a way as to compute quantiles giving the 90%, 95% and 99% significance level intervals (summarised in Table 2.1) of rejection of the null hypothesis $\mathsf{H_o}$ of $(X, Y)$ independence.

Rejection of $\mathsf{H_o}$ is easier for test-side correlations than for $c$ because the former deal with half of the data, below or above the median of $X$. The confidence regions of the information correlation are obtained from $I_g$. Thresholds for $k^{(2,1)}$ and $k^{(3,1)}$ are also computed. MCR tests are more appropriate than the MCG tests because they preserve the marginal distribution of working variables and their spatial correlations. This leads to less conservative criteria of $\mathsf{H_o}$ rejection in MCR, specially for the non-Gaussianity measures $I_{ng(EF)}$ and $\mathcal{J}_c$, as can be seen from the comparison of the $\mathsf{H_o}$- rejection thresholds in Table 2.1.

In all statistical maps the 90% MCR statistical significant regions are shaded. The fraction of statistically significant area (FS) must be larger than that occurring by mere chance (10%). Values of FS for MCR and MCG are denoted, respectively as FS-MCR and FS-MCG.

## 2.6 Results

### 2.6.1 Gaussian and asymmetric correlations

The correlation maps over NAE of the global correlation $c$, the Gaussian correlation $c_g$ and the rank correlation (not shown) between $X$ and $Y$ are rather similar over NAE. They only differ by no more than approximately $\pm 10\%$ at certain regions. The map of $c_g$, shown in Fig. 2.2a is mainly related to the northward (southward) shift of synoptic storm tracks in the NAO+ (NAO-), thus producing positive (negative) precipitation anomalies in the range $50^o$N$-70^o$N, east of $40^o$W, and negative (positive) precipitation anomalies in the range $35^o$N$-45^o$N, east of $60^o$W (Hurrell, 1995). Other negative correlation regions are the southeastern part of Greenland, Canadian Arctic and Labrador. There are some differences between the map of $c_g$ and those of test side correlations of Gaussian data, thus revealing asymmetry. This means that the response of precipitation to NAO is asymmetric and thus non-Gaussian. The gross contribution of $c_g$ (Fig. 2.2a) is due to the test central correlation $t_M(X_{\mathcal{G}}, Y_{\mathcal{G}})$ (Fig. 2.2b), which comes from the alignment slope between the NAO+ and NAO- centroids in the (NAO index, precipitation) space. Consequently, these maps are rather similar.

The differences between $t_-(X_{\mathcal{G}}, Y_{\mathcal{G}})$ and $t_+(X_{\mathcal{G}}, Y_{\mathcal{G}})$ reveal differences in sensitivities to the NAO+ and NAO- regimes. This fact is highlighted in Figs. 2.3a-2.3f, which show, for the six target locations 1) the distribution of the $(X_{\mathcal{G}}, Y_{\mathcal{G}})$

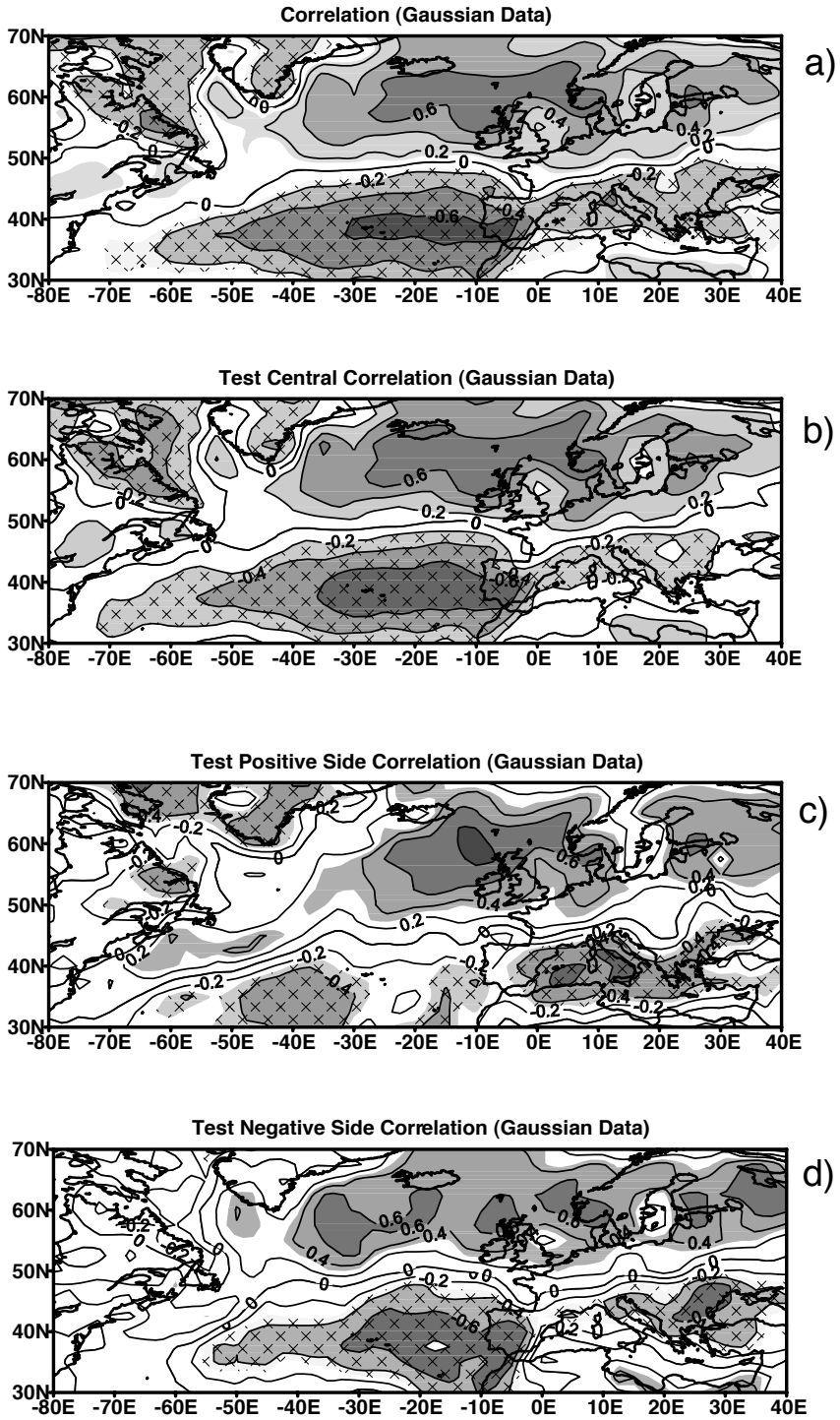



Figure 2.2: (a) Map of correlation $c$, (b) map of the test central correlation $t_M$, (c) map of the test positive side correlation $t_+$, and (d) map of the test negative side correlation $t_-$. All quantities are computed for Gaussian variables, i.e., subject to Gaussian anamorphosis. Contour interval (CI) = 0.2. The significant regions (SR) at $\alpha = 90\%$ are shaded. The 90% significant area fractions FS-MCR are 0.73, 0.68, 0.44, and 0.44 for (a), (b), (c), and (d), respectively. Nearly the same values hold for FS-MCG.

data, 2) the contours of the joint PDF obtained with the expansion (2.48), 3) the linear and nonlinear prediction of $Y_{\mathcal{G}}$ (2.61), and 4) the smoothed graphics of the $X_{\mathcal{G}}$ conditional mean square error (MSE) of the linear $Y_{\mathcal{G}}$(lin) and of the nonlinear $Y_{\mathcal{G}}$(nolin) (2.61) predictions, obtained in full cross validation mode over the $N = 159$ data.

For all six cases, the largest data concentration is seen near the origin ($X_{\mathcal{G}} = 0, Y_{\mathcal{G}} = 0$), where PDF contours are elliptic. The farthest contours are deformed due to extreme values. In the distribution, the negative and positive sides (NAO- and NAO+) can be quite different. Near the central Atlantic, where a large area of strong negative correlations is seen, the sensitivity of precipitation to the NAO index is negatively stronger in the NAO- (wetter) regime than in the NAO+ (drier) regime. This is in accordance with the values: $t_+ = -0.30$ and $t_- = -0.64$ for the ATL point (see Table 2.2). Therefore, as seen in Fig. 2.3a, those data and the PDF spread over a larger domain in the NAO+ regime than in the NAO- regime.

The improvement due to the nonlinear predictions is visible from the reduction of the cross-validated MSE of the NAO-downscaled precipitation, especially for the negative precipitation anomalous values (top of Fig. 2.3a and Table 2.2).

The larger sensitivity of precipitation in the wetter NAO regime is also verified north of Scotland, near the location of most positive correlation extremes, in accordance with $t_-(X_{\mathcal{G}}, Y_{\mathcal{G}}) = 0.43$ and $t_+(X_{\mathcal{G}}, Y_{\mathcal{G}}) = 0.82$ for the SCO point (see Fig. 2.3e and Table 2.2). This behaviour occurs because near the average storm tracking of NAO- and NAO+ regimes, the sensitivity of the precipitation response must be enhanced. The higher the sensitivity, the closer the phenomenon, i.e. a local source produces higher sensitivity than remote sources.

There are regions where the precipitation sensitivity is nearly restricted to

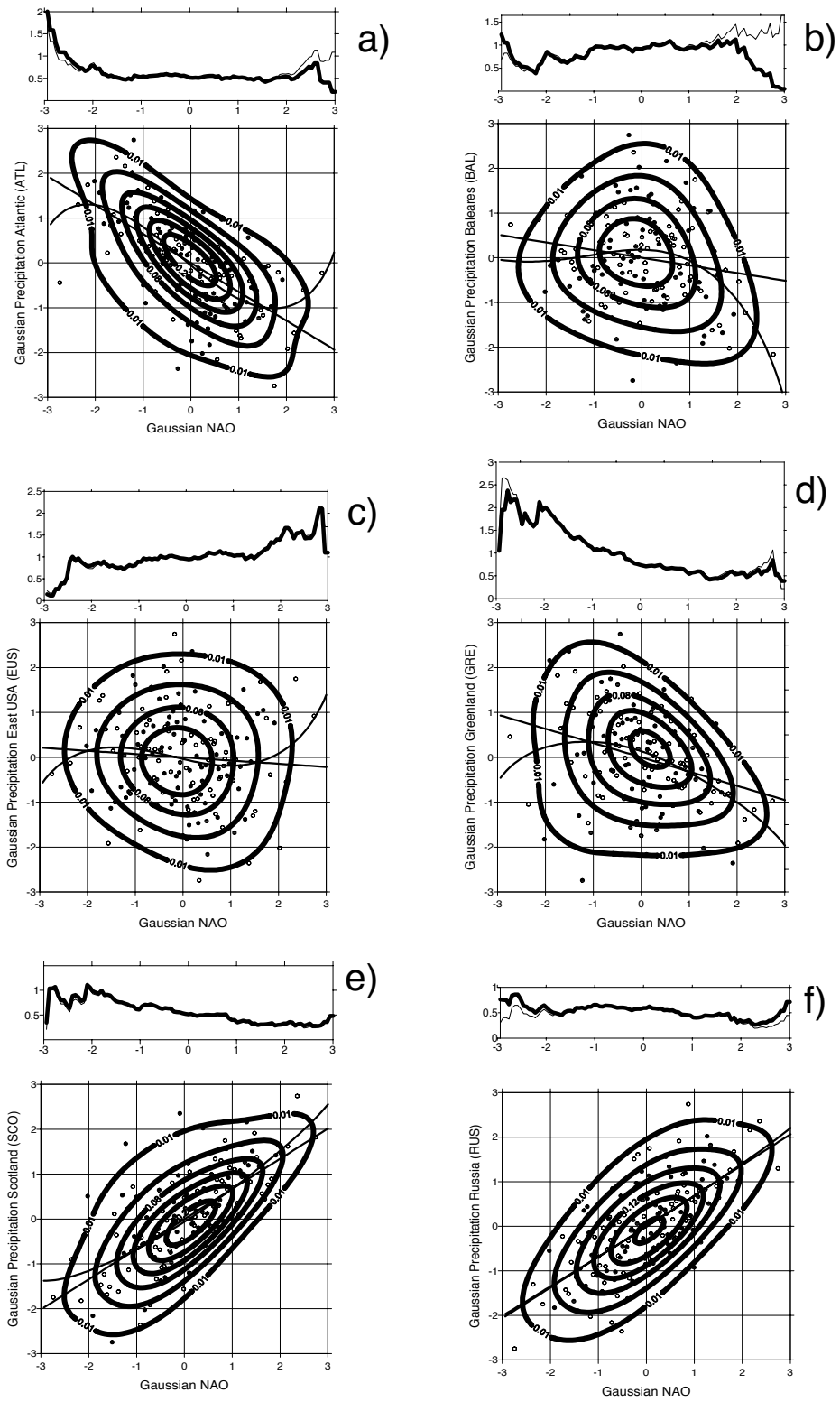



Figure 2.3: Composed graphics for the six selected points: (a) ATL, (b) BAL, (c) EUS, (d) GRE, (e) SCO, and (f) RUS. Each graphic contains 1) time series of the Gaussian precipitation ($Y_{\mathcal{G}}$) against the Gaussian NAO index ($X_{\mathcal{G}}$; filled circles for 1951−77, open circles for 1978−2003); 2) contours of the corresponding joint PDF; and 3) linear and nonlinear prediction of $Y_{\mathcal{G}}$. On top of each panel: Smoothed graphics of the conditional RMSE of the linear (thin curve) and nonlinear prediction (thick curve) of $Y_{\mathcal{G}}$ from $X_{\mathcal{G}}$.

Table 2.2: Values of correlation $c$, test-side correlations $(t_M, t_-, t_+)$, asymmetry test $\mathcal{J}_c$, Gaussian MI $I_g$ and non-Gaussian MI ($I_{ng(EF)}$ estimator), information correlation $c_{\text{inf}}$, integral $P_{neg}$ (see definition in text), and MSE ($E^2$) of the linear and nonlinear prediction for Gaussian variables, i.e., subject to Gaussian anamorphosis (G) in the six selected points. The values of $I_g$, $I_{ng(ME)}$, and $\mathcal{J}_c$ are also added for the original untransformed data (O).

| | ATL | SCO | BAL | GRE | EUS | RUS |
|---|---|---|---|---|---|---|
| (G) $c_g$ | −0.640 | 0.670 | −0.170 | −0.310 | −0.070 | 0.690 |
| (G) $t_M$ | −0.730 | 0.680 | −0.130 | −0.400 | −0.280 | 0.680 |
| (G) $t_-$ | −0.630 | 0.430 | 0.190 | 0.140 | 0.240 | 0.660 |
| (G) $t_+$ | −0.300 | 0.820 | −0.650 | −0.440 | 0.350 | 0.730 |
| (G) $\mathcal{J}_c$ | 0.040 | 0.050 | −0.080 | −0.060 | 0.010 | 0.010 |
| (G) $I_{ng(EF)}$ | 0.100 | 0.080 | 0.060 | 0.150 | 0.020 | 0.010 |
| (G) $I_g$ | 0.270 | 0.300 | 0.020 | 0.050 | 0.003 | 0.320 |
| (G) $c_{\text{inf}}$ | 0.690 | 0.700 | 0.370 | 0.470 | 0.230 | 0.700 |
| (G) $P_{neg}$ | −0.011 | −0.001 | −0.006 | −0.001 | −0.002 | −0.006 |
| (G) $E^2(Y_{\text{lin}})$ | 0.590 | 0.550 | 0.980 | 0.910 | 1.000 | 0.530 |
| (G) $E^2(Y_{\text{nolin}})$ | 0.560 | 0.550 | 0.910 | 0.890 | 0.990 | 0.540 |
| (O) $\mathcal{J}_c$ | 0.040 | 0.070 | −0.070 | −0.050 | −0.010 | 0.040 |
| (O) $I_g$ | 0.320 | 0.280 | 0.010 | 0.060 | 0.003 | 0.290 |

one of the regimes, i.e. where only one of the test side correlations ($t_-$ or $t_+$) is significantly different from zero. In particular, the negative side test correlation is particularly strong over Ukraine, Romania and former Yugoslavia ($t_- \approx -0.6$), whereas the corresponding positive one is quite small ($t_+ \approx -0.2$). In the Mediterranean region, south of 40°N, the positive test correlation $t_+$ is negative, whereas the corresponding one in the negative side practically vanishes. This means that, in the South Mediterranean, while in the NAO- regime the statistical mean response of precipitation to NAO is not significant, in the NAO+ regime it is favorable to strong extreme drought events. This is consistent with the values $t_- = 0.19$ and $t_+ = -0.65$ for the BAL point. This is also apparent from the shape of contours (Fig. 2.3b) and from the nonlinear prediction graphic. This situation occurs especially on the second half of the analysed time interval (i.e. 1978−2003), and may have a connection with the increasing desertification over Mediterranean regions during that same period. These results agree with the positive NAO trend over the last two decades (Hurrell et al., 2004), with higher positive extremes in the corresponding index.

In south Greenland and Baffin Bay the driest conditions are again especially favored in the NAO+ regime, whereas the NAO index in the negative regime practically does not have any average statistical influence on precipitation. Consistent values of $t_- = 0.14$ and $t_+ = -0.44$ are given at the GRE point. This is rather visible from the completely different shape of PDF contours in the positive and negative regimes (Fig. 2.3d). This may be due to a nonlinear influence of NAO on the systems influencing the precipitation in Greenland, which must be synoptically analysed in further studies. The precipitation in Greenland is also correlated with the presence of other regimes, such as the Greenland-Scandinavian regime, influencing the strength of the Icelandic low (Serreze et al., 1997). We

have studied another particular situation where $t_-$ and $t_+$ have the same signal, opposite to that of $t_M$. This holds at the EUS point (Fig. 2.3c) with $t_- = 0.24$, $t_+ = 0.35$ and $t_M = -0.28$. Unlike other points of strong correlation, the $(X_\mathcal{G}, Y_\mathcal{G})$ distribution in the RUS point is rather close to bi-Gaussianity, as it is clear from Fig. 2.3f and Table 2.2. At this point $c_g = 0.69$.

The asymmetry measure $\mathcal{J}_c(X_\mathcal{G}, Y_\mathcal{G})$ for Gaussian data is presented in Fig. 2.4a. The largest values are significant at $\sim 95\%$ significance level (see Table 2.1). Some coherent regions of significant $\mathcal{J}_c$ are visible in the map over the Mediterranean and the Central Atlantic, and near 40°N, South Greenland and the Canadian Arctic coast.

### 2.6.2 Cumulants and Edgeworth PDF applicability

The contribution to non-Gaussianity comes from high-order cumulants, easily expressed in terms of nonlinear correlations, leading to non-zero cumulant terms in the Edgeworth expansion of the joint PDF (2.48) and the Negentropy (2.60). We show maps of the main cumulant terms of the Gaussian variables: $k^{(2,1)}$, $k^{(3,1)}$ (Figs. 2.4b-2.4c), intervening the most in (2.49) and (2.60). They are proportional to the nonlinear correlations $cor(X_\mathcal{G}^2, Y_r)$ and $cor(X_\mathcal{G}^3, Y_r)$ respectively. The other cumulants – $k^{(1,2)}$, $k^{(1,3)}$ and $k^{(2,2)}$ – are only residual (not shown).

The first and second mentioned correlations express the correlation between the residues of the precipitation linear prediction and, respectively, the squares $(X_\mathcal{G}^2)$ and cubes $(X_\mathcal{G}^3)$ of the Gaussian NAO index $X_\mathcal{G}$. These correlations express, respectively, how a quadratic or a cubic function of NAO fits those residues. The cumulant $k^{(2,1)}$ is dominant over Greenland, the Mediterranean and the Atlantic Central Basin. The corresponding spatial dependence is closely related to the asymmetry test $\mathcal{J}_c$ map (Fig. 2.4a), with a map correlation of 0.93 over the

NAE region. This is explained because the nonlinear correlations $cor(X_{\mathcal{G}}^2, Y_r)$ and $cor(|X_{\mathcal{G}}|, Y_r)$, related with $\mathcal{J}_c$ (see sections 2.2.2 and 2.2.3), behave in the same way.

The cumulants $k^{(3,1)}$ and $k^{(2,1)}$ also contribute to the nonlinear prediction (2.62). Note, in particular for the Central Atlantic, that the stronger correlation in the negative side (NAO-) is consistent with a fitting predictive curve formed from a negative slope straight line [resulting from the negative correlation $c(X_{\mathcal{G}}, Y_{\mathcal{G}})$], a convex parabolic fitting curve [resulting from the positive $cor(X_{\mathcal{G}}^2, Y_r)$, Fig. 2.4b] and cubic curve dependent on $X_{\mathcal{G}}^3$ [resulting from the positive $cor(X_{\mathcal{G}}^3, Y_r)$, Fig. 2.4c]. Given the polynomial nature of the PDF-EDG estimator, it can exhibit negative values, thus violating the PDF-positiveness. In order to quantify this aspect, we compute $P_{neg}$ (integral of PDF-EDG in the domain of negative values) and show its map in Fig. 2.4d. The maxima of $|P_{neg}|$ are reached in the central Atlantic region ($\sim 0.014$) and south of Iceland ($\sim 0.008$). These small values constitute a necessary albeit not a sufficient condition for the EDF-PDF to be a good representation of the real PDF.

### 2.6.3 Mutual Information

The map of the Gaussian Mutual Information (MI) $I_g(X_{\mathcal{G}}, Y_{\mathcal{G}})$ (Fig. 2.5a) is obtained from that of $c_g$ with two cores of maxima reaching 0.4 *nat* near 60ºN, 10ºW, and 0.30 *nat* near 35ºN, 15ºW. The non-Gaussian MI is computed using the proposed estimators: the Edgeworth estimator of (2.60), $I_{ng(EF)}$, with the Gaussianised data thus preventing marginal Negentropies (Fig. 2.5b) and that obtained with the integral (2.57): $I_{ng(MI)}$, estimated by Gaussian quadrature. The correspondent field is practically identical to that of $I_{ng(EF)}$ in Fig. 2.5b.

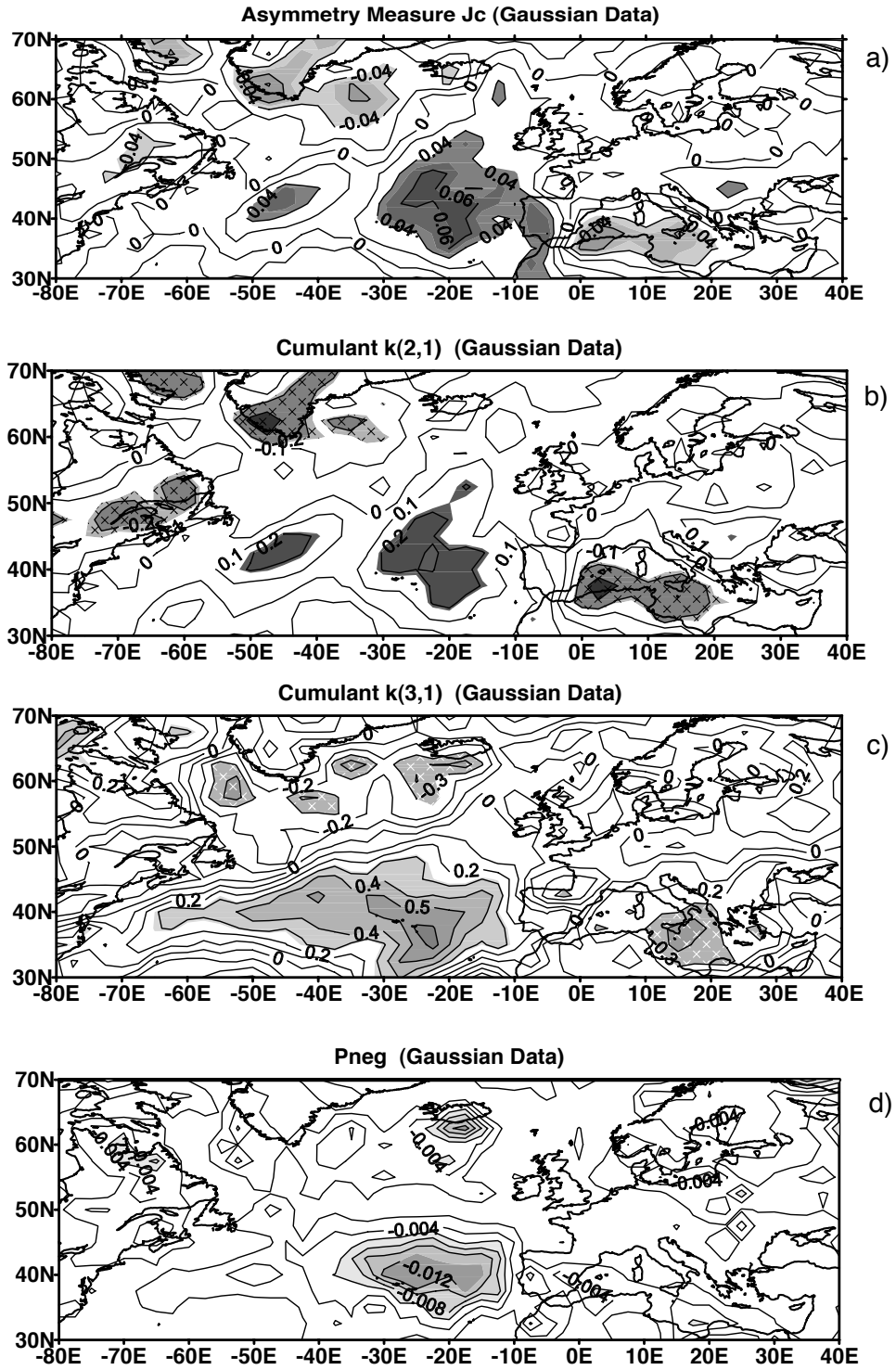



Figure 2.4: (a) Map of the asymmetry $\mathcal{J}_c$ (CI = 0.02; SR at $\alpha = 90\%$ shaded), (b) map of the cumulant $k^{(2,1)}$ (CI = 0.1; SR at $\alpha = 90\%$ shaded), (c) map of the cumulant $k^{(3,1)}$ (CI = 0.1; SR at $\alpha = 90\%$ shaded), and (d) map of $P_{neg}$ (CI = 0.002). All quantities computed are subjected to Gaussian anamorphosis. The 90% significant area fractions FS-MCR are 0.17, 0.17, and 0.19 for (a), (b), and (c), respectively. Corresponding values of FS-MCG are 0.16, 0.13, and 0.15.

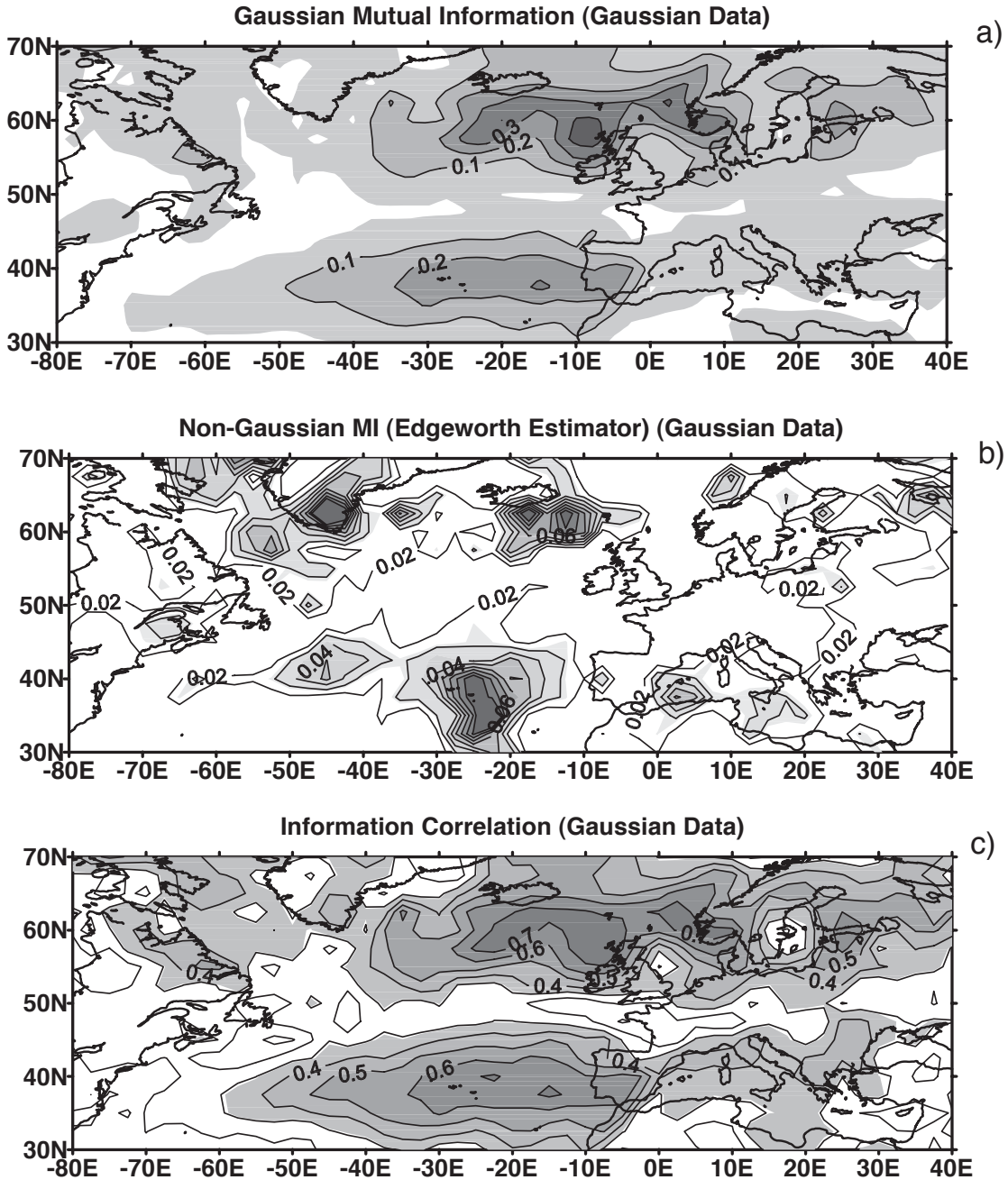



Figure 2.5: Map of the Gaussian mutual information $I_g$ (CI = 0.1; SR at $\alpha = 90\%$ shaded), (b) map of the non-Gaussian MI (Edgeworth estimator) (CI = 0.01; SF at 90% shaded), (c) map of the information correlation $c_{\text{inf}}$. (CI = 0.1; SR at 90% shaded). All quantities are computed for Gaussian variables. The 90% significant area fractions FS-MCR are 0.73, 0.23, 0.22, and 0.64 for (a), (b), (c), and (d), respectively. Corresponding values of FS-MCG are 0.73, 0.12, 015, and 0.57.

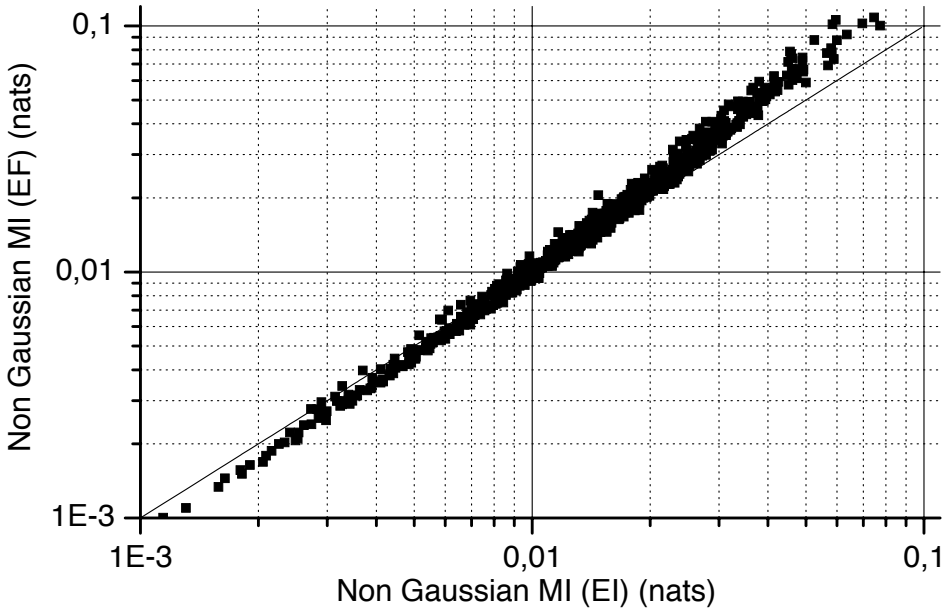



Figure 2.6: Non-Gaussian MI $I_{ng(EF)}$ estimator as function of the ascending order sorted values of non-Gaussian MI $I_{ng(EI)}$.

As expected, $I_{ng}$ and the asymmetry measure $\mathcal{J}_c$, share some common regions, since the asymmetry contributes for Non Gaussianity. In order to assess the effect of the Taylor approximation of the logarithm of the EDG-PDF (2.58), we have sorted, in ascending order, all the values of the integral value $I_{ng(MI)}$ in the 833 grid-points of the NAE region and plotted them against the correspondent values of $I_{ng(EF)}$ (Fig. 2.6a).

The estimators agree with each other within the error of the Gaussian quadrature ($\sim 10^{-4}$ *nat*) up to $I_{ng} \approx 0.02$ *nat*, followed by a slight overestimation by the Edgeworth equation (2.60) of MI [$I_{ng(EF)}$]. The map correlation between the two estimators is 0.98. The Central-Atlantic, South-Iceland and Greenland "Non-Gaussian" regions are retrieved by $I_{ng(EF)}$, where too "spiky" values are verified and $P_{neg}$ reaches the maximum values.

### 2.6.4 The Effect of Gaussian Anamorphosis

Near the centroids of maximum Gaussian correlation, 40°N, 20°W and 60°N, 10°W, we have verified that the side or asymmetric correlation is more intense in the "wet" NAO phase than in the "dry" NAO phase. If no Gaussian Anamorphosis (GA) is performed, there is an enhancement of the side correlation over the positively skewed half of the precipitation PDF in comparison with the marginally Gaussian variables. This can be seen through the larger intensity of the asymmetry measure $\mathcal{J}_c$ (Fig. 2.7a) as compared to the Gaussian case (Fig. 2.4a).

The Gaussian Anamorphosis is a nonlinear transformation that can either increase or decrease the absolute correlation $|c|$ between variables and thus the Gaussian MI. For example, GA, when performed with appropriate mixtures of two bivariate Gaussian PDFs, one with zero correlation and the other with correlation $\sim 1$, can convert a nearly zero correlation into a nearly 100% correlation and vice-versa. In our case, the change of correlation resulting from GA is not very high (maximum of $\sim 10\%$). However, given that the derivative of the Gaussian MI ($I_g$) grows from zero to infinity when the absolute correlation $|c|$ tends to one, a small change of $|c|$ can make a large difference in $I_g$. The $I_g$ difference between original and Gaussian data is plotted in Fig. 2.7b with a positive maximum of $\sim 0.05$ *nat* near 40°N,20°W, where $|c|$ grows from $\sim 0.64$ to $\sim 0.69$. Comparing with the case of marginally Gaussian data, the $I_{ng}$ in the Mediterranean is preserved in the original data (Fig. 2.7c) and another large area appears around the White Sea and Kola Peninsula at approximately 60°N, 35°E.

The invariance (2.43) of MI is hardly verified for the estimator $I_{ng(EF)}$ except for low values of the joint Negentropy, which was expectable given the limits of the Edgeworth expansion. The original highly skewed precipitation values render

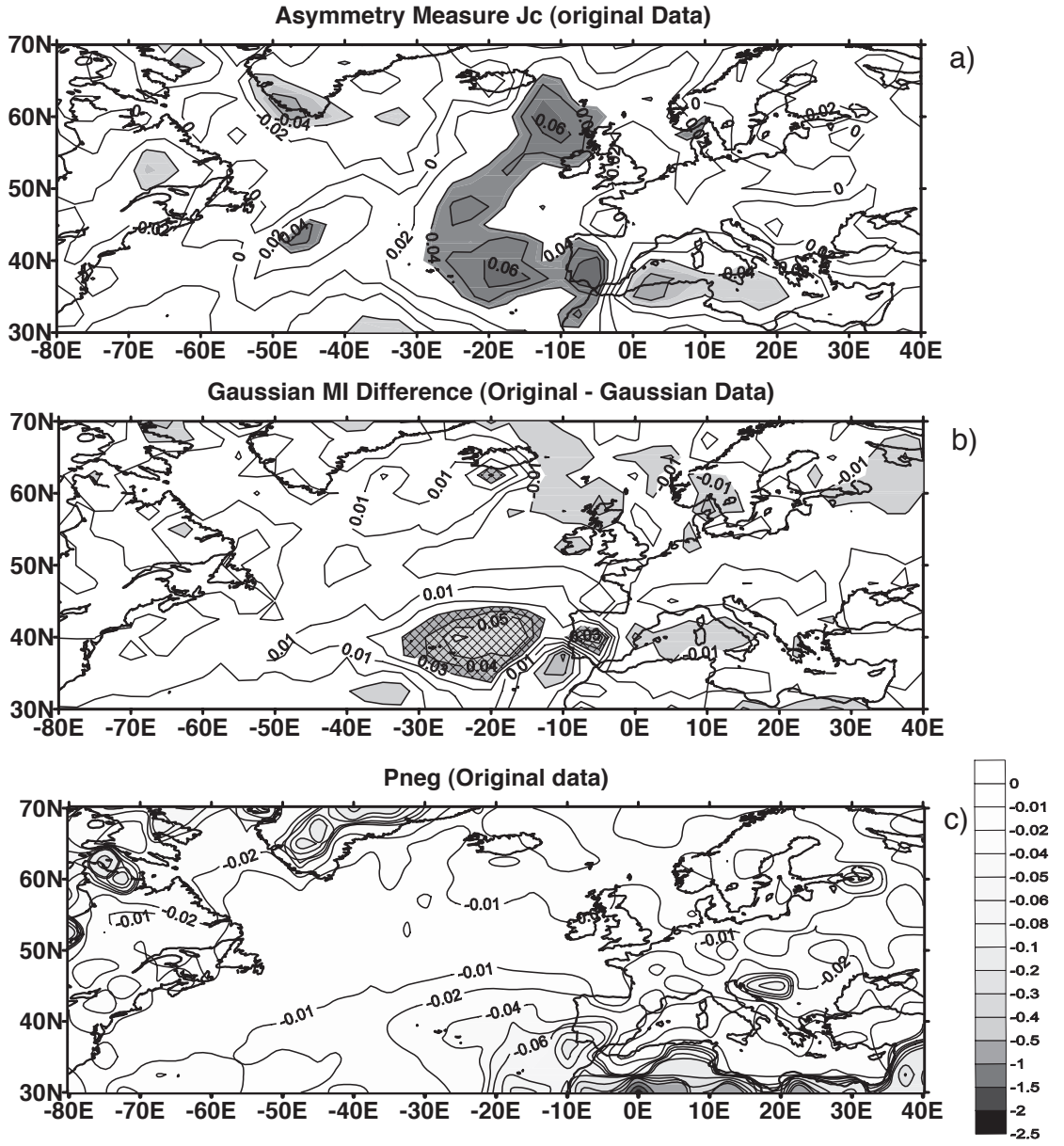



Figure 2.7: (a) Map of the asymmetry test $\mathcal{J}_c$ for original data (CI = 0.02; SR at $\alpha = 90\%$ shaded), (b) Gaussian MI difference between Gaussian and original data, (c) map of $P_{neg}$ for original data. The 90% significant area fractions FS-MCR is 0.17 for (a). The corresponding value of FS-MCG is 0.13.

the EDF-PDF inapplicable for much of the NAE region due to the much higher values of $|P_{neg}|$ (above 0.01), Fig. 2.7d, as compared to those of "Gaussianised" (previously rendered Gaussian) variables (Fig. 2.4d).

## 2.7 Discussion and Conclusions

An asymmetry measure of bivariate probability distributions has been constructed and presented in this chapter. This method computes the conditional correlations for each half of the sorted data, denoted in the thesis as asymmetric or side correlations. The comparison between side correlations and the global correlation, under the hypothesis of bivariate Gaussianity, has led us to the construction of several robust tests of PDF asymmetry. Asymmetry contributes to non-Gaussianity by giving extra information beyond that given by a Gaussian PDF. Consequently, we have aimed at evaluating Mutual Information (MI) and their Gaussian and non-Gaussian counterparts.

For that purpose, an estimator has been proposed based on the Edgeworth expansion (EDG) of the joint PDF and MI in terms of cumulants and Hermite polynomials, applicable only for sufficiently small values of the joint Negentropy and cumulants.

The Mutual Information term due to non-Gaussianity has been computed for two sets of variables: the original ones and those coming from the application of Gaussian Anamorphosis to the original variables in order to mitigate the effect of marginal outliers in the cumulants and thus render the Edgeworth method applicable in a much better way.

The methods have been applied to the PC-based NAO index as a large-scale predictor variable and to the gridpoint monthly DJF precipitation over the NAE as downscaled predictand variables.

From numerical computations we have verified that the response of monthly precipitation to NAO is asymmetric and non-Gaussian. Maps of both test side correlations, for NAO- and NAO+ regimes, show coherent regions and consistent regional differences, thus highlighting the asymmetric precipitation response to NAO. Within the main extreme correlation centers, between NAO and precipitation, the side correlation is stronger in the "wetter" NAO regime, and enhanced if no Gaussian anamorphosis is performed over the precipitation field. In other regions such as the Mediterranean or Greenland, the sensitivity of precipitation in the NAO- regime practically vanishes, whereas the correlation in the "drier" regime (NAO+) is significantly negative.

The Mutual Information (MI) provides a general measure of statistical redundancy. As far as the untransformed variables (NAO, precipitation) are concerned, the larger bulk of Mutual Information (MI) comes from its Gaussian part with some exceptions over areas such as Greenland where the non-Gaussian part of MI and nonlinearity are relevant. The Mutual Information term due to non-Gaussianity is relevant in some areas such as Mediterranean, the Southern part of Greenland, the southeast of Iceland, the area around 42$^o$N, 48$^o$W and regions around the White Sea and Kola Peninsula. For Gaussian data, a maximum of non-Gaussian MI appears in the Central Atlantic Basin due to local decrement of the global correlation produced by the Gaussian anamorphosis.

The EDG method can be generalised to multivariate data providing estimates of the joint and conditional PDF and moments. Extensions using higher order cumulants are also possible. However, in order to avoid overfitting, it is preferable that the PDF calibration and validation be made in cross-validation mode.

The EDG method is able to give an indication of non-Gaussianity if variables are previously constrained to significantly reduce the magnitude of cumulants.

# Chapter 3

# Nonlinear Dynamics and Predictability

## 3.1 Nature of the problem

It is widely believed among the general public and sectors within the scientific community that, given the state of a system at a certain instant, there must exist an algorithm that allows for the evolution of the system to be predicted indefinitely. This view is grounded on the *Newtonian paradigm* (Nicolis and Nicolis, 2007), whereby the universe is governed by precise physical laws allowing for the state of a system to be known everywhere and at all times.

With the onset of Quantum Physics, limitations to the classic approach were exposed in such situations as those evoked by the Heisenberg uncertainty principle (Gasiorowicz, 1996), whereby the precise position and momentum of a particle could not simultaneously be known.

However, even then the aforementioned paradigm prevailed, as a new phenomenology was brought up that allowed for the system to be accurately represented and indefinitely predicted through precise physical laws. In fact, the classical concept of a time-varying state of a particle – the dynamics of which would no longer be rigourously described – was replaced by a quantum state

rigourously characterised by a *wave function* containing all the information it is possible to obtain about the particle, and interpreted as a *probability amplitude of the particle's presence* (Cohen-Tanoudji et al., 1977).

While the Newtonian paradigm has proven its successes in extreme scales − very large, cosmological ones on one hand and very small, subatomic on the other − the same has not happened in the intermediate scales, such as those of concern in the geosciences and in our everyday lives (Nicolis and Nicolis, 2007). Here, nonlinear interactions and intrinsic structural instabilities severely limit the horizons of predictability. The problem is further aggravated as higher-dimensional systems are considered, thus imposing fundamental limits on the ability to forecast such relevant phenomena as geophysical flow.

### 3.1.1 The 'nemesis' of atmospheric predictability

It is widely admitted that there are two main reasons for which prediction in Meteorology is limited in time: the amplification in the course of the evolution of small uncertainties in the initial conditions used in a prediction scheme, usually referred as *initial errors*; and the presence of *model errors* reflecting the fact that a model is only an approximate representation of nature. Both the initial and the model errors have systematic and random components. While the first kind of error is indicative of the property of sensitivity to the initial conditions, suggesting that atmospheric dynamics share in this respect some key properties of deterministic chaos, the second one is an indicator of the property of sensitivity to the parameters and more generally of structural stability of the set of evolution laws governing the system at hand (Nicolis, Perdigão, and Vannitsem, 2009). A dynamical system is structurally stable at a parameter vector $\mu_p$ if the correspondent phase portrait remains topologically equivalent for small changes

of the parameter vector. On the contrary, we will have a bifurcation at $\mu_p$ and structural instability.

There exists an extensive literature on both initial and model errors, much of which is devoted to numerical experiments on large scale numerical forecasting models (Dalcher and Kalnay, 1987; Tribbia and Baumhefner, 1988, 2004; Reynolds et al., 1994; Schubert and Schang, 1996; Krishnamurti et al., 2004; Chu and Ivanov, 2007). While in practice the two sources of errors coexist and their respective effects on the results cannot be clearly identified, in most of the analyses reported they are treated separately (Lorenz, 1996; Nicolis, 1992; 2003; 2004).

The fundamental objective of the present chapter is to address this 'nemesis' of atmospheric predictability in a systematic, quantitative manner. For that purpose, we begin by reviewing basic concepts in the theory of dynamical systems (section 3.2). Then, we address some generic, model-independent features of the dynamics of prediction errors under the combined effect of initial and model errors and their connections with intrinsic system properties (section 3.3). Furthermore, the crossovers between the two kinds of errors and between the initial and intermediate time regimes are also considered. Illustrative case studies are then considered in section 3.4.

## 3.2 An overview on Dynamical Systems

Dynamical systems are often mistakenly regarded as eminently chaotic systems. In reality, systems exhibiting chaotic behaviour are just part of a broader family of dynamical systems. It is thus crucial to rigourously define what dynamical systems are after all.

Following Ott (2002), a *dynamical system* may be defined as a deterministic mathematical prescription for evolving the state of a system forward in time. One will thus be in the presence of a dynamical system when, for any *exact* initial state of the system, one can in principle solve the equations that govern its dynamics to obtain its future states − in a perfect model scenario, naturally.

Mathematically speaking, a dynamical system can be represented in the form:

$$\mathbf{x}(t) = \mathbf{F}[\mathbf{x}(0), \mu] \tag{3.1}$$

where $\mathbf{x}(t) = (x_1, \cdots, x_n)$ is the state at time $t$, $\mathbf{x}(0)$ is the initial state; $\mu$ represents a set of *control parameters* $(\mu_1, \cdots, \mu_n)$ through which the system communicates with the environment, and $\mathbf{F} = (F_1, \cdots, F_n)$ is a smooth function such that for each given $\mathbf{x}(0)$ there is only one $\mathbf{x}(t)$ (Nicolis and Nicolis, 2007).

At this stage one might consider a co-dimension (number of parameters) $n'$ not necessarily equal to the dimension $n$ of the system. However, for the sake of algorithmic simplicity in the upcoming applications, those are taken equal.

The space generated by all the possible system states is denoted as *phase space*, and the path in that space followed by the system as it evolves with time is referred to as an *orbit* or *trajectory*. In conservative systems the phase space volume is invariant. In dissipative systems, that volume decreases with time. Dissipative systems are typically characterised by the presence of attracting sets or *attractors* in phase space. The attracting sets are bounded subsets to which regions of initial conditions of nonzero phase space volume tend asymptotically as time increases (Ott, 2002). In particular, an attractor is an attracting set with a dense orbit.

An apparent contradiction can arise as one refers to chaotic behaviour within deterministic systems. In fact, it is commonly believed, outside of the specialised

community, that chaotic systems cannot be deterministic, or, conversely, that deterministic systems cannot exhibit chaotic behaviour. This mistaken belief is based on the naive perception that deterministic systems must be indefinitely predictable.

However, it is clear, from the dynamical system definition presented above, that it encompasses every system that can be deterministically predicted – from regular, periodic systems such as a harmonic oscillator, to complex deterministic systems exhibiting chaotic behaviour. In other words, if, in a deterministic dynamical system, the *exact* state of the system is known at a certain instant, then its future states can also be determined in a perfect-model scenario, i.e. with unperturbed control parameters – no matter how unstable the system is.

Nevertheless, the same will no longer occur if the knowledge of the initial state is not exact, even in a perfect model scenario and even if the difference between the measured state and the exact one is smaller than the precision of the most advanced method of experimental measurement and than the roundoff numerical errors. In fact, by taking such experimentally indistinguishable initial states, as time unfolds the upcoming trajectories will exponentially diverge in systems with instabilities. We are thus in the presence of *deterministic chaos* (Nicolis and Nicolis, 2007).

Why deterministic? For each exact initial condition and control parameters there is a corresponding exact state at a certain future instant. Why chaos? Small differences between initial conditions can lead to huge differences in corresponding future states. Small perturbations in the parameters can lead to completely different asymptotic behaviours (as the attractor properties depend on those parameters).

In the absence of parameter deviations, the dynamical behaviour subsequently

followed by initial perturbations as they evolve in time – e.g. whether they will grow or fade – is mainly driven by *Lyapunov exponents*, which give a means of characterising the stretching and contracting characteristics of attractors and other invariant sets (Ott, 2002).

The *Lyapunov exponents* can formally be defined as (Nicolis and Nicolis, 2007):

$$\lambda(\mathbf{x}_0) = \lim_{\substack{|\delta\mathbf{x}|\to 0 \\ t\to\infty}} \frac{1}{t} \ln \frac{|\delta\mathbf{x}(t)|}{|\delta\mathbf{x}(0)|} \tag{3.2}$$

where $|\delta\mathbf{x}(0)|$ is the distance between two considered initial conditions, and $|\delta\mathbf{x}(t)|$ the distance between the states to which those two initial conditions have evolved, governed by a dynamical system as in (3.1).

The Lyapunov exponents can be computed either for the whole attractor (global Lyapunov exponents), independently of the initial condition, or can be interpreted pointwise, where the growth ratio is evaluated for a finite time interval at $\mathbf{x}_0$ and $t$. Lyapunov vectors are vectors collinear to the small deviations $\delta x_0$. Successive Lyapunov exponents in descending order are obtained by constraining the small deviations $\delta x_0$ to be orthogonal to the previous Lyapunov vectors.

Negative Lyapunov exponents in a certain phase space direction will imply a convergence among trajectories in that direction (*stable direction*). Conversely, positive Lyapunov exponents will imply divergence (*unstable direction*). For chaos to occur, there must be at least one positive Lyapunov exponent in the system. In conservative systems, all Lyapunov exponents are null – in agreement with the volume-conserving dynamics of such systems.

While the Lyapunov exponents play a key role in the divergence of trajectories initially separated even by a small difference, trajectories can drift apart even if every trajectory begins at exactly the same point in phase space, if there are perturbations in the system itself. By considering the definition (3.1), such

perturbations are expressed as differences in control parameters $\mu$, which affect the *structural stability* of the system.

In order to better understand this concept, consider a reference phase trajectory $\Gamma_R$ describing a particular long-time behaviour of the system at hand. This trajectory lies necessarily on an invariant set like an attractor. If the system is perturbed, the trajectory effectively followed will be a displaced one, $\Gamma$, whose instantaneous displacement from $\Gamma_R$ is given by $\delta\mathbf{x}(t)$ at time $t$. Three situations may then occur:

a) If, for every bounded $\delta\mathbf{x}(0)$ the corresponding $\delta\mathbf{x}(t)$ will not be bounded no matter how small the initial displacement is, then the reference trajectory $\Gamma_R$ is *unstable*.

b) If, for every bounded $\delta\mathbf{x}(0)$ the corresponding $\delta\mathbf{x}(t)$ can be bounded, then the reference trajectory $\Gamma_R$ is *stable*.

c) If $\Gamma_R$ is *stable* and, in addition, the system eventually dampens the perturbations thereby returning to the reference state, $\Gamma_R$ is *asymptotically stable* (Nicolis and Nicolis, 2007).

From the overview on dynamical systems given in the present section, it is clear that prediction errors can be marred by perturbations in the initial conditions and in the control parameters intrinsic to the system.

The following step is to derive a systematic, analytical formulation that assesses the dynamics of prediction errors under the combined influence of initial condition and parameter errors. This is the subject of the next section.

## 3.3 Dynamics of prediction errors

The transient evolution of prediction errors in the short to intermediate time regime is hereby addressed, considering the combined effect of initial condition and model errors. Some generic features are brought out and connected with intrinsic system properties.

### 3.3.1 General formulation on a Hilbert space

Consider a nonlinear system governed by nonlinear evolution laws with continuous derivatives. The system is subjected to small perturbations relative to its underlying attractor and parameters. This is done in such a way that the variables representing the unperturbed and perturbed systems span the same manifold defined by a Hilbert space $\mathcal{H}$.

Let the unperturbed model, hereafter denoted as "nature", be driven by the following dynamics:

$$\frac{d\mathbf{x}_N}{dt} = \mathbf{f}(\mathbf{x}_N, \mu_N) \tag{3.3}$$

where $\mathbf{x}_N$ denotes the nature state-vector and $\mu_N$ the unperturbed parameters, and let the perturbed model, hereafter denoted simply as "model", be driven by:

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mu) + \delta\mathbf{f} \tag{3.4}$$

where $\mathbf{x}$ is the model state-vector, $\mu$ are the perturbed parameters and $\delta\mathbf{f}$ represents the additive deviation from the model formulation relative to the true dynamics (nature) it intends to represent. The functions $\mathbf{f}$ are generally nonlinear, though respecting smoothness properties as to have continuous derivatives in their whole domain.

We want to estimate the error between the solutions of the model, Eq. (3.4) and the evolution laws corresponding to nature, Eq. (3.3).

For that purpose, we begin by writing the relationships between model and nature variables and parameters as:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}_N + \mathbf{u} \\ \mu &= \mu_N + \delta\mu \end{aligned} \tag{3.5}$$

where $\mathbf{u}$ and $\delta\mu$ denote, respectively, errors in the variables and in the parameters.

As in Nicolis, Perdigão, and Vannitsem (2009), we place ourselves under the conditions of weakly imperfect model $|\delta\mu| \ll |\mu_N|$ and small initial errors $|\mathbf{u}_0| \ll |\mathbf{x}_N|$. We also assume structural stability of the underlying evolution laws. This entails, in particular, that the system is not crossing criticalities of any sort in the range of variations of the parameters caused by the error $\delta\mu$. As a corollary, the attractors of the model and reference system are close in phase space. As it will turn out, adopting this setting will allow us to conduct a systematic study and identify some features of error dynamics independent of the particular model considered, which could thus be qualified in this sense as "generic". In many operational-oriented numerical experiments on present day realistic weather prediction models these conditions (as well as some other more technically oriented ones enunciated in the sequel) may not be satisfied. Even so, having an idealised limiting case like the one considered here as a reference is helpful in the sense that possible deviations from the generic behaviours predicted by our analysis can be placed in the proper perspective and attributed to such factors as large initial errors or inadequacies in the parameterisations of some of the physical processes present.

The error evolution laws can then be obtained by expanding Eq. (3.4) in $\mathbf{u}$ and $\delta\mu$, subtracting (3.3) from the result and keeping the first nontrivial terms, leading us to:

$$\frac{d\mathbf{u}}{dt} = \mathbf{J}\mathbf{u} + \mathbf{p} \tag{3.6}$$

where

$$\begin{aligned} \mathbf{J} &= \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)_{\mathbf{x}_N,\mu_N} \\ \mathbf{p} &= \left(\frac{\partial \mathbf{f}}{\partial \mu}\right)_{\mathbf{x}_N,\mu_N} \delta\mu + \delta\mathbf{f} = \mathbf{\Phi}\delta\mu + \delta\mathbf{f} \end{aligned} \tag{3.7}$$

$\mathbf{J}$ is the Jacobian of $\mathbf{f}$ relative to the state vector, $\mathbf{\Phi}$ is is the Jacobian of $\mathbf{f}$ relative to the parameters, $\mathbf{p}$ is the tendency error, and the subscript $\mathbf{x}_N, \mu_N$ implies evaluation at $\mathbf{x} = \mathbf{x}_N$ and $\mu = \mu_N$.

The formal solution of the inhomogeneous (3.6) at time $t = \tau$ (forecasting time) is given by:

$$\mathbf{u}(\tau) = \mathbf{M}_{\tau,0}\mathbf{u}_0 + \int_0^\tau \mathbf{M}_{\tau,t}\,\mathbf{p}\,dt = \mathbf{u}_I(\tau) + \mathbf{d}(\tau) \equiv \mathbf{u}_I + \mathbf{d} \tag{3.8a}$$

where $\mathbf{u}_0 \equiv \mathbf{u}(\tau = 0)$ is the initial error, $\mathbf{u}_I(\tau)$ is the contribution due to error in the initial conditions and $\mathbf{d}(\tau)$ is that due to model error or drift.

The fundamental (resolvent) matrix $\mathbf{M}_{\tau,0}$ is the propagator operator satisfying the relation:

$$\frac{d\mathbf{M}_{\tau,t}}{dt} = \mathbf{J}\,\mathbf{M}_{\tau,t} \tag{3.8b}$$

the integral form of which is given by:

$$\mathbf{M}_{\tau,t} = \exp\left[\int_t^\tau \mathbf{J}(t')\,dt'\right], \quad t \leq \tau \tag{3.8c}$$

with $\mathbf{M}_{t,t}$ equal to the identity for all $t$.

The error formulation will now be expressed for a given metric. For this purpose, consider the inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^* \tag{3.9}$$

*(the superscript * denotes the complex conjugate)*

supported on a complete numerable basis of the orthonormalisation vector $\phi_\alpha$,

such that:

$$\begin{aligned} \langle \phi_\alpha, \phi_\alpha \rangle &= 1 \\ \langle \phi_\alpha, \phi_{\alpha'} \rangle &= 0, \text{ if } \alpha \neq \alpha' \end{aligned} \tag{3.10}$$

(e.g. the spherical harmonics over the sphere).

For the purpose of defining the square error, consider the norm:

$$N(\mathbf{u}) = ||\mathbf{u}^2(\tau)|| = \langle \mathbf{u}(\tau), \mathbf{A}\mathbf{u}(\tau) \rangle, \tag{3.11}$$

where $\mathbf{A}$ is a self-adjoint operator defined semi-positive. The semi-positiveness is useful for the case of defining the square error over a subspace of $\mathcal{H}$, such as when considering spectral truncation. The operator $\mathbf{A}$ is defined from the positive-defined operator $\mathbf{B}$ for which the same orthogonal basis can be defined:

$$\langle \phi_\alpha, \mathbf{B}\, \phi_{\alpha'} \rangle = 0, \text{ if } \alpha \neq \alpha' \tag{3.12}$$

Now let $\mathbf{\Pi}$ be a projection operator over a sub-space of the basis functions $\phi_\alpha$. $\mathbf{\Pi}$ is thus self-adjoint and commutes with $\mathbf{B}$. Therefore, $\mathbf{\Pi}\,\mathbf{B} = \mathbf{B}\,\mathbf{\Pi}$ is self-adjoint and defined semi-positive, just like $\mathbf{A}$. One can thus take $\mathbf{A} = \mathbf{\Pi}\,\mathbf{B} = \mathbf{B}\,\mathbf{\Pi}$.

For instance, when $\mathbf{u}$ is the perturbed streamfunction (as in chapter 4), one obtains the metric of square streamfunction for $\mathbf{B} = \mathbf{I}$ (identity), the metric of kinetic energy for $\mathbf{B} = -\frac{1}{2}\nabla^2$ and the metric of enstrophy for $\mathbf{B} = -\frac{1}{2}\nabla^4$.

From the decomposition in (3.8a), the square error becomes:

$$\begin{aligned} ||\mathbf{u}^2(\tau)|| &= \langle \mathbf{u}_I, \mathbf{A}\mathbf{u}_I \rangle + 2\,\Re\left(\langle \mathbf{u}_I, \mathbf{d} \rangle\right) + \langle \mathbf{d}, \mathbf{A}\mathbf{d} \rangle \\ &= \langle \mathbf{u}_I, \mathbf{A}\mathbf{u}_I \rangle + \langle \mathbf{d}, \mathbf{d} \rangle + 2\langle \mathbf{u}_I, \mathbf{A}\mathbf{d} \rangle \end{aligned} \tag{3.13}$$

where $\Re$ refers to the real part of the term onto which it operates. We see that the two parts $\mathbf{u}_I$ and $\mathbf{d}$ of the error interact on an instantaneous basis. The purpose of such methods as Model Output Statistics (MOS) is that they interact destructively, say $\langle \mathbf{u}_I, \mathbf{d} \rangle < 0$.

The above terms are written as:

$$\begin{aligned}\langle \mathbf{u}_I, \mathbf{A}\,\mathbf{u}_I\rangle &= \langle \mathbf{u}_0, \mathbf{M}^*_{\tau,0}\,\mathbf{A}\,\mathbf{M}_{\tau,0}\,\mathbf{u}_0\rangle \\ &= \langle \mathbf{u}_0, \mathbf{O}_{\tau,0,0}\,\mathbf{u}_0\rangle,\end{aligned} \tag{3.14}$$

where we define the generalised self-adjoint Oseledec operator, defined semi-positive in general (and positive for $\mathbf{A} = \mathbf{I}$):

$$\mathbf{O}_{\tau,t',t''} = \mathbf{M}^*_{\tau,t'}\,\mathbf{A}\,\mathbf{M}_{\tau,t''}, \tag{3.15}$$

where $\mathbf{M}^*_{\tau,t'}$ is the adjoint operator of $\mathbf{M}_{\tau,t'}$ with respect to the above-defined inner product $\langle\cdot,\cdot\rangle$, say $\langle \mathbf{a}, \mathbf{M}\,\mathbf{b}\rangle = \langle \mathbf{M}^*\,\mathbf{a}, \mathbf{b}\rangle,\ \mathbf{a},\mathbf{b} \in \mathcal{H}$. At $\tau = 0$, $\mathbf{O} = \mathbf{A}$.

The term referring to the interaction between errors due to initial conditions and those related to model imperfections becomes:

$$\begin{aligned}2\langle \mathbf{u}_I, \mathbf{d}\rangle &= 2\left\langle \mathbf{u}_0, \int_0^\tau \mathbf{O}_{\tau,0,t}\,\mathbf{p}(t)\,dt\right\rangle \\ &= 2\int_0^\tau \langle \mathbf{u}_0, \mathbf{O}_{\tau,0,t}\,\mathbf{p}(t)\rangle\,dt,\end{aligned} \tag{3.16}$$

where one introduces the integral of $\mathbf{O}$ due to the concavity of the propagator operator and its adjoint. If the series of orthogonal expansions are uniformly convergent, then one can permute the integral with the inner product in (3.16). It occurs for example in the expansion of continuous functions into spherical harmonics as we will consider in chapter 4.

Finally, the term due to the drift is given by:

$$\langle \mathbf{d}, \mathbf{d}\rangle = \int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), \mathbf{O}_{\tau,t',t''}\mathbf{p}(t'')\rangle dt'dt'' \tag{3.17}$$

where we have applied the linearity propagator of the inner product and assumed inner product of integrals as being equivalent to integrals of inner products.

### 3.3.2 Error dynamics along Lyapunov vectors

The analysis of the error dynamics along Lyapunov vectors allows us to outline the formulation for the error to be written in specific directions according to their local stability, e.g. in locally stable or unstable directions.

Let $\mathbf{\Lambda}_{\alpha,t''}$ be the set of finite normalised Lyapunov vectors (i.e. such that $\langle \mathbf{\Lambda}, \mathbf{\Lambda} \rangle = \mathbf{1}$), optimised between $t''$ and $\tau$, with $\alpha$ being the index referring to each vector.

Then:

$$\mathbf{O}_{\tau,t'',t''} \mathbf{\Lambda}_{\alpha,t''} = \eta_{\alpha,t''} \mathbf{\Lambda}_{\alpha,t''}, \tag{3.18}$$

where $\eta_{\alpha t''} = \exp\left[\lambda_{\alpha,t''}\left(\tau - t''\right)\right] = \eta_{\alpha^*_{\alpha t''}}$ (real eigenvalues), and $\lambda_{\alpha t''}$ is the finite Lyapunov exponent between $t''$ and $\tau$.

At this point, we take into account that the propagator $\mathbf{M}$ has the associative property, i.e.:

$$\begin{aligned} \mathbf{M}_{\tau t'} &= \mathbf{M}_{\tau t''} \mathbf{M}_{t'' t'} \\ \mathbf{M}^*_{\tau t'} &= \mathbf{M}^*_{t'' t'} \mathbf{M}^*_{\tau t''}. \end{aligned} \tag{3.19}$$

Therefore, the Oseledec operator becomes:

$$\begin{aligned} \mathbf{O}_{\tau,t',t''} &= \mathbf{M}^*_{\tau t'} \, \mathbf{A} \, \mathbf{M}_{\tau t''} \\ &= \mathbf{M}^*_{t'' t'} \left(\mathbf{M}^*_{\tau t''} \, \mathbf{A} \, \mathbf{M}_{\tau t''}\right) \\ &= \mathbf{M}^*_{t'' t'} \, \mathbf{O}_{\tau,t'',t'} \end{aligned} \tag{3.20}$$

Let us assume the expansion of $\mathbf{b}$ given by $\mathbf{b} = \sum_\alpha \mathbf{b}_\alpha \mathbf{\Lambda}_{\alpha,t''}$ with $\mathbf{b}_\alpha = \langle \mathbf{b}, \mathbf{\Lambda}_{\alpha,t''} \rangle$.

Then, by applying that expansion and taking into account Eqs. (3.18–3.20), the following will thus hold:

$$\begin{aligned} \langle \mathbf{a}, \mathbf{O}_{\tau,t',t''} \mathbf{b} \rangle &= \langle \mathbf{a}, \mathbf{M}^*_{t'' t'} \mathbf{O}_{\tau,t'',t'} \mathbf{b} \rangle \\ &= \langle \mathbf{M}_{t'' t'} \mathbf{a}, \mathbf{O}_{\tau,t'',t'} \mathbf{b} \rangle \\ &= \sum_\alpha \eta_{\alpha,t''} \mathbf{b}^*_\alpha \langle \mathbf{M}_{t'' t'} \mathbf{a}, \mathbf{\Lambda}_{\alpha t''} \rangle \end{aligned} \tag{3.21}$$

where $\mathbf{b}_\alpha*$ is the complex conjugate of $\mathbf{b}_\alpha$.

The perturbation $\mathbf{a}(t')$ evolves to $\mathbf{M}_{t''t'}\mathbf{a}$ at $t''$. As for $\mathbf{\Lambda}_{\alpha t''}$, they are Lyapunov vectors in $t''$, which will grow (or decrease) depending on whether they are unstable (or stable) until $t = \tau$.

That way, $\langle \mathbf{M}_{t''t'}\mathbf{a}, \mathbf{\Lambda}_{\alpha t''} \rangle$ provides the component of $\mathbf{M}_{t''t'}\mathbf{a}$ along the Lyapunov vector $\mathbf{\Lambda}_{\alpha t''}$. Moreover,

$$\langle \mathbf{M}_{t''t'}\mathbf{a}, \mathbf{\Lambda}_{\alpha t''} \rangle = \langle \mathbf{a}, \mathbf{M}^*_{t''t'}\mathbf{\Lambda}_{\alpha t''} \rangle = ||\mathbf{M}_{t''t'}\mathbf{a}|| \cos \sphericalangle (\mathbf{M}_{t''t'}\mathbf{a}, \mathbf{\Lambda}_{\alpha,t''}) \,. \tag{3.22}$$

The vector $\mathbf{a}$ is taken at $t'$. If it is projected along the Lyapunov vector $\mathbf{\Lambda}_{\alpha t'}$, then one obtains the inner product $\langle \mathbf{M}_{t't''}\mathbf{\Lambda}_{\alpha t'}, \mathbf{\Lambda}_{\alpha t''} \rangle$. The Lyapunov vector $\mathbf{\Lambda}_{\alpha t''}$, taken at $t''$, is not necessarily colinear with $\mathbf{M}_{t't''}\mathbf{\Lambda}_{\alpha t'}$ resulting from the application of the tangent linear evolution operator to the Lyapunov vector $\mathbf{\Lambda}_{\alpha t'}$ taken at $t'$. That colinearity only holds when the inverse operator of $\mathbf{M}_{t''t'}$ satisfies $(\mathbf{M}_{t''t'})^{-1} = \mathbf{M}^*_{t''t'}$, i.e. when it is a normal operator (where the adjoint equals the inverse). In general it does not hold in nonlinear systems because of the non-normal component of $\mathbf{M}$ due to the stretching, dilatation and compression of small parcels in the phase space.

### 3.3.3 Short-term error dynamics

The dynamics of prediction errors in the short-term scenario is hereby assessed, resorting to:

i) a Taylor expansion of the Oseledec operator in time $\tau$ and subsequently writing the square error terms as function of the $\tau$-expanded operator (subsection 3.3.3.1);

ii) a Taylor expansion of the square error in time $\tau$, directly from the governing equations for the error, Eq. (3.6) (subsection 3.3.3.2).

#### 3.3.3.1 Via the Oseledec operator

Our first step is to determine the first time derivatives of the Oseledec operator defined in Eq. (3.15). For that purpose, we begin by considering the time-derivative of $\mathbf{M}$ in Eq. (3.8b) or, equivalently for its adjoint,

$$\frac{d\mathbf{M}^*}{dt} = \mathbf{M}^*\mathbf{J}^* \,. \tag{3.23}$$

We introduce now, for a generic operator $\mathbf{J}$, the operator function $\mathbf{K_J} = \mathbf{M^{-1}J\,M}$ with adjoint $\mathbf{K_J^*} = \mathbf{M}^*\mathbf{J}^*(\mathbf{M^{-1}})^*$.

The first time-derivative of the Oseledec operator can thus be written as:

$$\frac{d\mathbf{O}_{\tau,t',t''}}{d\tau} = \mathbf{O}\,\mathbf{K}_{\mathbf{J}t''} + \mathbf{K}^*_{\mathbf{J}t'}\mathbf{O}\,. \tag{3.24}$$

For $\tau = 0$ and since $0 \leqslant t', t'' \leqslant \tau$, the Jacobian and its derivatives are computed at $t' = t'' = 0$ in the derivatives of $\mathbf{O}$ at $\tau = 0$, which are given by:

$$\left(\frac{d\mathbf{O}}{d\tau}\right)_{\tau=0} = \mathbf{A}\,\mathbf{J}_{t''} + \mathbf{J}^*_t\mathbf{A} \equiv \mathbf{O}_1\,. \tag{3.25}$$

The notation used for the time derivatives of $\mathbf{O}_{\tau,t',t''}$ is presented in the form $(\mathbf{O}_k)_{t't''} = \left(\frac{d^k\mathbf{O}_{\tau,t',t''}}{d\tau^k}\right)_{\tau=0}$ or simply $\mathbf{O}_k = \left(\frac{d^k\mathbf{O}}{d\tau^k}\right)_{\tau=0}$.

We proceed by deriving the second time derivative of $\mathbf{O}$.

Let $\mathbf{J}_l = \frac{d^l\mathbf{J}}{dt^l}$ be the $l$-th order time derivative of $\mathbf{J}$. Since $\mathbf{J}$ depends on the current state $\mathbf{x}(t)$, the Jacobian depends implicitly on time $t$. Then, we obtain $\mathbf{J}_1 = \frac{d\mathbf{J}}{dt} = \frac{\partial\mathbf{J}}{\partial\mathbf{x}}\mathbf{f}(x)$. In turn, $\frac{d\mathbf{K_J}}{d\tau} = \mathbf{K}_{\mathbf{J}1} = \mathbf{M^{-1}J_1M}$, from the fact that $\frac{d\mathbf{M}^{-1}}{d\tau} = -\mathbf{M}^{-1}\mathbf{J}$. At $\tau = 0$, $(\mathbf{K}_{\mathbf{J}1})_{\tau=0} = \mathbf{J}_1$.

In general:

$$\frac{d\mathbf{K_A}}{d\tau} = \mathbf{K}_{[\mathbf{A},\mathbf{J}]} + \mathbf{K}_{\frac{d\mathbf{A}}{d\tau}}\,, \tag{3.26}$$

where $[\mathbf{A},\mathbf{J}] = \mathbf{A}\,\mathbf{J} - \mathbf{J}\,\mathbf{A}$ is the commutation operator.

The second time derivative is then given by:

$$\frac{d^2\mathbf{O}}{d\tau^2} = \frac{d\mathbf{O}}{d\tau}\mathbf{K}_{\mathbf{J}t''} + \mathbf{K}^*_{\mathbf{J}t''}\frac{d\mathbf{O}}{d\tau} + \mathbf{O}\left(\mathbf{K}_{\mathbf{J}_1}\right)_{t''} + \left(\mathbf{K}^*_{\mathbf{J}_1}\right)_{t'}\mathbf{O}\,. \tag{3.27}$$

At $\tau = 0$:

$$\left(\frac{d^2\mathbf{O}}{d\tau^2}\right)_{\tau=0} = \left(\mathbf{A}\,\mathbf{J}_{t''} + \mathbf{J}^*_t\,\mathbf{A}\right)\mathbf{J}_{t''} + \mathbf{J}^*_t\left(\mathbf{A}\,\mathbf{J}_{t''} + \mathbf{J}^*_t\,\mathbf{A}\right) + \mathbf{A}\,\mathbf{J}_{1t''} + \mathbf{J}^*_{1t''}\mathbf{A} \equiv \mathbf{O}_2\,. \tag{3.28}$$

The third derivative is given by:

$$\frac{d^3\mathbf{O}}{d\tau^3} = \frac{d^2\mathbf{O}}{d\tau}\mathbf{K}_{\mathbf{J}t''} + 2\frac{d\mathbf{O}}{d\tau}\left(\mathbf{K}_{\mathbf{J}_1}\right)_{t''} + \mathbf{K}^*_{\mathbf{J}t'}\frac{d^2\mathbf{O}}{d\tau^2} + 2\left(\mathbf{K}^*_{\mathbf{J}_1}\right)_{t'}\frac{d\mathbf{O}}{d\tau} + \mathbf{O}\left(\mathbf{K}_{\mathbf{J}_2}\right)_{t''} + \left(\mathbf{K}^*_{\mathbf{J}_2}\right)_{t'}\mathbf{O}\,, \tag{3.29}$$

by taking into account that:

$$\begin{aligned}\mathbf{K}_{\mathbf{J}_2} = \frac{d\mathbf{K}_{\mathbf{J}1}}{d\tau} &= \mathbf{K}_{[\mathbf{J}_1,\mathbf{J}]} + \mathbf{K}_{\mathbf{J}_2} \\ &= \mathbf{K}_{\mathbf{J}\,\mathbf{J}_1} - \mathbf{K}_{\mathbf{J}_1\mathbf{J}} + \mathbf{K}_{\mathbf{J}_2}\,.\end{aligned} \tag{3.30}$$

At $\tau = 0$:

$$\begin{aligned}\mathbf{K}_{\mathbf{J}_2} &= \mathbf{J}\,\mathbf{J}_1 - \mathbf{J}_1\,\mathbf{J} + \mathbf{J}_2 \\ &= [\mathbf{J}, \mathbf{J}_1] + \mathbf{J}_2\,.\end{aligned} \tag{3.31}$$

Therefore:

$$\begin{aligned}\left(\frac{d^3\mathbf{O}}{d\tau^3}\right)_{\tau=0} &= \left[\mathbf{A}\,\mathbf{J}^2_{t'} + 2\mathbf{J}^*_{t'}\mathbf{A}\,\mathbf{J}_{t''} + \left(\mathbf{J}^*_{t'}\right)^2\mathbf{A} + \mathbf{A}\,\mathbf{J}_{1t''} + \mathbf{J}^*_{t'}\,\mathbf{A}\right]\mathbf{J}_{t''} + 2\left(\mathbf{A}\,\mathbf{J}_{t''} + \mathbf{J}^*_{t'}\mathbf{A}\right)\mathbf{J}_{1t''} \\ &+ \mathbf{J}^*_{t'}\left[\mathbf{A}\mathbf{J}^2_{t''} + \left(\mathbf{J}^*_{t'}\right)^2\mathbf{A} + \mathbf{A}\mathbf{J}_{1t''} + \mathbf{J}^*_{1t'}\mathbf{A} + 2\mathbf{J}^*_{t'}\,\mathbf{A}\,\mathbf{J}_{t''}\right] + 2\mathbf{J}^*_{\mathbf{1}t'}\left(\mathbf{A}\,\mathbf{J}_{t''} + \mathbf{J}^*_{t'}\mathbf{A}\right) \\ &+ \mathbf{A}\left(\mathbf{J}_{t''}\mathbf{J}_{1t''} - \mathbf{J}_{1t''}\mathbf{J}_{t''} + \mathbf{J}_{2t''}\mathbf{A}\right) \equiv \mathbf{O}_3\,.\end{aligned} \tag{3.32}$$

The Oseledec operator will thus admit the $\tau$-expansion:

$$\mathbf{O} = \mathbf{A} + \left(\mathbf{O}_1\right)_{t't''}\tau + \left(\mathbf{O}_2\right)_{t't''}\frac{\tau^2}{2} + \left(\mathbf{O}_3\right)_{t't''}\frac{\tau^3}{6} + \mathcal{O}(\tau^4)\,, \tag{3.33}$$

where $\mathbf{O}_1, \mathbf{O}_2, \mathbf{O}_3$ depend on $t', t''$.

Therefore, the square error term relative to errors in the initial conditions becomes:

$$\begin{aligned}\langle \mathbf{u}_I, \mathbf{A}\mathbf{u}_I\rangle = \langle \mathbf{u}_0, \mathbf{A}\mathbf{u}_0\rangle + \langle \mathbf{u}_0, (\mathbf{O}_1)_{00}\mathbf{u}_0\rangle\tau + \langle \mathbf{u}_0, (\mathbf{O}_2)_{00}\mathbf{u}_0\rangle\frac{\tau^2}{2} \\ + \langle \mathbf{u}_0, (\mathbf{O}_3)_{00}\mathbf{u}_0\rangle\frac{\tau^3}{6} + \mathcal{O}(\tau^4)\,,\end{aligned} \tag{3.34a}$$

where $(\mathbf{O}_1)_{00}, (\mathbf{O}_2)_{00}, (\mathbf{O}_3)_{00}$ are taken at $t' = t'' = \tau = 0$ and therefore they depend exclusively from the initial condition $\mathbf{x}(t=0)$.

Equivalently, (3.34a) can be written as:

$$\begin{aligned}&\langle \mathbf{u}_I, \mathbf{A}\mathbf{u}_I\rangle = \langle \mathbf{u}_0, \mathbf{Z}(\tau)\mathbf{u}_0\rangle \\ &\text{with} \\ &\mathbf{Z}(\tau) = \mathbf{A} + (\mathbf{O}_1)_{00}\tau + (\mathbf{O}_2)_{00}\frac{\tau^2}{2} + (\mathbf{O}_3)_{00}\frac{\tau^3}{6} + \mathcal{O}(\tau^4)\,.\end{aligned} \tag{3.34b}$$

The term referring to the interaction between initial condition and model errors becomes:

$$\begin{aligned}\langle \mathbf{u}_I, \mathbf{d}\rangle &= \left\langle \mathbf{u}_0, \int_0^\tau \mathbf{A}\,\mathbf{p}(t)\,dt \right\rangle + \left\langle \mathbf{u}_0, \int_0^\tau (\mathbf{O}_1)_{0t}\,\mathbf{p}(t)\,dt \right\rangle \tau \\ &+ \left\langle \mathbf{u}_0, \int_0^\tau (\mathbf{O}_2)_{0t}\,\mathbf{p}(t)\,dt \right\rangle \frac{\tau^2}{2} + \left\langle \mathbf{u}_0, \int_0^\tau (\mathbf{O}_3)_{0t}\,\mathbf{p}(t)\,dt \right\rangle \frac{\tau^3}{6} + \mathcal{O}(\tau^4) \\ &= \int_0^\tau \langle \mathbf{u}_0, \mathbf{A}\,\mathbf{p}(t)\rangle\,dt + \int_0^\tau \langle \mathbf{u}_0, (\mathbf{O}_1)_{0t}\,\mathbf{p}(t)\rangle\,dt\,\tau \\ &+ \int_0^\tau \langle \mathbf{u}_0, (\mathbf{O}_2)_{0t}\,\mathbf{p}(t)\rangle dt\,\frac{\tau^2}{2} + \int_0^\tau \langle \mathbf{u}_0, (\mathbf{O}_3)_{0t}\,\mathbf{p}(t)\rangle\,dt\,\frac{\tau^3}{6} + \mathcal{O}(\tau^4)\,.\end{aligned} \tag{3.35}$$

The term representing the contribution of model error (drift term) becomes:

$$\begin{aligned}\langle \mathbf{d}, \mathbf{d}\rangle &= \int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), \mathbf{A}\mathbf{p}(t'')\rangle\,dt'\,dt'' \\ &+ \int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), (\mathbf{O}_1)_{t't''}\mathbf{p}(t'')\rangle\,dt'\,dt''\,\tau \\ &+ \int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), (\mathbf{O}_2)_{t't''}\mathbf{p}(t'')\rangle\,dt'\,dt''\,\frac{\tau^2}{2} \\ &+ \int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), (\mathbf{O}_3)_{t't''}\mathbf{p}(t'')\rangle\,dt'\,dt''\,\frac{\tau^3}{3} + \mathcal{O}(\tau^4)\,.\end{aligned} \tag{3.36}$$

Unlike the term relative to errors in the initial conditions, (3.34a), the terms (3.35) and (3.36) depend, for each time $\tau$, on the accumulated effect of model errors until time $\tau$.

The short-term approximation for the square error is then obtained by applying Eqs. (3.34a–3.36) onto Eq. (3.13).

The terms in $\langle \mathbf{u}_I, \mathbf{A}\mathbf{u}_I \rangle$ are of the type $\langle \mathbf{u}_0, \mathbf{Z}\mathbf{u}_0 \rangle$, where $\mathbf{Z}$ is an operator. The terms in $\langle \mathbf{u}_I, \mathbf{d} \rangle$ are of the type $\int_0^\tau \langle \mathbf{u}_0, \mathbf{Z}_t \mathbf{p}(t) \rangle \, dt$. Finally, the terms in $\langle \mathbf{d}, \mathbf{d} \rangle$ are of the type $\int_0^\tau \int_0^\tau \langle \mathbf{p}(t'), \mathbf{Z}_{t't''} \, \mathbf{p}(t'') \rangle \, dt' \, dt''$, where $\mathbf{Z} = \mathbf{Z}(\mathbf{x}_0)$, $\mathbf{Z}_t = \mathbf{Z}(\mathbf{x}(t))$ for $t \geq 0$, $\mathbf{Z}_{t't''} = \mathbf{Z}(\mathbf{x}(t'), \mathbf{x}(t''))$ for $t', t'' \geq 0$.

The errors $\mathbf{u}_0$ and $\mathbf{p}(t)$ have a constant term, another depending on $\mathbf{x}$ and a random one. Over the above referred terms one shall perform an averaging on $\mathbf{u}_0, \mathbf{p}$ and over the attractor. By assuming the error $\mathbf{p}(t)$ to be uncorrelated in time, it does not interfere on the a posteriori model error. This assumption can be relaxed by admitting that there is a non-vanishing correlation time between random model errors as assumed by Vannitsem and Toth (2002).

In the reanalysis mode, in which the analysis error $\mathbf{u}_0$ depends also on numerical model simulations prior to the initial time $\tau = 0$, $\mathbf{p}(t)$ is no longer independent from $\mathbf{u}_0$. The model errors $\mathbf{p}(t')$ and $\mathbf{p}(t'')$ can have memory and thus be mutually dependent.

The various terms are of the type $\langle \mathbf{a}, \mathbf{Z}\,\mathbf{b} \rangle$, where $\mathbf{a}$ and $\mathbf{b}$ are $\mathbf{u}_0$ or $\mathbf{p}$, whereas $\mathbf{Z}$ is linear and depends on $\mathbf{x}$.

One can decompose $\mathbf{a}$ and $\mathbf{b}$ on the orthonormal basis $\phi_\alpha$ relative to the inner product $\langle \cdot, \cdot \rangle$ defined in Eq. (3.9). This way:

$$\langle \mathbf{a}, \mathbf{Z}\,\mathbf{b} \rangle = \sum_{\alpha,\alpha'} \langle \mathbf{a}_\alpha \, \phi_\alpha, \mathbf{Z}\,\mathbf{b}_{\alpha'} \phi_{\alpha'} \rangle = \sum_{\alpha,\alpha'} \mathbf{a}_\alpha \mathbf{b}^*_{\alpha'} \langle \phi_\alpha, \mathbf{Z}\,\phi_{\alpha'} \rangle \,, \tag{3.37}$$

where $\mathbf{a}_\alpha = \langle \mathbf{a}, \phi_\alpha \rangle$, $\mathbf{b}_{\alpha'} = \langle \mathbf{b}, \phi_{\alpha'} \rangle$ are the components of $\mathbf{a}, \mathbf{b}$ on the referred basis.

By averaging $\mathbf{a}$ and $\mathbf{b}$ over an ensemble of initial conditions, one has

$$\overline{\langle \mathbf{a}, \mathbf{Z}\,\mathbf{b} \rangle} = \sum_{\alpha,\alpha'} \overline{\mathbf{a}_\alpha \mathbf{b}^*_\alpha} \langle \phi_\alpha, \mathbf{Z}\,\Phi_{\alpha'} \rangle = \sum_{\alpha,\alpha'} \mathbf{C}_{\alpha\alpha'}\, \mathbf{Z}_{\alpha\alpha'}\,, \tag{3.38}$$

where $\mathbf{Z}_{\alpha\alpha'} = \langle \phi_\alpha, \mathbf{Z}\,\phi_{\alpha'} \rangle$ and $\mathbf{C}_{\alpha\alpha'} = \overline{\langle \mathbf{a}, \phi_\alpha \rangle \langle \phi_{\alpha'}, \mathbf{b} \rangle}$ is the mean of the product of $\mathbf{a}_\alpha$ by $\mathbf{b}^*_{\alpha'}$.

$\mathbf{C}_{\alpha\alpha'}$ is obtained from the covariance operator $\mathbf{C}$ between $\mathbf{a}$ and $\mathbf{b}$:

$$\begin{aligned} \mathbf{C}_{\alpha\alpha'} = \overline{\mathbf{a}_\alpha \mathbf{b}^*_{\alpha'}} = \overline{\langle \mathbf{a}, \phi_\alpha \rangle \langle \phi_\alpha, \mathbf{b} \rangle} = \\ = \langle \mathbf{C_{ab}} \phi_{\alpha'}, \phi_\alpha \rangle = \langle \phi_{\alpha'}, \mathbf{C_{ba}} \phi_\alpha \rangle\,. \end{aligned} \tag{3.39}$$

The adjoint of $\mathbf{C_{ab}}$ is $\mathbf{C^*_{ab}} = \mathbf{C_{ba}}$.

When $\mathbf{a} = \mathbf{b}$, the operator is self-adjoint, i.e. $\mathbf{C^*_{aa}} = \mathbf{C_{aa}}$.

The matrices $\mathbf{C}_{\alpha\alpha'}$ refer to means over products of errors $\mathbf{u}_0$ on the initial conditions and errors $\mathbf{p}$ on the tendencies of the linear tangent model.

In case there is only a random, unbiased component in the errors, the terms $\mathbf{C}_{\alpha\alpha'}$ are covariances. These covariances can be constant or be dependent on the position in the phase-space. The covariances can admit simplifications, such as when being homogeneous and isotropic on the sphere, in which case the matrix $\mathbf{C}$ is diagonal with elements solely dependent on the total wavenumber $n$ (Boer and Shepherd, 1983).

The term $\overline{\langle \mathbf{a}, \mathbf{Z}\,\mathbf{b} \rangle}$ admits also the expansion:

$$\overline{\langle \mathbf{a}, \mathbf{Z}\,\mathbf{b} \rangle} = \sum_{\alpha,\alpha'} \langle \mathbf{C_{ab}}\, \phi_{\alpha'}, \phi_\alpha \rangle \langle \phi_\alpha, \mathbf{Z}\,\phi_{\alpha'} \rangle \tag{3.40}$$

This expansion exhibits the separation between the covariance structure of the error in the initial conditions on the covariance operator $\mathbf{C}$ and the dynamics (present in the operator $\mathbf{Z}$).

When $\mathbf{Z} = \mathbf{I}$:

$$\overline{\langle \mathbf{a}, \mathbf{b} \rangle} = \sum_{\alpha,\alpha'} \langle \mathbf{C_{ab}}\, \phi_{\alpha'}, \phi_\alpha \rangle \delta_{\alpha\alpha'} = \sum_{\alpha} \langle \mathbf{C_{ab}} \phi_\alpha, \phi_\alpha \rangle = \sum_{\alpha} \overline{\mathbf{a}_\alpha \mathbf{b}^*_\alpha}\,. \tag{3.41}$$

If $\mathbf{a} = \mathbf{b}$ and it has a Gaussian distribution, then $\mathbf{C}$ is nuclear, i.e. $tr\,\mathbf{C} = \sum_\alpha \langle \mathbf{C_{ab}} \phi_\alpha, \phi_\alpha \rangle$ is well defined and independent from the chosen basis.

#### 3.3.3.2 Direct expansion

The Taylor expansion can also be obtained directly from the differential equation governing $||\mathbf{u}^2||$.

We begin by defining the initial square error:

$$U_0 \equiv \langle \mathbf{u}_0, \mathbf{A}\, \mathbf{u}_0 \rangle\,. \tag{3.42}$$

Now, let $\frac{d\mathbf{u}}{d\tau} = \mathbf{J}\mathbf{u} + \mathbf{p} \equiv \mathbf{R}$, $\mathbf{R}_{\tau=0} = \mathbf{J}\mathbf{u}_0 + \mathbf{p}_0 \equiv \mathbf{R}_0$. Then:

$$\begin{aligned} \frac{\partial ||\mathbf{u}^2||}{\partial \tau} = \frac{\partial}{\partial \tau} \langle \mathbf{u}, \mathbf{A}\mathbf{u} \rangle = \langle \mathbf{R}, \mathbf{A}\, \mathbf{u} \rangle + \langle \mathbf{u}, \mathbf{A}\, \mathbf{R} \rangle = \langle \mathbf{R}, \mathbf{A}\, \mathbf{u} \rangle + \langle \mathbf{A}\, \mathbf{u}, \mathbf{R} \rangle = \\ = \Re \left( 2 \langle \mathbf{A}\, \mathbf{u}, \mathbf{R} \rangle \right) = 2 \langle \mathbf{u}, \mathbf{A}\, \mathbf{R}_0 \rangle\,, \end{aligned} \tag{3.43}$$

since $\mathbf{u}, \mathbf{R} \in \mathbb{R}$.

At $\tau = 0$ (3.43) becomes:

$$\left( \frac{\partial ||\mathbf{u}^2||}{\partial \tau} \right)_{\tau=0} = 2 \langle \mathbf{u}_0, \mathbf{A}\, \mathbf{J}\, \mathbf{u}_0 \rangle + 2 \langle \mathbf{u}_0, \mathbf{A}\, \mathbf{p}_0 \rangle = 2 \langle \mathbf{u}_0, \mathbf{A}\, \mathbf{R}_0 \rangle \equiv U_1\,, \tag{3.44}$$

where all quantities are evaluated at $\tau = 0$.

The second-order time derivative is given by:

$$\frac{\partial^2 ||\mathbf{u}^2||}{\partial \tau^2} = 2 \langle \mathbf{R}, \mathbf{A}\, \mathbf{R} \rangle + 2 \langle \mathbf{u}, \mathbf{A}\, \mathbf{R}_1 \rangle\,, \tag{3.45}$$

where $\mathbf{R}_1 = \frac{\partial \mathbf{R}}{\partial \tau} = \mathbf{J}_1 \mathbf{u} + \mathbf{J}\, \mathbf{R} + \mathbf{p}_1$, $\mathbf{p}_1 = \frac{\partial \mathbf{p}}{\partial \tau}$.

Therefore, at $\tau = 0$ (3.45) becomes:

$$\begin{aligned} \left( \frac{\partial^2 ||\mathbf{u}^2||}{\partial \tau^2} \right)_{\tau=0} = 2 \langle \mathbf{J}\, \mathbf{u}_0 + \mathbf{p}_0, \mathbf{A}\, (\mathbf{J}\mathbf{u}_0 + \mathbf{p}_0) \rangle + 2 \langle \mathbf{u}_0, \mathbf{A}\, \mathbf{p}_1 \rangle + 2 \langle \mathbf{p}_0, \mathbf{A}\mathbf{p}_0 \rangle \\ + 2 \langle \mathbf{u}_0, \mathbf{A}\, \left( \mathbf{J}_1 + \mathbf{J}_1^2 \right) \mathbf{u}_0 + \mathbf{A}\, \mathbf{J}\, \mathbf{p}_0 + \mathbf{A}\mathbf{p}_1 \rangle \equiv U_2\,, \end{aligned} \tag{3.46}$$

where all quantities are evaluated at $\tau = 0$.

The first term is part of $\langle \mathbf{u}_I, \mathbf{A}\,\mathbf{u}_I \rangle$, the second and fourth are in $2\langle \mathbf{u}_I, \mathbf{d} \rangle$ and the third in $\langle \mathbf{d}, \mathbf{d} \rangle$.

The third-order time derivative is given by:

$$\begin{aligned} \frac{\partial^3 ||\mathbf{u}^2||}{d\tau^2} &= 2\langle \mathbf{R}, \mathbf{A}\,\mathbf{R}_1 \rangle + 2\langle \mathbf{R}_1, \mathbf{A}\,\mathbf{R} \rangle + 2\langle \mathbf{u}, \mathbf{A}\,\mathbf{R}_2 \rangle = \\ &= 6\langle \mathbf{R}, \mathbf{A}\,\mathbf{R}_1 \rangle + 2\langle \mathbf{u}, \mathbf{A}\,\mathbf{R}_2 \rangle\,, \end{aligned} \tag{3.47}$$

where $\mathbf{R}_2 = \mathbf{J}_2\mathbf{u} + 2\,\mathbf{J}_1\mathbf{R} + \mathbf{J}\,\mathbf{R}_\mathbf{1} + \mathbf{p}_2$, $\mathbf{p}_2 = \frac{\partial^2 \mathbf{p}}{\partial \tau^2}$.

At $\tau = 0$:

$$\begin{aligned} (\mathbf{R}_2)_{\tau=0} &= \mathbf{J}_2\mathbf{u}_0 + 2\mathbf{J}_1\left(\mathbf{J}\,\mathbf{u}_0 + \mathbf{p}_0\right) + \mathbf{J}\left[\mathbf{J}_1\mathbf{u}_0 + \mathbf{J}\left(\mathbf{J}\mathbf{u}_0 + \mathbf{p}_0\right) + \mathbf{p}_1\right] + \mathbf{p}_2 \\ &= \left(\mathbf{J}_2 + 2\mathbf{J}_1\mathbf{J} + \mathbf{J}\mathbf{J}_1 + \mathbf{J}_3\right)\mathbf{u}_0 + \left(2\mathbf{J}_1 + \mathbf{J}^2\right)\mathbf{p}_0 + \mathbf{J}\mathbf{p}_1 + \mathbf{p}_2\,, \end{aligned} \tag{3.48}$$

where all quantities are evaluated at $\tau = 0$.

Therefore the third-order time derivative at $\tau = 0$ becomes:

$$\begin{aligned} \left(\frac{\partial^3 ||\mathbf{u}^2||}{\partial \tau^3}\right)_{\tau=0} &= 6\left\langle \left(\mathbf{J}\mathbf{u}_0 + \mathbf{p}_0\right), \mathbf{A}\left[\left(\mathbf{J}_1 + \mathbf{J}^2\right)\mathbf{u}_0 + \mathbf{J}\mathbf{p}_0 + \mathbf{p}_1\right]\right\rangle + 2\langle \mathbf{u}_0, \mathbf{A}\,\mathbf{R}_2 \rangle = \\ &= \left\langle \mathbf{u}_0, \left\{6\left[\mathbf{J}^*\mathbf{A}\left(\mathbf{J}_1 + \mathbf{J}^2\right)\right] + 2\mathbf{A}\left(\mathbf{J}_2 + 2\mathbf{J}_1\mathbf{J} + \mathbf{J}\mathbf{J}_1 + \mathbf{J}^3\right)\right\}\mathbf{u}_0\right\rangle + \\ &+ \left\langle \mathbf{u}_0, \left\{6\,\mathbf{J}^*\mathbf{A}\,\mathbf{J} + 6\left[\mathbf{J}_1^* + (\mathbf{J}^*)^2\right]\mathbf{A} + 2\,\mathbf{A}\left(2\,\mathbf{J}_1 + \mathbf{J}^2\right)\right\}\mathbf{p}_0\right\rangle + \\ &+ \langle \mathbf{u}_0, \left(6\,\mathbf{J}^*\mathbf{A} + 2\,\mathbf{A}\,\mathbf{J}\right)\mathbf{p}_1\rangle + \langle \mathbf{u}_0, 2\mathbf{A}\mathbf{p}_2\rangle + \langle \mathbf{p}_0, 6\mathbf{A}\,\mathbf{J}\,\mathbf{p}_0\rangle + \langle \mathbf{p}_0, 6\mathbf{A}\,\mathbf{p}_1\rangle \\ &\equiv U_3 \end{aligned} \tag{3.49}$$

The Taylor expansion around $\tau = 0$ (Mac-Laurin expansion) of the square error can then be written as:

$$||\mathbf{u}^2(\tau)|| = U_0 + U_1\tau + \frac{1}{2}U_2\tau^2 + \frac{1}{6}U_3\tau^3 + \mathcal{O}(\tau^4)\,. \tag{3.50}$$

Here, $U_0$ naturally features only errors in the initial conditions (3.14), $U_1$ features contributions of initial condition errors and of their interaction with model-related errors (3.16), and $U_i$ for $i \geqslant 2$ has contributions from the aforementioned errors, along with that from the drift term (3.17).

In the absence of errors in the initial conditions, the dominant term in (3.50) is the quadratic one, which is consistent with Vannitsem and Toth (2002) and Nicolis, Perdigão, and Vannitsem (2009).

We will be interested in the behaviour of the square error averaged over the attractor and over an ensemble of initial conditions. By taking such averages in Eq. (3.50), we are led to:

$$\overline{[\![||\mathbf{u}^2(\tau)||]\!]} = \overline{[\![U_0]\!]} + \overline{[\![U_1]\!]}\tau + \frac{1}{2}\overline{[\![U_2]\!]}\tau^2 + \frac{1}{6}\overline{[\![U_3]\!]}\tau^3 + \mathcal{O}(\tau^4)\,, \tag{3.51}$$

where $[\![\cdots]\!]$ denotes averaging over the attractor and $\overline{\cdots}$ averaging over an ensemble of initial conditions.

An additional averaging should eventually be taken over the model-related error terms in case these have a random component apart from the constant, systematic one.

Simplifications to $[\![U_i]\!]$ occur in particular cases to be considered in subsequent sections.

### 3.3.4 Beyond Taylor expansions: a more robust analytical approximation

Given the terms of the Mac-Laurin expansion of the mean square error up to the third order in $\tau$, we can take the Padé approximant (Baker and Graves-Morris, 1996) of order $[2:1]$, in the form:

$$\mathcal{P}\left(\overline{[\![\,||\mathbf{u}^2(\tau)||\,]\!]}\right) = \frac{c_1 + c_2\tau + c_3\tau^2}{c_4 + c_5\tau}\,, \tag{3.52}$$

where the coefficients $c_k$ are determined in such a way that the $\tau$-expansion of $\mathcal{P}\left(\overline{[\![\,||\mathbf{u}^2(\tau)||\,]\!]}\right)$ matches that of $\overline{[\![\,||\mathbf{u}^2(\tau)||\,]\!]}$ up to the third order in $\tau$.

This way,

$$\begin{aligned} c_1 &= T_2 T_0 \,, \\ c_2 &= T_2 T_1 - T_3 T_0 \,, \\ c_3 &= T_2^2 - T_3 T_1 \,, \\ c_4 &= T_2 \,, \\ c_5 &= -T_3 \,, \end{aligned} \tag{3.53}$$

where $T_i$ is the coefficient of the $\tau^i$ term in the Taylor expansion. For instance, by considering Eq. (3.51), $T_i = \overline{[\![U_i]\!]}/i!$.

Analogously, the Padé approximant for the square error, $\mathcal{P}\left(||\mathbf{u}^2(\tau)||\right)$, can be derived from Eq. (3.50)). In this case, $T_i = U_i/i!$.

Note that there is a fundamental caveat underneath the Padé approximant: it becomes infinite when the denominator in (3.52) becomes null, i.e. when $\tau = -c_4/c_5 = T_2/T_3$. Nevertheless, in the various practical applications considered in the thesis and related research papers the singularity does not occur in the short to intermediate time scale, where the approximation holds.

### 3.3.5 Short to intermediate time expansion in finite dimensional systems

Most of the developments and results presented for the remainder of this chapter have been included in an article prepared under close supervision of Prof. Doctor Catherine Nicolis and Doctor Stéphane Vannitsem at the Royal Meteorological Institute of Belgium, and published at the Journal of Atmospheric Sciences from the American Meteorological Society (Nicolis, Perdigão, and Vannitsem, 2009). The structure of the upcoming subsections within this chapter will thus follow the aforementioned article, though now complemented with some formal generalisations.

In the present subsection, a systematic expansion of the solutions of Eq. (3.6) in the short to intermediate time regime is carried out in the practical case of a generic $n$-dimensional system.

Some general, model-independent features are brought out such as the role of the mechanisms of error transfer between a particular initial direction to components along other directions, the existence of a minimum in the error evolution and the relative importance of the IC errors and model errors in the global evolution.

Consider generic $n$-dimensional systems, the variables of which now span a $n$-dimensional phase space and a $n'$-dimensional parameter vector $\mu$ along with the corresponding rectangular $\mathbf{\Phi}(n \times n')$ matrix, with $n, n' \in \mathbb{N}$.

We hereby derive the behaviour of the error vector $\mathbf{u}$ and its norm $N(\mathbf{u})$, in the regime of short to intermediate times.

The term $\delta\mathbf{f}$ is considered small enough to be neglected ($|\delta\mathbf{f}| \ll |\mathbf{\Phi}\delta\mu|$), which leads to $\mathbf{p}$ in Eq. (3.7) becoming $\mathbf{p} = \mathbf{\Phi}\delta\mu$.

We begin by expanding $\mathbf{u}$ around $t = 0$, keeping terms up to $\mathcal{O}(t^3)$:

$$\mathbf{u}(t) = \mathbf{u}(0) + \left(\frac{d\mathbf{u}}{dt}\right)_0 t + \frac{1}{2}\left(\frac{d^2\mathbf{u}}{dt^2}\right)_0 t^2 + \frac{1}{6}\left(\frac{d^3\mathbf{u}}{dt^3}\right)_0 t^3 + \cdots . \tag{3.54}$$

By using Eq. (3.6) we obtain straightforwardly (see Appendix C for details)

$$u_i(t) = u_i + A_i t + \frac{1}{2}B_i t^2 + \frac{1}{6}C_i t^3 + \cdots , \tag{3.55}$$

where the coefficients $A_i$, $B_i$, $C_i$ are given by (being understood that all quantities involved are to be evaluated at $t = 0$ on the reference attractor):

$$A_i = \sum_j \left(J_{ij}u_j + \Phi_{ij}\delta\mu_j\right) \tag{3.56a}$$

$$B_i = \sum_{jk} J_{ij} J_{jk} u_k + \sum_j \frac{dJ_{ij}}{dt} u_j + \sum_{jk} J_{ij} \Phi_{jk} \delta\mu_k + \sum_j \frac{d\Phi_{ij}}{dt} \delta\mu_j \tag{3.56b}$$

$$\begin{aligned} C_i = & \sum_{jkl} J_{ij} J_{jk} J_{kl} u_l + 2 \sum_{jk} \frac{dJ_{ij}}{dt} J_{jk} u_k + \sum_{jk} J_{ij} \frac{dJ_{jk}}{dt} u_k, \\ & + \sum_j \frac{d^2 J_{ij}}{dt^2} u_j + \sum_{jkl} J_{ij} J_{jk} \Phi_{kl} \delta\mu_l \\ & + \sum_{jk} \left( 2 \frac{dJ_{ij}}{dt} \Phi_{jk} + J_{ij} \frac{d\Phi_{jk}}{dt} \right) \delta\mu_k + \sum_k \frac{d^2 \Phi_{ik}}{dt^2} \delta\mu_k \,. \end{aligned} \tag{3.56c}$$

These terms generalise those presented in Nicolis, Perdigão, and Vannitsem (2009) to the more general case in which $\delta\mu$ is a vector. $A_i$, $B_i$ and $C_i$ correspond to the subsection 3.3.3.2 terms $\mathbf{R}_0$, $\mathbf{R}_{1\tau=0}$ and $\mathbf{R}_{2\tau=0}$, respectively.

We next turn to the computation of the norm $N(\mathbf{u})$, to the same order. By taking a general class of norms, we write the quadratic error $N(\mathbf{u}) = ||\mathbf{u}^2(t)||$ as:

$$\begin{aligned} ||\mathbf{u}^2(t)|| &= \sum_{ij} g_{ij} u_i(t) u_j(t) \\ &= \sum_{ij} g_{ij} \left( u_i + A_i t + \frac{1}{2} B_i t^2 + \frac{1}{6} C_i t^3 + \cdots \right) \\ &\quad \cdot \left( u_j + A_j t + \frac{1}{2} B_j t^2 + \frac{1}{6} C_j t^3 + \cdots \right) \end{aligned} \tag{3.57}$$

and keep all terms up to $t^3$. This yields:

$$\begin{aligned} ||\mathbf{u}^2(t)|| = \sum_{ij} g_{ij} \Bigg\{ & u_i u_j + (A_i u_j + u_i A_j) t \\ & + \left[ A_i A_j + \frac{1}{2} (B_i u_j + u_i B_j) \right] t^2 \\ & + \left[ \frac{1}{2} (A_i B_j + B_i A_j) + \frac{1}{6} (C_i u_j + u_i C_j) \right] t^3 \Bigg\} \,. \end{aligned} \tag{3.58}$$

The positive-defined matrix $g_{ij}$ plays the role of $\mathbf{A}$ in (3.11).

By taking, as the aforementioned $\mathbf{A}$, $g_{ij}$ to be self-adjoint, $g_{ij} = g_{ji}$ and, consequently, Eq. (3.58) becomes:

$$||\mathbf{u}^2(t)|| = \sum_{ij} g_{ij} \Bigg\{ u_i\, u_j + 2\, u_i\, A_j\, t + (A_i\, A_j + u_i\, B_j)\, t^2 + \left( A_i\, B_j + \frac{1}{3}\, u_i\, C_j \right) t^3 \Bigg\}. \tag{3.59}$$

This expression is thus equivalent to (3.50) under the conditions of the present section (proof in Appendix E).

For the remainder of this chapter, we shall use the Euclidean metric, where $g_{ij} = \delta_{ij}^{\mathrm{kr}}$, where $\delta_{ij}^{\mathrm{kr}}$ is the Kronecker Delta. Other metrics shall be considered in the next chapter.

Under the Euclidean metric, Eq. (3.57) then becomes

$$\begin{aligned} ||\mathbf{u}^2(t)|| &= \sum_i u_i^2(t) \\ &= \sum_i \left( u_i + A_i t + \frac{1}{2} B_i t^2 + \frac{1}{6} C_i t^3 + \cdots \right)^2 \end{aligned} \tag{3.60}$$

and (3.58) becomes

$$||\mathbf{u}^2(t)|| = \sum_i u_i^2 + 2 \sum_i A_i u_i t + \sum_i \left( A_i^2 + B_i u_i \right) t^2 + \sum_i \left( A_i B_i + \frac{1}{3} C_i u_i \right) t^3. \tag{3.61}$$

By inspecting the previous equations, the following observations can thus be made:

(i) At the level of $u_i(t)$, in addition to contributions due to the evolution of the initial error $u_j$ by the Jacobian matrix $\mathbf{J}$ and to the model error *per se*, there are terms arising from their combined effect. These terms show up as products of elements of the Jacobian matrix or derivatives thereof and of components of the model error source term $\mathbf{\Phi}$ and its derivatives (all derivatives being evaluated at $t = 0$ on the reference attractor).

(ii) At the level of $||\mathbf{u}^2(t)||$, in addition to the aforementioned contributions there are "direct" coupling terms as well, in which the initial error components $u_i$ themselves multiply contributions containing the model error source term $\mathbf{\Phi}$.

(iii) According to Eqs (3.55)-(3.56b) there is a cascade mechanism by which an initial error acting solely along a particular component is transferred in the course of time in phase space to affect eventually components along other (initially error-free) directions as well.

Since local quantities are subjected to large fluctuations, to proceed further we place ourselves in the perspective of a *statistical ensemble of forecasts* and perform an average of the square error (3.61) in order to get information independent of the initial condition chosen. This averaging involves two kinds of processes. At first, over the reference (nature's) attractor, whose structure enters in Eqs. (3.8a) through the state dependence of $\mathbf{J}$ and $\mathbf{\Phi}$. The notation for such average is taken as $[\![\cdots]\!]$. And a second one, over the possible orientations and magnitudes of the initial error vector $\mathbf{u}(0)$ *per se*, with notation $\overline{\cdots}$ as in the previous section. In doing so we assume initially unbiased and uncorrelated error components $\left(\overline{[\![u_i]\!]} = 0, \overline{[\![u_i u_j]\!]} = \overline{[\![u_i^2]\!]}\delta_{ij}^{\mathrm{kr}}\right)$, keeping otherwise the full form of the associated probability distribution general.

It is worth noting, however, that even for biased errors and a non-diagonal initial error covariance matrix $\mathbf{C}$, one may consider a linear non-homogeneous tranformation of variables by performing the diagonalization of $\mathbf{C}$. Equal variances in all components are still possible by applying appropriate scaling. Finally, after obtaining the model equations for the transformed variables, one may proceed with the formalism by assuming uncorrelated, unbiased and equal variance initial condition errors without any loss of generality.

The usefulness of the averaging over the possible orientations and magnitudes

of the initial error vector is to provide hints on the overall predictive skill of a forecasting model. A good illustration is provided by Lorenz's pioneering study (Lorenz, 1982; Simmons and Hollingsworth, 2002) that led to the estimate of the $\sim 2$ day predictability horizon of the ECMWF weather forecasts in the early 1980's.

Taking into account that in an ergodic system the average of the time derivative of a bounded function is zero we obtain after a straightforward calculation (see Appendix D for details):

$$\begin{aligned}\overline{[\![||\mathbf{u}^2(t)||]\!]} = &\sum_i \overline{[\![u_i^2]\!]} + 2\sum_i \overline{[\![J_{ii}]\!]}\,\overline{[\![u_i^2]\!]}t \\ &+ \left[\sum_{ij} \overline{[\![J_{ij}^2]\!]}\,\overline{[\![u_j^2]\!]} + \sum_{ij} \overline{[\![J_{ij}J_{ji}]\!]}\,\overline{[\![u_i^2]\!]} + \sum_{ij} \overline{[\![\Phi_{ij}^2]\!]}\,\overline{[\![\delta\mu_j^2]\!]}\right] t^2 \\ &+ \left[\sum_{ijk} \overline{[\![J_{ik}J_{ij}J_{jk}]\!]}\,\overline{[\![u_k^2]\!]} + \frac{1}{3}\left(\sum_{ijk} \overline{[\![J_{ij}J_{jk}J_{ki}]\!]} + \sum_{ij} \overline{\left[\!\left[\frac{dJ_{ij}}{dt}J_{ji}\right]\!\right]}\right)\overline{[\![u_i^2]\!]} \right. \\ &\left. + \sum_k \overline{\left[\!\left[\sum_i \Phi_{ik}\sum_j J_{ij}\Phi_{jk}\right]\!\right]}\,\overline{[\![\delta\mu_k^2]\!]}\right] t^3. \end{aligned} \tag{3.62}$$

The case of biased initial errors is briefly considered in the end of subsection 3.4.3. We recognise (last term of the $t^2$ part) the short time behaviour of model error found in previous work (Nicolis, 2003, 2004). We also see that in the process of averaging, all direct coupling terms between initial and model errors have been cancelled. There subsists, however, a single contribution (last term in the $t^3$ part) where the error source term is evolved by the Jacobian matrix. Notice that the Jacobian matrix enters in Eq. (3.62) both through its diagonal elements and its non-diagonal ones. As a rule these elements are not related to each other, as the matrix needs not be symmetric.

So far our formulation accounts for an arbitrary distribution of the magnitudes of the individual initial error components. It is now instructive to consider the limit where these errors are distributed isotropically in phase space, a property translated by $\overline{[\![u_i^2]\!]} = \epsilon^2$ independently of $i$. In this limit the only surviving term in Eq. (3.62) still involving the time derivative of a phase space function will vanish. Furthermore, the coefficient of the $t$ part – which constitutes the dominant contribution for short times – displays the average of the sum of the diagonal elements of $\mathbf{J}$, which is known to be equal to the rate of change of phase space volumes as the dynamics is proceeding. In dissipative systems – a class which encompasses most of the systems encountered in Meteorology related problems – phase space volumes contract on average, entailing that the sum in question is negative. This leads us to the rather general conclusion that the mean quadratic error is bound to decrease for short times. The magnitude of $\sum_i J_{ii}$ or the relative rate of variation of phase space volume will determine the extent of this decreasing stage and, at the same time, the range of validity of the $t$-expansion. In particular, in near-conservative systems where the sum is close to zero the expansion is expected to provide an adequate description for an appreciable period of time.

If the system's dynamics is unstable the above mentioned decreasing trend will eventually be reversed as the unstable modes will gradually take over, even in the absence of model error. There is thus bound to be, in such systems, a minimum of the mean square error as a function of time attained at some value $t_m$ for which the time derivative of the right hand side of Eq. (3.62) vanishes. As seen in subsections 3.4.1 to 3.4.3 the $t$-expansion of Eq. (3.62) provides in many cases reasonable estimates of this time which can be further improved by alternative (more global) approximation schemes like the Padé approximants, as

discussed further below.

Another interesting type of estimate afforded by our formulation is that of the relative importance of the contributions of initial ($\mathbf{\Phi}$-independent parts) and model ($\mathbf{\Phi}$-dependent parts) errors. Clearly, as the contributions of model errors start as terms of $\mathcal{O}(t^2)$ there is bound to be a time interval (which in certain cases may be quite short − see examples in the following subsections) during which initial errors dominate model ones. As the latter are gradually building up, the question arises as to (a), their effect on the minimum attained at $t_m$; (b), the existence of a crossover time $t_c$ for which the magnitudes of the contributions of initial and model errors match each other. Eq. (3.62) allows one to check these points and, in particular, to evaluate $t$ and compare it to the value of the time $t_m$ where the mean square error attains a minimum. Beyond the time $t_c$ error dynamics will be dominated by model errors. Notice that a crossover, if any, does not imply that the mean quadratic error vanishes: its two constituent parts (initial and model errors) reach equal magnitudes, but do not cancel each other (they would need for this to be of opposite signs). Strict cancellation, if any, can only occur at the level of Eq. (3.55) for the error itself prior to averaging, for certain particular initial conditions. We stress again that all estimates above can be carried out systematically and in a quantitative manner in the limit where both initial and parameter errors are small. Beyond this limit they become system dependent and need to be considered on a case by case basis.

The general formulation and in particular, the $t$-expansion outlined above carry through in essentially the same form for the class of norms generalising the Euclidean one by the presence of a more general metric as seen in (3.57) and (3.58). Nevertheless, the question of occurrence of minimum and crossover times is subtler, since the Jacobian matrix elements $J_{ij}$ are now weighted by the $g_{ij}$'s.

As an example, the $t$-term in Eq. (3.62) is replaced by $\sum_i \overline{[\![J_{ij}]\!]}\, g_{ii}\, \overline{[\![u_i^2]\!]}\, t$ or, in the limit where initial errors are isotropically distributed, by $\epsilon^2 \sum_i \overline{[\![J_{ii}]\!]}\, g_{ii}\, t$. In a sense, because of the subsistence of the weighting factors $g_{ii}$ multiplying $\overline{[\![J_{ii}]\!]}$ the error dynamics in the isotropic case under such a norm is mapped into the error dynamics in the anisotropic case under an Euclidean norm. Although no general statement can be made, one might expect (cf. also comment in the last paragraph of subsection 3.4.2 below) that some of the results on minimum and crossover times will subsist as long as the $g_{ii}$'s along the stable directions retain a sufficiently significant value. Finally, when norms that are not quadratic in $\mathbf{u}$ are adopted like, for instance, the magnitude $|\mathbf{u}|$ of the error vector, the linear in $\mathbf{u}$ terms do not cancel in Eq. (3.62) and the model error grows in a subquadratic fashion. Extracting generic features becomes now more laborious, owing to the non-analytic dependencies introduced by the absolute value function.

We stress that the assumptions of unbiased and uncorrelated errors used to derive Eq. (3.62), as well as the one on isotropically distributed errors used in much of the discussion following this equation, apply only for the initial errors. As the system evolves errors will not only grow but will become, as a rule, strongly correlated by the dynamics. They will also develop in an anisotropic way and, for sufficiently long times, they will tend to be oriented along the leading Lyapunov vectors. In numerical weather prediction models used for operational purposes correlated and anisotropic "initial" errors show up through the use of short forecasts in the process of data assimilation. As long as these errors remain small, they can be accounted for by the averaged versions of Eqs. (D.1), (D.3) and (D.4) where no assumptions of randomness and isotropy as used in deriving Eq. (3.62) are made. As a counterpart no conclusions of comparable generality as that of our earlier ones can now be drawn, as one needs to specify the kinds of

correlations and anisotropies that may be present. This can only be done on a case by case basis.

Finally, by using the procedure leading to Eq. (3.62) one may also compute the error along one particular phase space direction $\alpha$ ($\alpha = 1, \cdots, n$):

$$\begin{aligned}\overline{[\![u_\alpha^2(t)]\!]} =& \overline{[\![u_\alpha^2]\!]} + 2\overline{[\![J_{\alpha\alpha}]\!]}\,\overline{[\![u_\alpha^2]\!]}\,t \\ &+ \left[\sum_j \left(\overline{[\![J_{\alpha j}^2]\!]}\,\overline{[\![u_j^2]\!]} + \overline{[\![J_{\alpha j}J_{j\alpha}]\!]}\,\overline{[\![u_\alpha^2]\!]}\right) + \overline{[\![\Phi_{\alpha j}^2]\!]}\,\overline{[\![\delta\mu_j^2]\!]}\right] t^2 \\ &+ \left[\sum_{jk} \overline{[\![J_{\alpha k}J_{\alpha j}J_{jk}]\!]}\,\overline{[\![u_k^2]\!]} + \frac{1}{3}\left(\sum_{jk}\overline{[\![J_{\alpha j}J_{jk}J_{k\alpha}]\!]} + \sum_j \overline{\left[\!\left[\frac{dJ_{\alpha j}}{dt}J_{j\alpha}\right]\!\right]}\right)\overline{[\![u_\alpha^2]\!]} \right. \\ &\left. + \sum_k \overline{\left[\!\left[\Phi_{\alpha k}\sum_j J_{\alpha j}\Phi_{jk}\right]\!\right]}\,\overline{[\![\delta\mu_k^2]\!]}\right] t^3 .\end{aligned} \tag{3.63}$$

This relation can be applied to evaluate the effect, on a particular component $\alpha$, of initial errors acting selectively along a particular phase space direction $\beta$, or a combination of such directions. If the latter happens to be associated with a positive Lyapunov exponent the growth of error will occur from the very start of the evolution.

## 3.4 Error dynamics in illustrative systems

### 3.4.1 Bistable systems

Switching between simultaneously stable steady spaces is a common phenomenon in Meteorology in connection, for instance, with the transitions between different regimes of atmospheric circulation patterns (Charney and DeVore, 1979). The canonical form of evolution equation for a bistable system is (Nicolis, 1995)

$$\frac{dx}{dt} = \mu x - x^3 . \tag{3.64}$$

It admits the trivial steady state solution $x_0 = 0$, which is stable if $\mu < 0$ and, if $\mu > 0$, two additional ones given by $x_\pm = \pm\mu^{1/2}$, which are stable in this range of $\mu$ values, bifurcating from $x_0 = 0$ when $\mu$ crosses the value $\mu = 0$. By setting $x = x_N + u$, $\mu = \mu_N + \delta\mu$ the linearised error equation (3.6) becomes

$$\frac{du(t)}{dt} = \left(\mu_N - 3x_N^2\right) u(t) + \delta\mu\, x_N\,. \tag{3.65}$$

The solution of this equation for an initial error $u_0 = u$ is (we choose $\mu_N > 0$, $x_N = \mu_N^{1/2}$):

$$u(t) = \frac{\delta\mu}{2\mu_N^{1/2}} + \exp\left(-2\mu_N t\right)\left(u - \frac{\delta\mu}{2\mu_N^{1/2}}\right)\,. \tag{3.66}$$

Notice that in the absence of model error the initial error decreases here monotonously with time, owing to the choice of one of the stable fixed points as reference attractor. By squaring this expression and averaging over all values of $u$ sampled from a uniform distribution with $\overline{[\![u]\!]} = 0$ and $\overline{[\![u^2]\!]} = \epsilon^2$, one obtains the mean quadratic error:

$$\overline{[\![u^2]\!]} = \exp\left(-4\mu_N t\right)\left(\epsilon^2 + \frac{\delta\mu^2}{4\mu_N}\right) - \frac{\delta\mu^2}{2\mu_N}\exp\left(-2\mu_N t\right) + \frac{\delta\mu^2}{4\mu_N}\,. \tag{3.67}$$

By equalling to zero the time derivative of this expression we obtain the time $t_m$ at which a minimum is attained:

$$t_m = \frac{1}{2\mu_N}\ln\left(1 + 4\mu_N\frac{\epsilon^2}{\delta\mu^2}\right)\,. \tag{3.68}$$

We notice that $t_m$ is anticipated as the model error source term increases and delayed as the initial error source term increases. In fact the existence of a minimum at a finite time $t = t_m$ is here due entirely to the presence of model error, since $t_m \to \infty$ as $\delta\mu \to 0$ for $\epsilon$ fixed. This is a peculiarity of the class of models described by (3.64)-(3.65), where there is no coexistence of stable and unstable motions.

We come next to the $t$-expansion. By setting $J = -2\mu_N$ and $\Phi = \mu_N^{1/2}$ in Eq. (3.62) we get:

$$\begin{aligned}\overline{[\![u^2(t)]\!]} = \epsilon^2 \left(1 - 4\mu_N t + 8\mu_N^2 t^2 - \frac{32}{3}\mu_N^3 t^3\right) \\ + \delta\mu^2 \left(\mu_N t^2 - 2\mu_N^2 t^3\right) + \cdots .\end{aligned} \tag{3.69}$$

This is just the expansion of Eq. (3.67) to the order $t^3$, thereby illustrating the consistency of the general formulation of subsection 3.3.1 and in particular of subsection 3.3.5. The time $t_m$ to attain the minimum satisfies now a quadratic equation in $t$ whose (unique) positive solution reduces to (3.68) in the limit $\mu_N \to 0$, i.e. close to the bifurcation point. This provides an illustration of the discussion following Eq. (3.62), since in this limit the Jacobian tends to zero and the system becomes weakly dissipative.

Eq. (3.64) can also be solved exactly in the fully nonlinear regime. For $\mu > 0$ this solution reads:

$$x(t) = \frac{x(0)\mu^{1/2}}{\sqrt{x^2(0) + e^{-2\mu t}\left(\mu - x^2(0)\right)}} . \tag{3.70}$$

By evaluating this solution for a reference system $x_N$ where $\mu = \mu_N$, $x(0) = x_N(0)$ and a model system $x$ where $\mu = \mu_N + \delta\mu$, $x(0) = x_N(0) + u$, one has then access to the full nonlinear evolution of the instantaneous quadratic error $u^2(t)$. Fig. 3.1 depicts the time evolution of this quantity as computed from the exact expression with $x_N(0) = \mu_N^{1/2}$ (full line), the solution of the linearised equation for the error (Eq. (3.67), dashed line) and the $t$-expansion (Eq. (3.69), dotted line) for a small value of $\mu_N$. As can be seen the agreement is quite satisfactory, as expected from the comments made in connection with Eq. (3.62) and Eqs (3.68)-(3.69). The situation is quite different from $\mu_N$ values far away from the bifurcation point (Fig. 3.2). The $t$-expansion (dotted line) remains here close to the exact solution (full

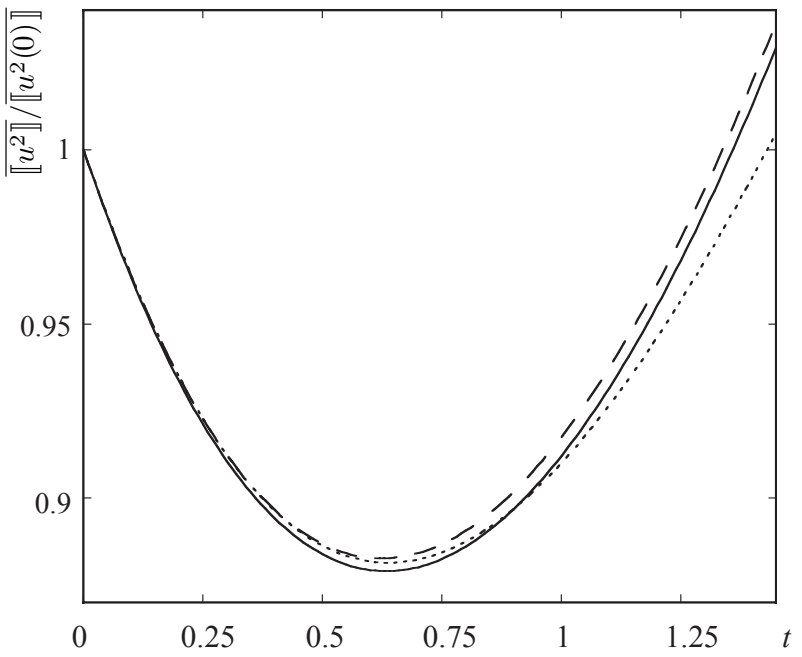



Figure 3.1: Time evolution of the mean quadratic error in the presence of both initial condition and model errors in the case of a bistable system Eq. (3.64) with $\mu = 0.1$. Initial condition errors are randomly sampled from a uniform distribution around the steady state solution of the exact system with $\epsilon^2 = 0.33 \times 10^{-6}$ and $\delta\mu = 10^{-3}$. The full line depicts the exact solution, the dashed line the linearised solution Eq. (3.67) and the dotted line the result of the $t$-expansion. The number of realisations considered is $2 \times 10^4$.

line) for a short period of time but shows subsequently a qualitatively different behaviour, missing the minimum versus time altogether.

The expansion can be improved significantly (Fig. 3.2, dashed-dotted line) by performing a partial resummation using a Padé approximant [Eq. (3.52)]. This approximation improves the $t^3$ expansion and prolongs its range of validity while ensuring at the same time its positivity.

We turn now to the computation of the crossover time $t_c$ introduced in the end of subsection 3.3.5 . By requiring that the contributions in $\epsilon^2$ and $\delta\mu^2$ in Eq. (3.67) reach the same value at $t = t_c$ we obtain a quadratic equation for

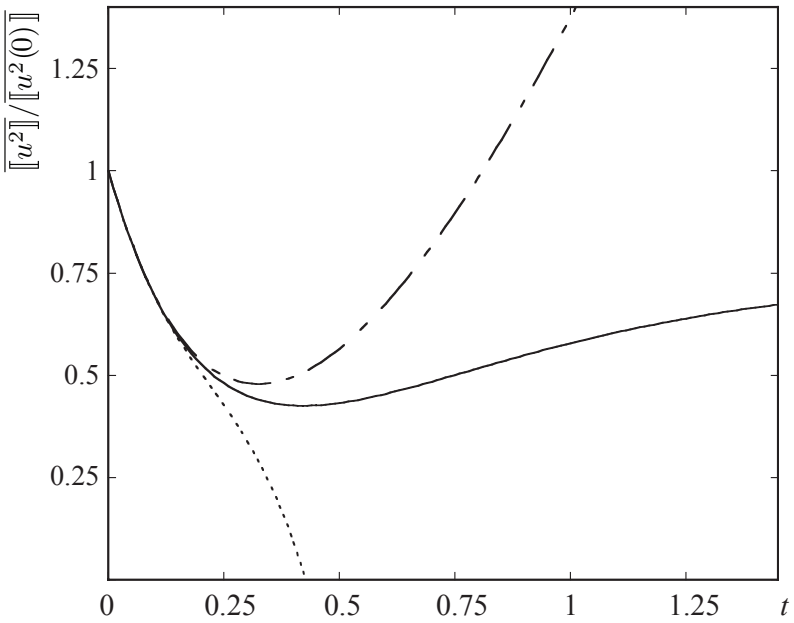



Figure 3.2: As in Fig. 3.1 but for $\mu = 1$. The dashed-dotted line stands for a Padé approximant of the $t$-expansion [Eq. (3.52)].

$\exp(-2\mu_N t_c)$, the solution of which yields:

$$t_c = \frac{1}{2\mu_N} \ln \left( \frac{\frac{\delta\mu^2}{2\mu_N} - 2\epsilon^2}{\frac{\delta\mu^2}{2\mu_N} - 2\epsilon} \right) . \tag{3.71}$$

In Fig. 3.3 the time dependencies of the $\epsilon^2$ and $\delta\mu^2$ parts of Eq. (3.67) and of its $t$-expansion [Eq. (3.69)] are plotted against time (full and dotted lines, respectively). In both cases crossover is found at a $t_c$ value close to the estimate of Eq. (3.71), which turns out to be significantly longer than the time $t_m$ for the total mean error $\overline{[\![u^2(t)]\!]}$ to attain the minimum as computed for the same parameter values (cf. Fig. 3.1). Notice that $t_c$ exists only as long as the model error, as measured by $\delta\mu^2/(2\mu_N)$, is sufficiently large as compared to the initial error.

The simplicity of the model studied in this subsection allows one to identify further the nature of the balance realised at the crossover time $t_c$. Clearly at

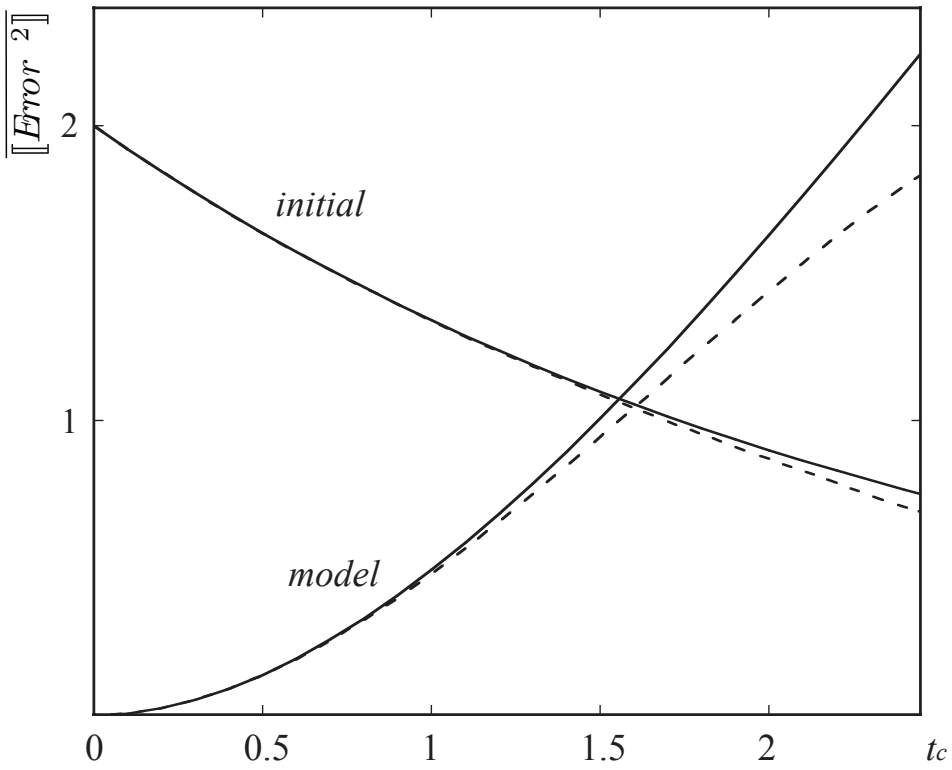



Figure 3.3: Time evolution and crossover time $t_c$ of the contributions of initial condition and model errors (normalised by $\epsilon^2$) considered separately from the linearised expression (3.67) (full lines) and from the $t$-expansion Eq. (3.69) (dashed lines) for the parameter values of Fig. 3.1.

$t = t_c$ the mean quadratic error does not vanish, in agreement with the general comment made in the end of subsection 3.3.5, since the right hand side of Eq. (3.67) does not admit a real-valued root: the two parts of the mean quadratic error do not cancel each other but, rather, attain equal magnitudes. It is only at the level of Eq. (3.66) for the (non-averaged) error itself that a cancellation is possible, provided that parameter $\delta\mu$ and initial errors $u$ have opposite signs. As a by-product, for such realisations the quadratic error $|u(t)|^2$ would possess a minimum at $u = 0$. In other words, the minimum of total error and matching of its two components are linked for this class of realisations. This is not so any longer for realisations in which $\delta\mu$ and $u$ have the same sign and, as a corollary, for the mean quadratic error itself.

### 3.4.2 Error dynamics around a saddle point

In the system considered in the preceding subsection the error evolution was taking place around a single, stable steady state solution of the reference system. Now, one of the signatures of the complexity of atmospheric dynamics is the coexistence of stable motions reflecting the presence of underlying regularities, and of unstable motions interrupting these regularities in a seemingly erratic way (Nicolis and Nicolis, 1995). In order to capture some of the aspects of this property we project Eq. (3.6) along the stable and unstable directions of a fixed point of a saddle type mimicking in this way, locally, what is expected to be happening globally in a strongly unstable, hyperbolic dynamical system. In the minimal case of a two-dimensional dynamics Eq. (3.6) becomes then:

$$\begin{aligned} \frac{du_1}{dt} &= \mu_N u_1 + \delta\mu x_N \\ \frac{du_2}{dt} &= -\lambda_N u_2 \,. \end{aligned} \tag{3.72}$$

Here $x_N$ is the coordinate of the saddle point along the $x$ axis, $\mu_N$ ($\mu_N > 0$) plays the role of both the control parameter subjected to uncertainty and of the positive Lyapunov exponent and $-\lambda_N$ ($\lambda_N > 0$) is the negative Lyapunov exponent, supposed not to be subjected to uncertainty. To satisfy the dissipativity condition we require $\lambda_N > \mu_N$.

The solution of Eqs. (3.72) reads:

$$\begin{aligned} u_1(t) &= u_1 e^{\mu_N t} + \frac{x_N}{\mu_N}\left(e^{\mu_N t - 1}\right)\delta\mu \\ u_2(t) &= u_2 e^{-\lambda_N t}\,, \end{aligned} \tag{3.73}$$

from which the quadratic error can be deduced:

$$\begin{aligned} u^2(t) =& u_1^2 e^{2\mu_N t} + u_2^2 e^{-2\lambda_N t} + \frac{x_N^2}{\mu_N^2}\left(e^{\mu_N t} - 1\right)^2 \delta\mu^2 \\ &+ 2\frac{u_1 x_N}{\mu_N} e^{\mu_N t}\left(e^{\mu_N t} - 1\right)\delta\mu\,. \end{aligned} \tag{3.74}$$

By averaging over random initial errors and across the attractor ( $\overline{[\![u_1]\!]} = \overline{[\![u_2]\!]} = 0$, $\overline{[\![u_1^2]\!]} = \overline{[\![u_2^2]\!]} = \epsilon^2$ ) we obtain:

$$\overline{[\![u^2(t)]\!]} = \epsilon^2\left(e^{2\mu_N t} + e^{-2\lambda_N t}\right) + \frac{x_N^2}{\mu_N^2}\left(e^{\mu_N t} - 1\right)^2 (\delta\mu)^2 \tag{3.75a}$$

and the corresponding $t$-expansion up to $O\left(t^3\right)$, the analog of Eq. (3.62) with $J_{11} = \mu_N$, $J_{22} = -\lambda_N$, $J_{12} = J_{21} = 0$ and $\Phi_1 = x_N$, $\Phi_2 = 0$:

$$\begin{aligned} \overline{[\![u^2(t)]\!]} =& 2\epsilon^2 + 2\left(\mu_N - \lambda_N\right)\epsilon^2 t + \left[2\left(\mu_N^2 + \lambda_N^2\right)\epsilon^2 + x_N^2(\delta\mu)^2\right] t^2 \\ &+ \left[\frac{4}{3}\left(\mu_N^3 - \lambda_N^3\right)\epsilon^2 + x_N^2\mu_N(\delta\mu)^2\right] t^3 + \cdots\,. \end{aligned} \tag{3.75b}$$

By setting the time derivative of $\overline{[\![u^2(t)]\!]}$ to zero one can evaluate, from Eqs. (3.75b), the time $t_m$ for a minimum to occur. Owing to the presence of the contributions due to $-\lambda_N$ a minimum is bound to exist, even in the absence of model error. As pointed out earlier this is a general feature of systems in which stable and unstable motions coexist (in this respect, the example of subsection 3.4.1 is an exception). However, since model error gives a positive contribution to the

time derivative it tends to anticipate $t_m$ in such a way that the contribution containing $\exp(-2\lambda_N t_m)$ can still cancel those containing the positive exponentials. This further confirms the trend found in the previous subsections. Figs. 3.4a-3.4b depict the time dependence of $\overline{[\![u^2(t)]\!]}$ according to the full expression [Eq. (3.75a), full line] and its $t$-expansion [Eq. (3.75b), dotted line] for two different values of $\mu_N$ while keeping the other parameters fixed. They confirm the second trend identified in the previous subsections, namely that as $\lambda_N$ tends to $\mu_N$ the agreement between the full and the approximate form of Eqs. (3.75) is improved.

We proceed by estimating the crossover time, $t_c$, at which the values of the contributions of initial and model errors are matching each other. Dividing through by $x_N^2\delta\mu^2/\mu_N^2$ in Eq. (3.75a) and expanding the square term one sees that at the crossover time $t_c$ the following condition must be satisfied:

$$\left(\frac{\epsilon^2\mu_N^2}{x_N^2\delta\mu^2}-1\right)e^{2\mu_N t_c}+\frac{\epsilon^2\mu_N^2}{x_N^2\delta\mu^2}e^{-2\lambda_N t_c}=1-2e^{\mu_N t_c}\,. \tag{3.76a}$$

Since the right hand side in (3.76a) is negative and the second term in the left hand side is positive, crossover requires that:

$$\frac{\epsilon^2\mu_N^2}{x_N^2\delta\mu^2}<1\,. \tag{3.76b}$$

In other words [see also comment at the end of the paragraph following Eq. (3.71)], model error needs to exceed a certain value that depends on initial error and on the parameters of the reference system. Figs. 3.5a,3.5b depict the initial and model error contributions as functions of time for the exact [Eq. (3.75a), full lines] and approximate [Eq. (3.75b), dotted lines] expression and for parameter values for which crossover is allowed and forbidden, respectively. Notice that in the first case $t_c > t_m$.

The above conclusions subsist in the case of anisotropically distributed initial errors [$\overline{[\![u_1^2]\!]} \neq \overline{[\![u_2^2]\!]}$ in Eq. (3.74)] provided that the magnitude of the error along

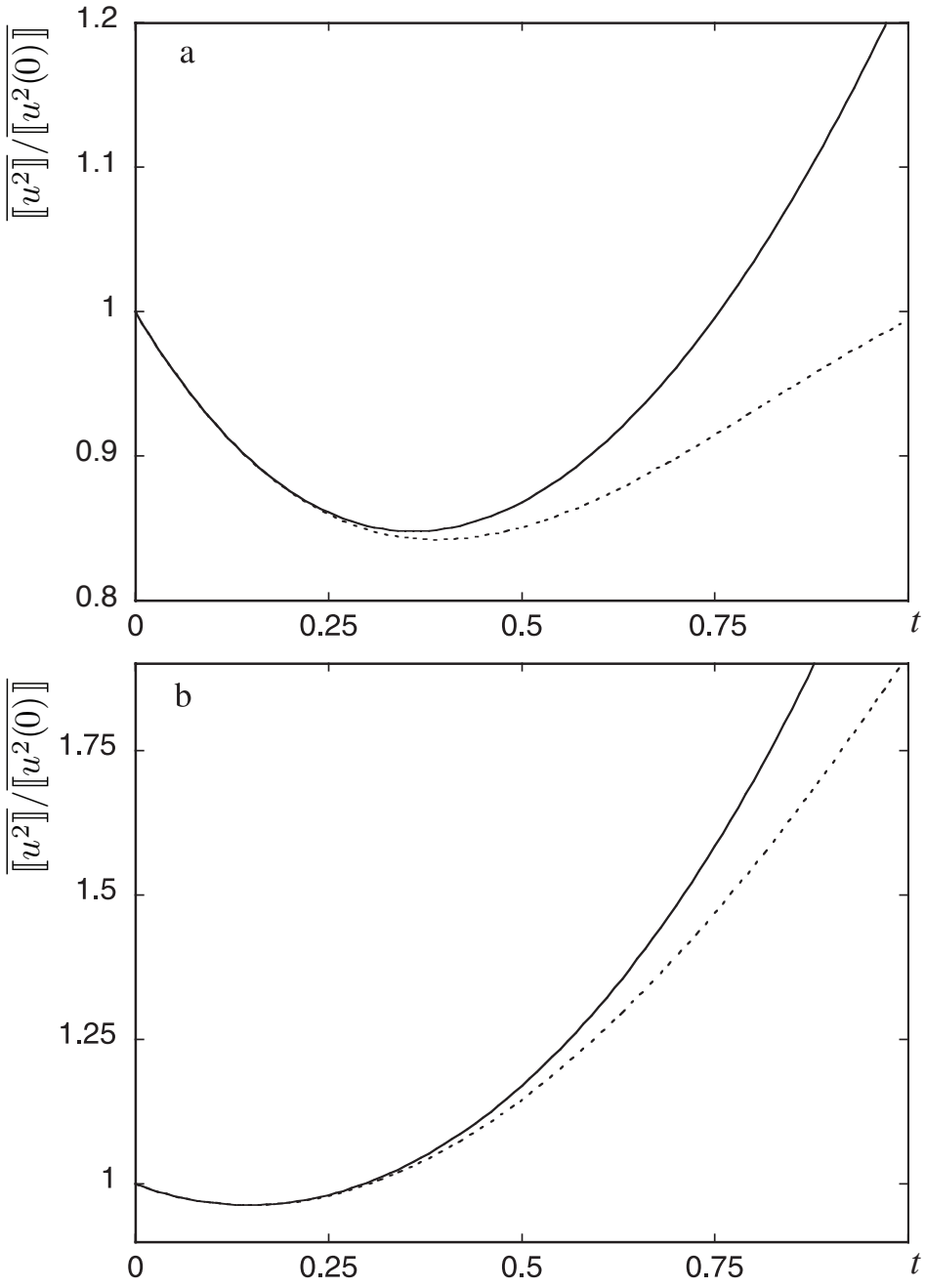



Figure 3.4: Time evolution of the mean quadratic error in the presence of both initial condition and model errors around a saddle point. Full lines stand for the exact expression (3.75a) and dotted lines for the corresponding $t$-expansion (3.75b). Parameter values are $\epsilon = \delta\mu = 10^{-3}$, $x_N = 1$, $\lambda_N = 1$, $\mu_N = 0.1$ (a), $\mu_N = 0.5$ (b).

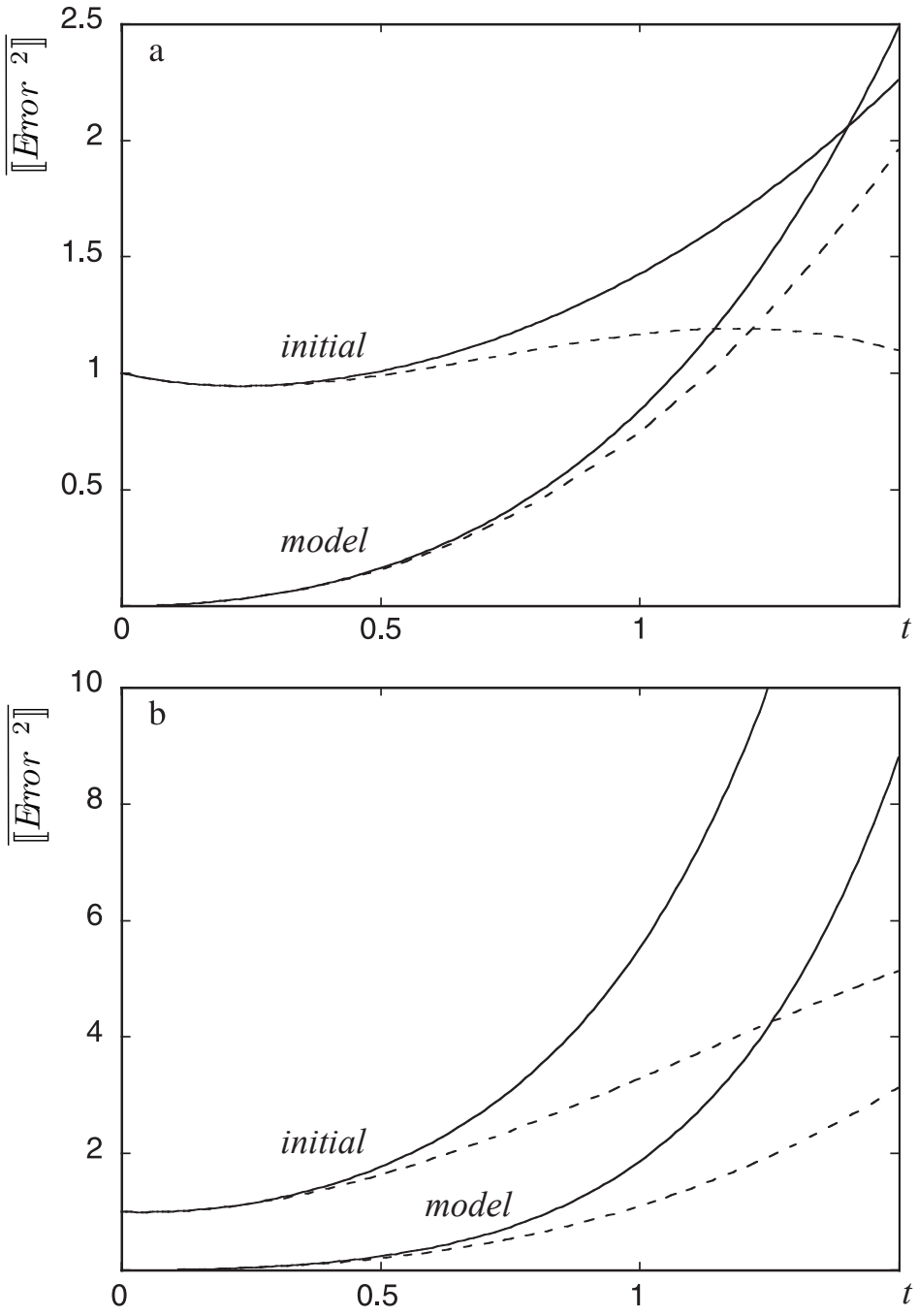



Figure 3.5: As in Fig. 3.3 but for the saddle point case. Parameter values as in Fig. 3.4 but $\lambda_N = 1$, $\mu_N = 0.5$ (a) and $\lambda_N = 1.5$, $\mu_N = 1.2$ (b).

the stable direction does not fall below some critical value, which depends on the ratio of the positive to the negative Lyapunov exponent.

### 3.4.3 Low order systems with chaotic dynamics

In this section we summarise results on the combined initial and model error dynamics in the light of the analysis of subsection 3.3.5, for low-order systems giving rise to deterministic chaos.

#### 3.4.3.1 The "Lorenz (1963)" system

We begin with the Lorenz's three-mode truncation of the Boussinesq equations of thermal convection (Lorenz, 1963):

$$\begin{aligned} \frac{dx}{dt} &= \sigma\left(-x+y\right) \\ \frac{dy}{dt} &= rx-y-xz \\ \frac{dz}{dt} &= xy-bz \end{aligned} \tag{3.77}$$

where $x$ measures the rate of convective (vertical) turnover, $y$ the horizontal temperature variation and $z$ the vertical temperature variation. Parameters $\sigma$ and $b$ account, respectively, for the intrinsic properties of the material and for the geometry of the convective pattern. In what follows we focus on the role of parameter $r$, the (reduced) Rayleigh number, which provides a measure of the strength of the thermal constraint to which the system is subjected and is the main responsible for the thermal convection instability occurring in the system. Here $b$ and $\sigma$ are taken constant, therefore the parameter vector has only one component ($\mu = r$). The matrix $\mathbf{\Phi}$ is $3 \times 1$ and given by $\mathbf{\Phi} = (0, x, 0)^{\mathsf{T}}$, therefore the model error term in Eqs. (3.62) and (3.63) reduces to:

$$\mathbf{\Phi}\delta\mu = (0, x\delta r, 0)^{\mathsf{T}} \ . \tag{3.78}$$

As for the Jacobian matrix in (3.62) and (3.63), it now becomes:

$$\mathbf{J} = \begin{bmatrix} -\sigma & \sigma & 0 \\ r - z(t) & -1 & -x(t) \\ y(t) & x(t) & -b \end{bmatrix} . \tag{3.79}$$

In the computation of the mean square error, the state vector $(x, y, z)$ and the matrices $\mathbf{J}$ and $\mathbf{\Phi}$ are computed for points on the attractor after a long integration of 100 time units of the Lorenz system from an arbitrary initial condition. A small perturbation is then added to the obtained point in the attractor. The procedure is performed $10^5$ times. The attractor is chosen to correspond to the typical values $r = 28$, $\sigma = 10$, $b = 8/3$, under which chaos occurs. Notice that $\sum_i \overline{[\![J_{ii}]\!]} = -(\sigma + b + 1)$ takes here a constant (state-independent) strongly negative value, reflecting the highly dissipative character of the ongoing dynamics.

Fig. 3.6 summarises results pertaining to the position of the minimum in the time evolution of the mean square error. The full lines are obtained by direct solution of the full Eqs. (3.77) and the dashed ones stand for the analytic results provided by the Padé approximant (3.52), corresponding to Eq. (3.62). The standard deviation of the initial error is $\overline{[\![u^2]\!]}^{1/2} = 10^{-3}$. As can be seen the position of the minimum in the absence of model error is displaced to the left (anticipated) as the model error is increased, thus confirming further the trend found in the preceding subsections. Furthermore, the numerical and analytic results are practically indistinguishable well beyond the minimum for model errors considerably exceeding the initial ones.

The relative importance of initial and model errors in the course of time is illustrated in Fig. 3.7. In both cases a crossover is predicted, located (as in subsections 3.4.1 and 3.4.2) after the time at which the total error attains its minimum. For large model errors these times are very small – see also Fig. 3.6 with $\delta r = 5 \times 10^{-3}$. The situation is different for model errors comparable to

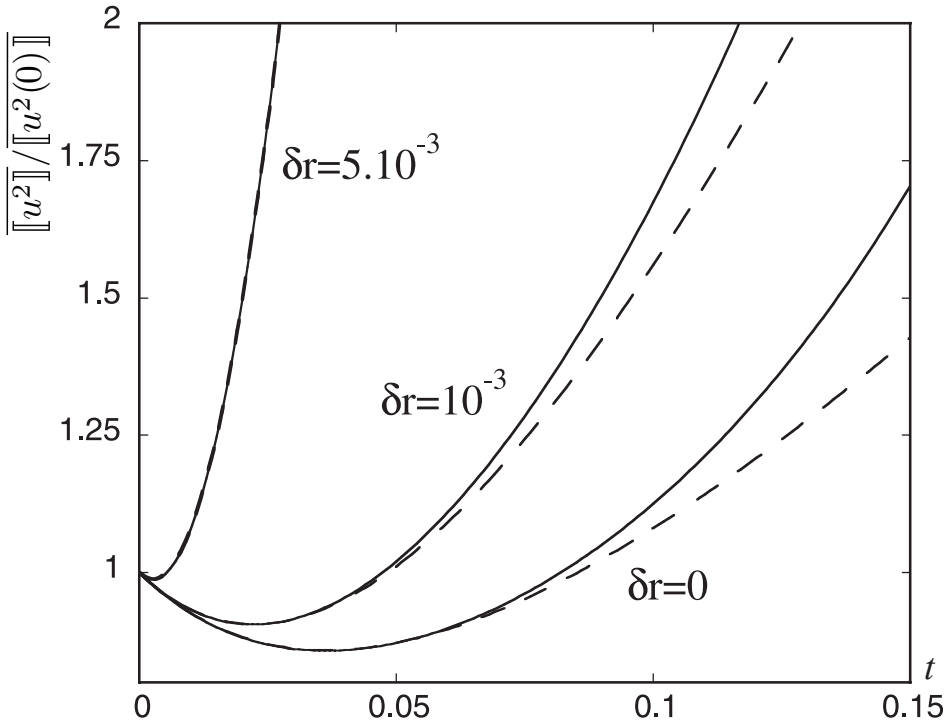



Figure 3.6: Short time behaviour of the mean square error, normalised by its initial value, as obtained numerically (full lines) from model (3.68) with $r = 28 + \delta r$, $\sigma = 10$, and $b = 8/3$ for three different values of $\delta r$. Dashed lines stand for the Padé approximant [Eq. (3.52)] of the corresponding third order analytic expansion [Eq. (3.62)]. Initial condition of the mean square error is $\overline{[\![||\mathbf{u}||^2]\!]} \equiv \overline{[\![u^2]\!]} = 10^{-6}$ and number of realisations for the averaging is $10^5$.

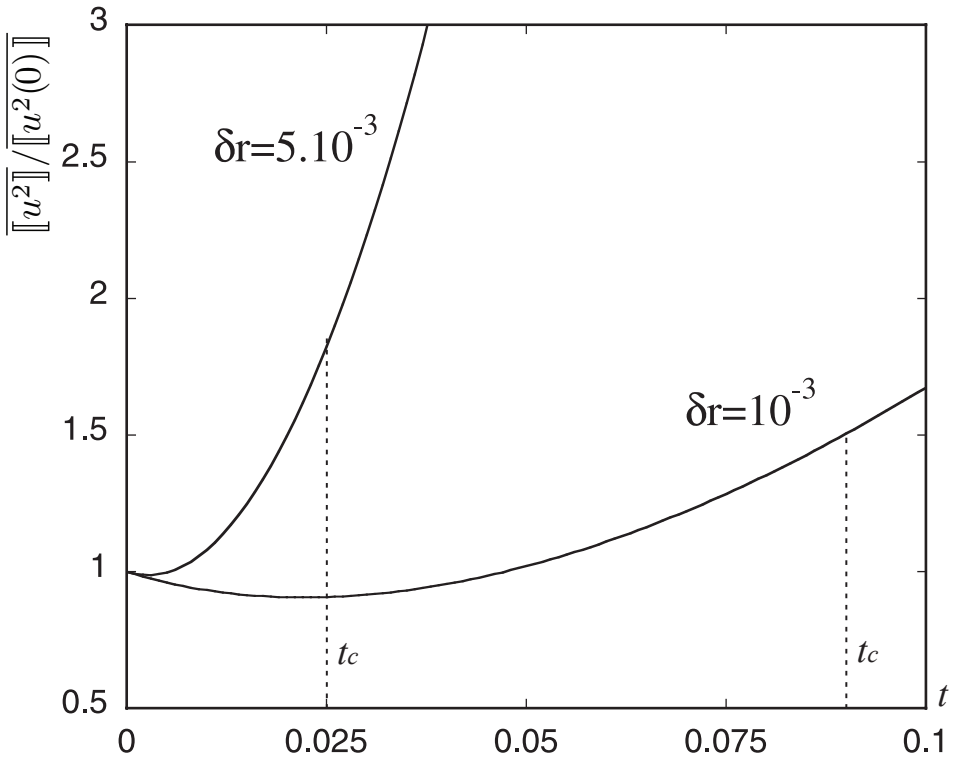



Figure 3.7: Crossover times, $t_c$, of initial and model errors (dotted lines) for two different magnitudes of error in the parameter $r$ of the model [Eq. (3.77)] as obtained from the corresponding analytic third order $t$-expansion [Eq. (3.62)]. The other parameters are as in Fig. 3.6.

initial ones, although in this case the initial error evolution becomes increasingly unsatisfactory when limited to $\mathcal{O}(t^3)$ terms.

As a further indicator of the relative roles of initial and model errors we depict, in Fig. 3.8, the structure and transient evolution of the error probability distributions in the absence ($\delta r = 0$, full lines) and in the presence ($\delta r = 5 \times 10^{-3}$, dashed lines) of model error. In both cases the initial errors of the variables $x$, $y$ and $z$ are sampled from a uniform distribution of zero mean and variance equal to $3.3 \times 10^{-4}$. For $\delta r = 0$ the bulk of probability density remains confined to a fairly narrow internal of values. At the same time a certain asymmetry is manifested, reflected by the tendency to develop a (rather modest) tail in the direction of large error values. The situation changes considerably under the combined action of initial and model errors: the distribution is now much broader and displays, transiently, a bimodal structure at times that are considerably longer than the minimum or the crossover times (Figs. 3.6 and 3.7). We are probably dealing here with an intermediate to long time type of effect, reflecting the increasing delocalisation of the system in phase space induced by the presence of model error.

Up to now, random (unbiased) initial condition errors were used in the numerical experiments. However, there is evidence that systematic initial condition errors are present in the analyses used for operational forecasts. These systematic errors can arise either from systematic observational biases coming, for instance, from a progressive degradation of the quality of a measurement device (see e.g. Kalnay, 2003) or from the data assimilation procedure, which uses an imperfect model displaying some systematic drifts. The presence of such biases has been amply demonstrated during the reanalysis experiments performed at the NCEP (Kistler et al., 2001) or at the ECMWF (Simmons et al., 2004), through a detailed

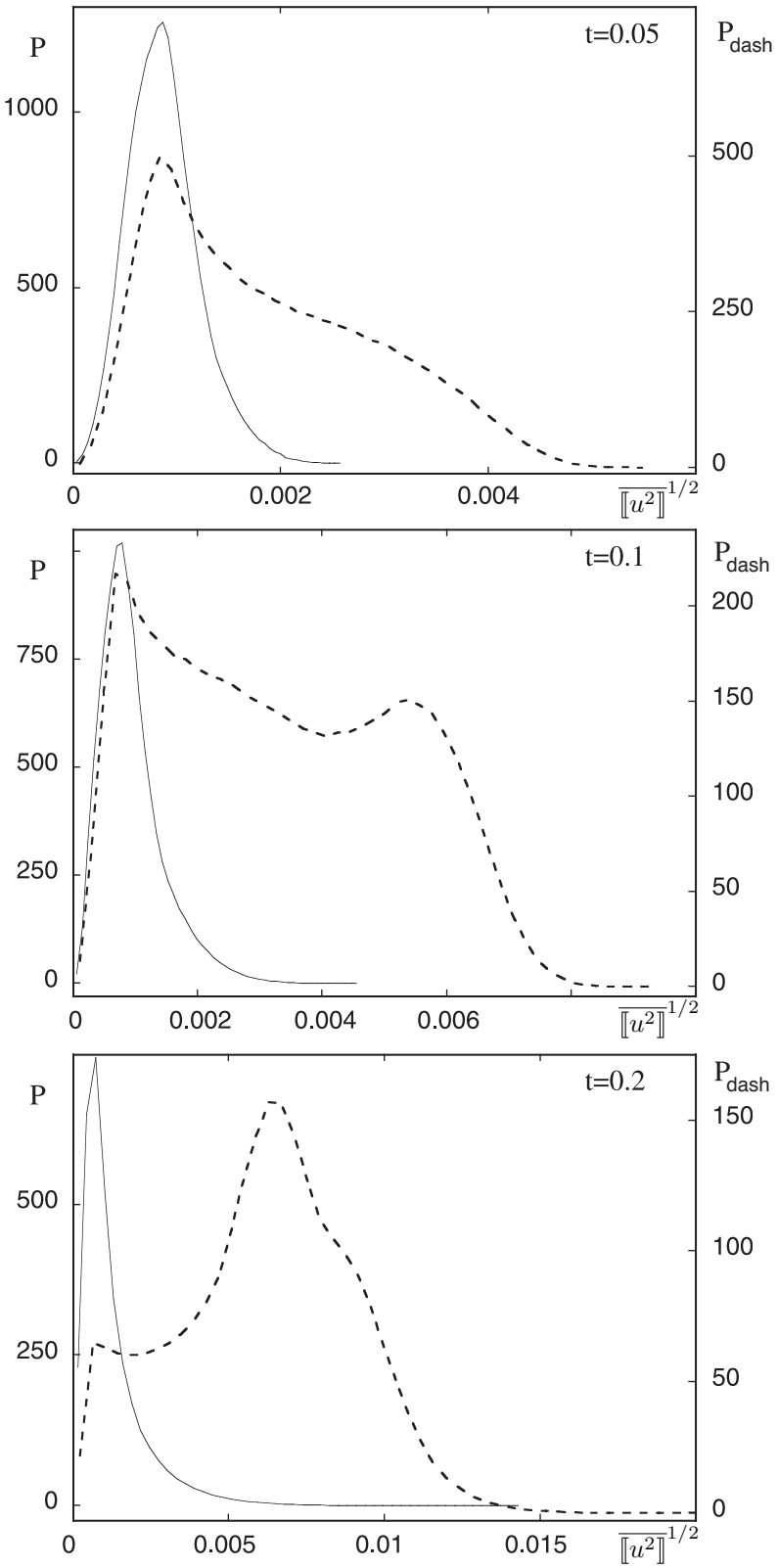



Figure 3.8: Short to intermediate probability density of $\overline{[\![u^2]\!]}^{1/2}$ in the absence of model error (full lines) and in the presence of model error, $r = 28 + \delta r$ with $\delta r = 5 \times 10^{-3}$ (dashed lines). Other parameters are as in Fig. 3.6 and number of realisations is $10^6$.

comparison with observed data.

A natural question to be raised concerns the impact of these systematic errors in operational forecasting on the predictability of the system at hand under the simultaneous presence of model errors. This point is briefly addressed here by introducing a systematic initial error for one of the variables of the Lorenz model, namely the variable $z$. The amplitude of this systematic error has been taken as equal to the standard deviation of the random part of error added to each model variable at the initial time.

Fig. 3.9a depicts the time of the minimum for the experiments, with and without systematic errors, as a function of the amplitude of the model error perturbation $\delta r$. An interesting feature is that the time of the minimum for positive values of $\delta r$ is now shifted toward larger values. In addition, the minimum (normalised by the initial value of the error) is deepening as compared with the case where systematic errors are absent (Fig. 3.9b). Notice that when the amplitude of the systematic error is increased further, the deepening of the minimum and its shift are also increased.

In the notation of subsection 3.3.5 and Appendix D, the numerical experiments summarised above correspond to $\overline{[\![\mathbf{u}]\!]} = (0, 0, s)$, $(s > 0)$ with $\mathbf{\Phi}$ as in Eq. (3.78). Under these conditions the coupling between initial and model errors is absent at the level of the first order term of the short time expansion [Eq. (D.1)] but gives a contribution at the level of the second order term [$3^{rd}$ last term in Eq. (D.2)]. Using the explicit form of the Jacobian we can see that this term yields a negative contribution, under the conditions of Fig. 3.9a. In fact, this contribution is equal to $(2J_{yz} + J_{zy})\,\Phi_y u_z \delta r = -x_A^2 s \delta r$. By discarding at this stage the $\mathcal{O}(t^3)$ term, this will thus tend to increase the value of the time of minimum, which will be

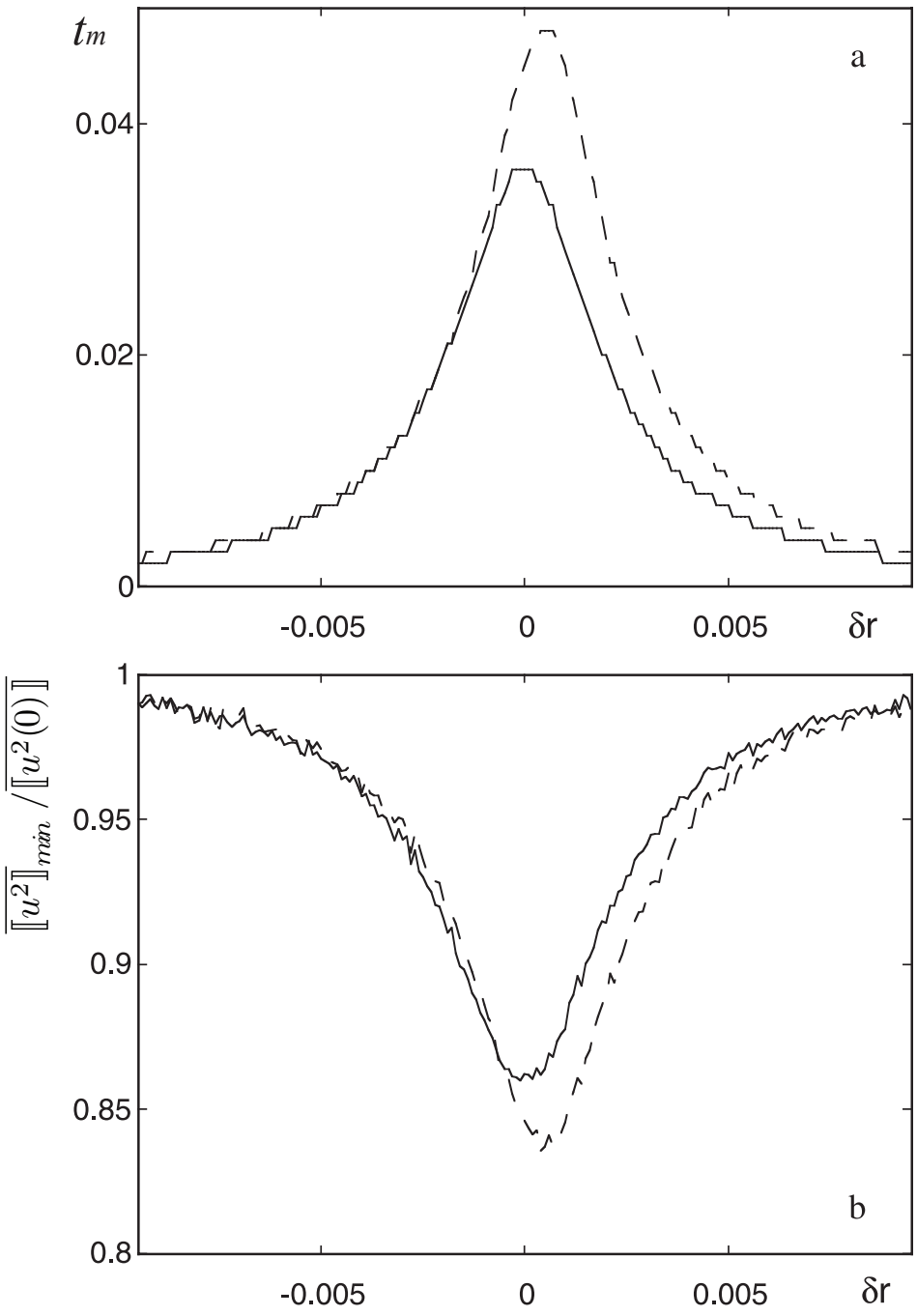



Figure 3.9: Time when mean square errors attain their minimum, $t_m$ (a), and relative size of the minimum (b), against the magnitude of the model error perturbation $\delta r$. Full lines stand for the case of unbiased initial condition errors and dashed lines for the case of biased ones. The amplitude of the bias is equal to the standard deviation of the random part of the initial error of the variables of model (3.77). Parameters are as in Fig. 3.6 and number of realisations is $10^5$.

given by:

$$t_m \approx \frac{-(\text{coefficient of } t - \text{term})}{2(\text{coefficient of } t^2 - \text{term})}\,, \tag{3.80}$$

in agreement with Fig. 3.9a. By applying this into the expression of the error one sees likewise that the value of the error at its minimum tends to decrease, in agreement with Fig. 3.9b.

In summary, a rich variety of behaviours can be found in the dynamics of the error in the Lorenz system for biased initial errors. In particular, a deepening of the error minimum and a shift of this minimum toward large times is obtained for some specific model and systematic errors. These features could have considerable operational implications since a model subjected to certain types of model errors could display different predictability properties depending on the presence or not of systematic errors in the initial conditions.

#### 3.4.3.2 A simple model for atmospheric circulation (Lorenz, 1984)

A low-dimensional model for long-term atmospheric circulation proposed by Lorenz (1984) is hereby considered as another example of a relevant dynamical system with chaotic behaviour. This model, obtained by a Galerkin projection of an infinite-dimensional model, is given by the following three-dimensional system of ordinary differential equations:

$$\begin{aligned} \frac{dx}{dt} &= -a\,x - y^2 - z^2 + a\,F \\ \frac{dy}{dt} &= -y + x\,y - b\,x\,z + G \\ \frac{dz}{dt} &= -z + b\,x\,y + x\,z \end{aligned} \tag{3.81}$$

This system has been thoroughly analysed in such papers as Lorenz (1986) and Broer et al. (2002) in the context of dynamical systems. Here, we briefly present the meaning of the terms involved in Eq. (3.81).

As far as variables are concerned, $x$ represents the intensity of the mean zonal flow, taking into account the intensity of the middle-latitude westerly wind current and the cross-latitude temperature gradient in each hemisphere, with wind and temperature fields assumed to be in permanent geostrophic balance; $y$ and $z$ represent the strength of the cosine and sine phases of a chain of vortices superposed on the zonal flow transporting the heat poleward.

As for the parameters, $F$ and $G$ are thermal forcings, $F$ representing the symmetric cross-latitude solar heating contrast and $G$ the asymmetric heating contrast between oceans and continents, varying with longitude. The coefficient $a$ measures the ratio of the damping rates of the westerly wind current to those of the vortices; that way, if $a < 1$ the westerly wind current is allowed to damp less rapidly than the vortices. The constant $b$ measures the ratio of the transport rate to the amplification rate; the terms where it is present ($b\,x\,y$ and $b\,x\,z$) represent displacement of the vortices due to interaction with the westerly wind.

The Jacobian matrix of the system with respect to the variables $x, y, z$ is given by:

$$\mathbf{J} = \begin{bmatrix} -a & -2\,y(t) & -2\,z(t) \\ y - b\,z(t) & x(t) - 1 & -b\,x(t) \\ b\,y(t) + z(t) & b\,x(t) & x(t) - 1 \end{bmatrix} . \tag{3.82}$$

Setting $a$ as a constant, the Jacobian of the system with respect to the parameters $\mu = (F, G, b)$ becomes:

$$\mathbf{\Phi} = \begin{bmatrix} a & 0 & 0 \\ 0 & 1 & -x(t)\,z(t) \\ 0 & 0 & x(t)\,z(t) \end{bmatrix} . \tag{3.83}$$

In the computation of the mean square error, the state vector $(x, y, z)$ and the matrices $\mathbf{J}$ and $\mathbf{\Phi}$ are computed for points on the attractor after a long integration of 1000 time steps of the system [Eq. (3.81)] from an arbitrary initial condition. A small perturbation is then added to the obtained point in the attractor. For

the numerical approach, numerical integration proceeds onward, using the Runge-Kutta method of fourth-order with a fixed step size $\delta t = 0.001$. For the analytical approach, Eq. (3.52) is invoked, with $\mathbf{J}$ and $\mathbf{\Phi}$ given respectively by Eqs. (3.82) and (3.83) above.

The procedure is performed $10^4$ times for "perfect model" (unperturbed system) and "imperfect model" (perturbed system) scenarios. The attractor is chosen to correspond to parameter values $a = 0.25$, $b = 4$, $F = 8$, $G = 0.9$, under which chaos occurs.

Fig. 3.10 depicts the short-to-intermediate time behaviour of the mean square error, normalised by its initial value, as obtained numerically (full lines) from model (3.81) with $a = 0.25$, $b = 4$, $F = 8$, $G = 0.9$ for three different values of $\delta G$: a) $\delta G = 0.04$; b) $\delta G = 0.02$; c) $\delta G = 0$. Dashed lines stand for the Padé approximant [Eq. (3.52)] of the corresponding third order analytic expansion [Eq. (3.62)]. The initial condition of the mean square error is $\overline{[\![||\mathbf{u}||^2]\!]} \equiv \overline{[\![u^2]\!]} = 10^{-6}$ and the number of realisations for the averaging is $10^4$. The short time regime is that for which $t < 0.07$, 0.07 being the inverse of the modulus of the largest Lyapunov exponent of this system under the current choice of parameters.

Within the short time regime the numerical and analytical approaches to the mean square error are rather coincident, thus showing the adequateness of the analytical approximation to replicating results obtained numerically. Divergence only occurs after that stage, i.e. beyond the limits of validity of the analytical approximation. The larger the parameter error, the greater and the sooner that divergence occurs.

Even though the graphical representation does not clearly show it, there is a slight initial decrease in the mean square error for the "perfect model" scenario ($\delta G = 0$), as noted by data inspection. As $\delta G$ increases, the mean square error

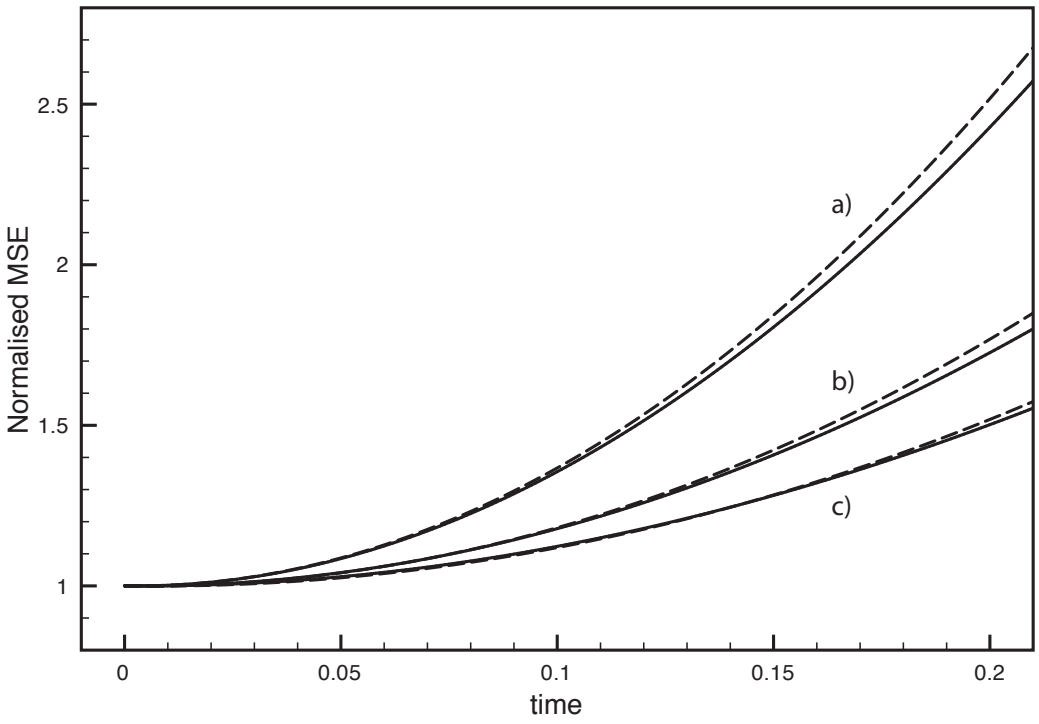



Figure 3.10: Short-to-intermediate time behaviour of the mean square error, normalised by its initial value, as obtained numerically (full lines) from model (3.81) with $a = 0.25$, $b = 4$, $F = 8$, $G = 0.9$ for three different values of $\delta G$: a) $\delta G = 0.04$; b) $\delta G = 0.02$; c) $\delta G = 0$. Dashed lines stand for the Padé approximant [Eq. (3.52)] of the corresponding third order analytic expansion [Eq. (3.62)]. Initial condition of the mean square error is $\overline{[\![||\mathbf{u}||^2]\!]} \equiv \overline{[\![u^2]\!]} = 10^{-6}$ and number of realisations for the averaging is $10^4$.

minimum is anticipated and its depth weakened even further.

The initial mean square error decrease is barely noticeable as a result of the trace of the Jacobian matrix in (3.82) being only very slighly negative, close to zero, under the conditions considered in the experiments the result of which are presented in Fig. 3.10. In particular, $\overline{Tr\,\mathbf{J}} \approx a = 0.25$ therefore $\tau_{min}$ (time at which mean square error is minimum) is within the order of the time-step.

A word is due on the relative importance of initial and model errors in the course of time, along with the crossover times between mean square errors stemming from those contributions. These are shown in Fig. 3.11, which depicts the time evolution of the mean square error, normalised by $10^{-6}$, of the analytical approximation to the model (3.81) with $a = 0.25$, $b = 4$, $F = 8$, $G = 0.9$, for a case in which only errors in the initial conditions are considered: $\overline{[\![\mathbf{u}^2]\!]} = 10^{-6}$ and $\delta G = 0$ (full line); and three scenarios without errors in the initial conditions: $\overline{[\![\mathbf{u}^2]\!]} = 0$ and $\delta G = 0.02$ (dotted line), $\delta G = 0.04$ (dashed line) and $\delta G = 0.06$ (dashed-dotted line). Crossovers occur when the error curve stemming from model errors crosses that from errors in the initial conditions ($t_c \approx 0.175$ for $\delta G = 0.06$, $t_c \approx 0.3$ for $\delta G = 0.04$). The number of realisations considered in the averaging is $10^4$.

Here, as in the systems considered previously, crossovers occur after the time at which the mean square error (under the combined influence of both initial conditions and model-related errors, see Fig. 3.10) attains its minimum. Moreover, they exist only as long as the model error contribution is sufficiently large as to eventually outweigh that of the initial conditions. Still in line with results for the previous systems, the time at which it occurs, $t_c$, is anticipated with the increase in the considered parameter errors. This is natural and intuitive, as the higher the model error contribution, the quicker its growth and thus the sooner it will

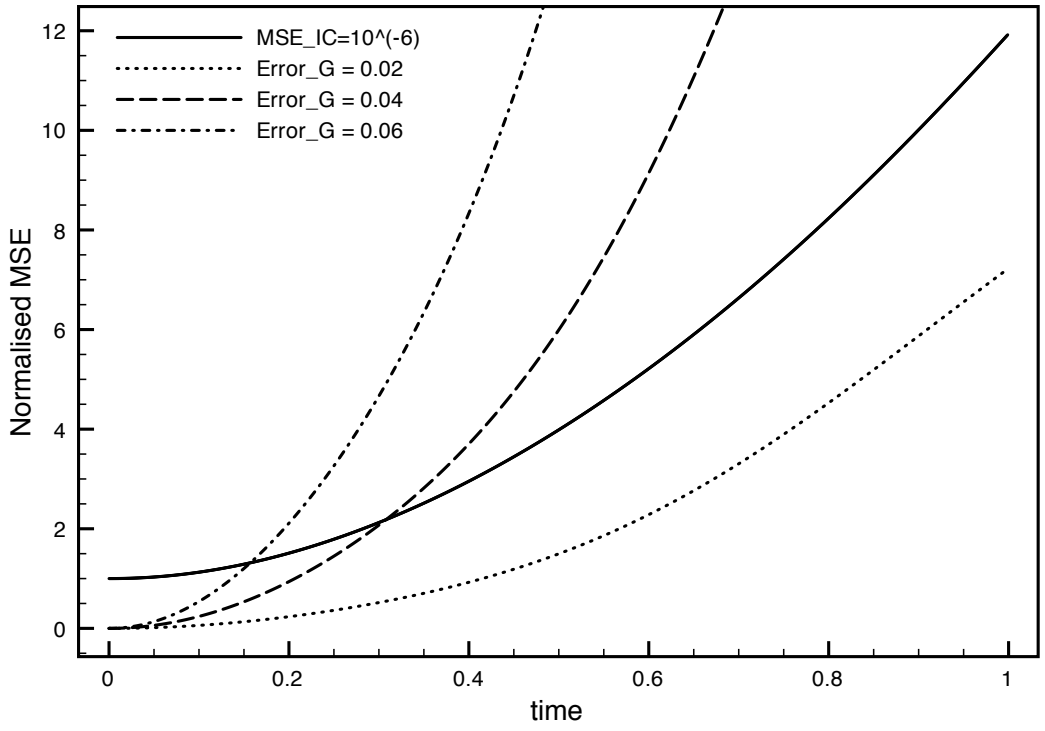



Figure 3.11: Time evolution of the mean square error, normalised by $10^{-6}$, of the analytical approximation to the model (3.81) with $a = 0.25$, $b = 4$, $F = 8$, $G = 0.9$, for a case in which only errors in the initial conditions are considered: $\overline{[\![\mathbf{u}^2]\!]} = 10^{-6}$ and $\delta G = 0$ (full line); and three scenarios without errors in the initial conditions: $\overline{[\![\mathbf{u}^2]\!]} = 0$ and $\delta G = 0.02$ (dotted line), $\delta G = 0.04$ (dashed line) and $\delta G = 0.06$ (dashed-dotted line). Crossovers occur when the error curve stemming from model errors crosses that from errors in the initial conditions ($t_c \approx 0.175$ for $\delta G = 0.06$, $t_c \approx 0.3$ for $\delta G = 0.04$). The number of realisations considered in the averaging is $10^4$.

match and outweigh the error contribution from initial errors.

Further experiments on this system reveal a significant redundancy in error features with respect to those observed in the previous subsection and are therefore not shown.

#### 3.4.3.3 A minimalist chaotic system (Rossler, 1976)

Our final example features one of the most emblematic "Rossler systems" (Rossler, 1976), introduced by Otto Rossler as a simple, minimalist model to address continuous-time chaos, motivated by the search for chaotic behaviour in far-from-equilibrum chemical kinetics. The chosen system is given by:

$$\begin{aligned} \frac{dx}{dt} &= -y - z \\ \frac{dy}{dt} &= x + a\,y \\ \frac{dz}{dt} &= b + z\,(x - c) \end{aligned} \tag{3.84}$$

Its Jacobian matrix with respect to the state variables $(x, y, z)$ is given by:

$$\mathbf{J} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & a & 0 \\ z(t) & 0 & x(t) - c \end{bmatrix}. \tag{3.85}$$

and the Jacobian of the system with respect to the parameters $(a, b, c)$:

$$\mathbf{\Phi} = \begin{bmatrix} 0 & 0 & 0 \\ y(t) & 0 & 0 \\ 0 & 1 & -z(t) \end{bmatrix}. \tag{3.86}$$

Depending on the chosen parameter values, this system has stationary, periodic, quasiperiodic and chaotic attractors. We will be interested in the latter.

The minimalism of such a system as a "recipe" for chaos is patent in its phase space having the minimal dimension three for chaos to occur, its nonlinearity being minimal as only a single quadratic term is featured, and its minimalist chaotic attractor having a single lobe.

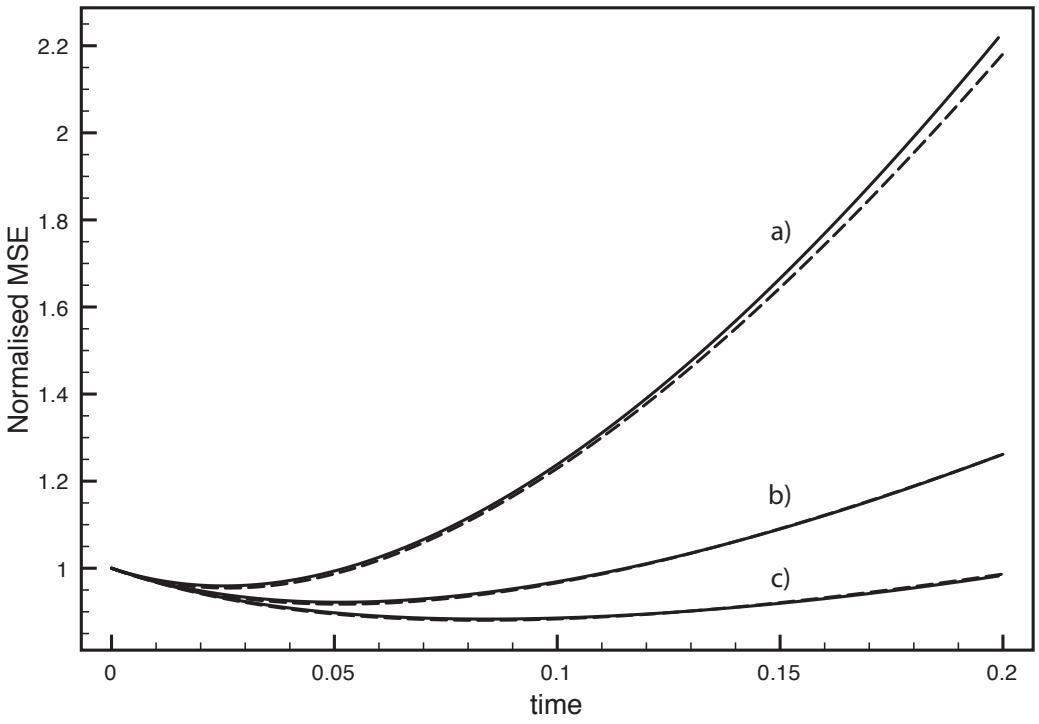



Figure 3.12: Short-to-intermediate time behaviour of the mean square error, normalised by its initial value, as obtained numerically (full lines) from model (3.84) with $a = b = 0.2$, $c = 5.7$ for three different values of $\delta c$: a) $\delta c = 0.018$; b) $\delta c = 0.012$; c) $\delta c = 0.006$. Dashed lines stand for the Padé approximant [Eq. (3.52)] of the corresponding third order analytic expansion [Eq. (3.62)]. Initial condition of the mean square error is $\overline{[\![||\mathbf{u}||^2]\!]} \equiv \overline{[\![u^2]\!]} = 10^{-6}$ and number of realisations for the averaging is $10^4$.

In the computation of the mean square error, the state vector $(x, y, z)$ and the matrices $\mathbf{J}$ and $\mathbf{\Phi}$ are computed for points on the attractor after a long integration of 1000 time steps of the system [Eq. (3.84)] from an arbitrary initial condition. A small perturbation is then added to the obtained point in the attractor. For the numerical approach, numerical integration proceeds onward, using the Runge-Kutta method of fourth-order with a fixed step size $\delta t = 0.001$. For the analytical approach, Eq. (3.52) is invoked, with $\mathbf{J}$ and $\mathbf{\Phi}$ given respectively by Eqs. (3.85) and (3.86) above. The procedure is performed $10^4$ times for "perfect model" (unperturbed system) and "imperfect model" (perturbed system) scenarios with $\delta c = 0.006$, 0.012 and 0.018. The attractor is chosen to correspond to well-known typical parameter values $a = b = 0.2$ and $c = 5.7$, under which chaos occurs. The results are shown in Fig. 3.12.

There, a fair degree of overlap between analytical and numerical solutions is noticeable within the validity of the short time regime ($t < 0.1$) in all three cases a)-c), with a remarkable overlap extending well into the intermediate regime when the parameter error is smaller ($\delta c = 0.012$ and $\delta c = 0.006$). This further stresses the adequateness of the analytical approximation as a relevant tool in assessing the dynamics of the mean square error.

A look is now taken at the relative importance of initial and model errors in the course of time, along with the crossover times between mean square errors stemming from those contributions. These are shown in Fig. 3.13, which depicts the time evolution of the mean square error, normalised by $10^{-6}$, of the analytical approximation to the model (3.84) with $a = b = 0.2$, $c = 5.7$, for a case in which only errors in the initial conditions are considered: $\overline{[\![\mathbf{u}^2]\!]} = 10^{-6}$ and $\delta c = 0$ (full line); and three scenarios without errors in the initial conditions: $\overline{[\![\mathbf{u}^2]\!]} = 0$ and $\delta c = 0.006$ (dotted line), $\delta c = 0.012$ (dashed line) and $\delta c = 0.018$ (dashed-dotted

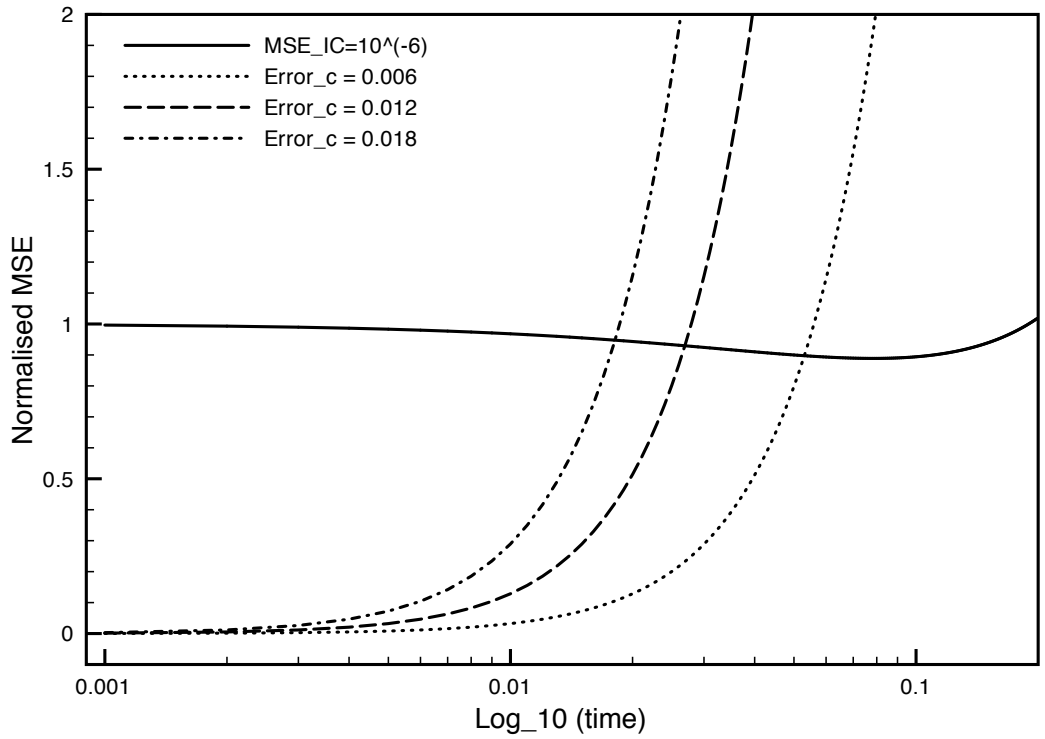



Figure 3.13: Time evolution of the mean square error, normalised by $10^{-6}$, of the analytical approximation to the model (3.84) with $a = b = 0.2$, $c = 5.7$, for a case in which only errors in the initial conditions are considered: $\overline{[\![\mathbf{u}^2]\!]} = 10^{-6}$ and $\delta c = 0$ (full line); and three scenarios without errors in the initial conditions: $\overline{[\![\mathbf{u}^2]\!]} = 0$ and $\delta c = 0.006$ (dotted line), $\delta c = 0.012$ (dashed line) and $\delta c = 0.018$ (dashed-dotted line). Crossovers occur when the error curve stemming from model errors crosses that from errors in the initial conditions ($t_c \approx 0.018$ for $\delta c = 0.018$, $t_c \approx 0.027$ for $\delta c = 0.012$ and $t_c \approx 0.055$ for $\delta c = 0.006$). The time is presented in logarithmic scale and the number of realisations considered in the averaging is $10^4$.

line). Crossovers occur when the error curve stemming from model errors crosses that from errors in the initial conditions. The crossover times are $t_c \approx 0.018$ for $\delta c = 0.018$, $t_c \approx 0.027$ for $\delta c = 0.012$ and $t_c \approx 0.055$ for $\delta c = 0.006$. The number of realisations considered in the averaging is $10^4$. Times are presented in logarithmic scale given how early and close in time crossovers occur, at times $t_c \ll 1$.

Here, unlike systems considered previously, crossovers occur after the time at which the mean square error (under the combined influence of both initial condition and model-related errors, Fig. 3.12) attains its minimum, as the errors stemming from the contribution of $\delta c$ are already fairly large, rapidly outweighing the contribution from errors in the initial conditions. In line with results for the previous systems, the time at which the crossovers occur, $t_c$, is anticipated with the increase in the considered parameter errors.

## 3.5 Conclusions

In this chapter some generic properties of the transient evolution of prediction errors under the combined effect of initial condition and of model errors have been derived, in the limit of small initial and parameter errors. We have begun by deriving a general formulation of the error dynamics on a Hilbert space with a generic metric, so as not to be formally limited to the low-order systems considered in the chapter, but rather being ready for tackling systems of higher complexity such as the case study in chapter 4. Given that in practical applications we consider finite-dimensional systems, we have particularised the formulation to a generic $n$-dimensional system.

A further particularisation has taken place by considering the short to intermediate time regime, as reflected by carrying out a power series expansion of the

error and its norm limited to the $\mathcal{O}(t^3)$ terms, along with the corresponding Padé approximant. The expansion accounts for arbitrary types of initial and model errors beyond the usually considered case of unbiased (random) uncorrelated ones and brings out clearly the mechanisms by which an initial error acting along a particular phase space direction ends up contaminating, in the course of time, phase space directions that were initially error-free.

Under the additional assumption of uncorrelated and unbiased initial errors a simplified expression has been derived in such a way as to allow us to identify conditions for the existence of a time at which mean quadratic errors attain a minimum, a crossover time at which the effects of initial conditions and of model errors match each other, or, possibily, the occurrence of inflexion points. In each case the role of the intrinsic dynamics and, in particular, its dissipative character and the interplay between stability and instability have been brought out.

The general properties have been tested and illustrated on a number of generic low-order models of atmospheric dynamics. In all cases considered (except for the Rossler system) the crossover time was shown to exceed the time of the minimum. Some quantitative relations have been obtained showing how the time of minimum is shifted as the magnitude of the model error is increased. The case of biased (systematic) initial errors has been considered on a model giving rise to deterministic chaos and shown to be responsible for some qualitatively new properties.

In the next chapter a more general case study is considered for a spatially extended system with non-Euclidean metrics and eventual anisotropies in the initial conditions. The general formulation derived in the present chapter will thus be tested and implemented in a scenario of greater complexity.

# Chapter 4

# Error dynamics in a Quasi-Geostrophic model

## 4.1 Introduction

In the present chapter a systematic assessment on the short term dynamics of prediction errors is given for a Quasi-Geostrophic (QG) model, considering the combined influence of both initial condition and model-related errors and their relation to intrinsic properties of the underlying system. For this purpose, the generic analytical approach to the error dynamics presented in chapter 3 is applied and validated with numerical experiments. While some generic features are identified that come in agreement with those seen in lower-order systems, further properties of physical relevance, stemming from the higher complexity of the model, are also unveiled.

A brief overview on a QG model well known in Dynamic Meteorology (Marshall and Molteni, 1993) is given in section 4.2, followed by the error dynamics formulation in the context of the aforementioned QG model in section 4.3. In section 4.4 the forecasting-time evolution of prediction errors is assessed for the considered Quasi-Geostrophic model both analytically and numerically, thus allowing for some interesting features to be already brought out. The way the

minimum of the mean square error varies relative to parameter error variations is assessed in section 4.5. Section 4.6 outlines a probabilistic approach on the error dynamics, providing insight not only on the mean of the square error but also on other aspects of its statistical distribution. Finally, section 4.7 closes the present chapter with a general discussion and concluding remarks.

## 4.2 Quasi-Geostrophic (QG) model

A well-known three-level (L3) global Quasi-Geostrophic spectral model with triangular truncation T21 is considered in the present study, as in several references (e.g. Marshall and Molteni, 1993).

The forecasting-time evolution of the prediction errors in the QG model is to be analysed. For that purpose, we begin by briefly considering the Quasi-Geostrophic prognostic equation in Potential Vorticity:

$$\frac{\partial \mathbf{q}}{\partial t} = \mathbf{J}_s(\mathbf{q}, \mathbf{\Psi}) - \left[\mathbf{D}\mathbf{\Psi} + c_H \nabla^8 \xi_p + \mathbf{E}\mathbf{\Psi} + \mathbf{S}\right] = \mathbf{G}(\mathbf{q})\,. \tag{4.1}$$

The symbols present in (4.1) follow, with minor changes, the notation for QG used in such typical references as Marshall and Molteni (1993) or Holton (2004), where a comprehensive explanation is given on the topic. Still, for the sake of clarity and instructiveness, a brief review is hereby given on each term.

To begin with, the advective term $\mathbf{J}_s(\mathbf{q}, \mathbf{\Psi})$ is the Jacobian of the 2D field over the sphere, given in the tangent space $(x, y)$ by:

$$\mathbf{J}_s(\mathbf{q}_i, \mathbf{\Psi}_i) = \frac{\partial(\mathbf{q}_i, \mathbf{\Psi}_i)}{\partial(x, y)} = \frac{\partial \mathbf{q}_i}{\partial x}\frac{\partial \mathbf{\Psi}_i}{\partial y} - \frac{\partial \mathbf{q}_i}{\partial y}\frac{\partial \mathbf{\Psi}_i}{\partial x}\,, \tag{4.2}$$

where $\mathbf{\Psi}$ represents the Streamfunction, $\mathbf{q}$ the Potential Vorticity and $i$ represents the level at which the operation is performed (levels 1,2 and 3, corresponding respectively to 200 hPa, 500 hPa and 800 hPa).

As for the dissipative term **D**, it can be written as:

$$\mathbf{D} = \begin{pmatrix} -a_1 & a_1 & 0 \\ a_1 & -a_1 - a_2 & a_2 \\ 0 & a_2 & -a_2 \end{pmatrix}, \tag{4.3}$$

where $a_1 = \tau_R^{-1} R_1^{-2}$ and $a_2 = \tau_R^{-1} R_2^{-2}$, $R_j$ being the Rossby radii of deformation ($R_1 = 700$ km and $R_2 = 450$ km) appropriate to the 200-500 hPa and 500-800 hPa levels respectively and $\tau_R$ the radiative scale, taken as 25 days.

The term **E**, representing the Ekman dissipation, is given by:

$$\mathbf{E} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & k\nabla^2 + \nabla k \cdot \nabla \end{pmatrix}, \tag{4.4}$$

where $k(\lambda, \varphi, h) = \tau_E^{-1} \left[1 + \alpha_1 LS(\lambda, \varphi) + \alpha_2 FH(h)\right]$, $\tau_E = 3$ days, $\alpha_1 = \alpha_2 = 0.5$, $LS(\lambda, \varphi)$ is the fraction of land within a grid box at longitude $\lambda$ and latitude $\varphi$ and $FH(h) = 1 - \exp\left[-h/h_{\text{ref}}\right]$, $h_{\text{ref}} = 10^3$ m. $FH(h)$ models the dependency of the drag coefficient on height $h$. The symbol $\nabla$ denotes the 2D gradient operator over the sphere. $\nabla^2$ is thus the 2D Laplacian operator over the sphere.

Here and henceforth for the remainder of this chapter, the coordinates $\lambda$, $\varphi$ and $h$ denote, respectively, the longitude, latitude and height.

As for the coefficient $c_H$, it is taken as $c_H = \tau_{DIF}^{-1} r_E^8 [NT \cdot (NT+1)]^{-4}$, with diffusive time scale $\tau_{DIF} = 2$ days and $r_E$ being the average radius of the Earth.

The term $\xi_p$ is the Potential Vorticity in the $p$-system (pressure coordinate system), given by $\xi_p = \nabla^2 \mathbf{\Psi}$.

Finally, the forcing term is found in **S**. This time-independent yet spatially varying term constrains the solution of the model to an average, statistically stable, observed winter climatology.

The Potential Vorticity field at the three-level model, $\mathbf{q} = (q_1, q_2, q_3)^\mathsf{T}$, is given in algebraic form as:

$$\mathbf{q} = \mathbf{A}\boldsymbol{\Psi} + \mathbf{b} \tag{4.5}$$

in terms of the streamfunction field $\boldsymbol{\Psi} = (\Psi_1, \Psi_2, \Psi_3)^\mathsf{T}$, where the subscripts indicate the corresponding level. Here, $\mathbf{A} = (\nabla^2 - \mathbf{R})$, $\mathbf{R}$ being given by:

$$\mathbf{R} = \begin{pmatrix} c_1 & -c_1 & 0 \\ -c_1 & c_1 + c_2 & -c_2 \\ 0 & -c_2 & c_2 \end{pmatrix} , \tag{4.6}$$

where $c_1 = R_1^{-2}$, $c_2 = R_2^{-2}$, and the time-independent $\mathbf{b}$

$$\mathbf{b} = \begin{pmatrix} f + f_0 p_T/\Delta p \\ f \\ f(1 + h/H_0) \end{pmatrix} , \tag{4.7}$$

where $f = 2\Omega \sin\varphi$, $p_T$ is the pressure on top of the troposphere, $\Delta p$ is the pressure difference between levels and $H_0 = (\rho_0 g/\Delta p)^{-1} = 9 \times 10^3$ m.

The Streamfunction $\boldsymbol{\Psi}$ can be straightforwardly obtained from Eq. (4.5) with invertible $\mathbf{A}$:

$$\boldsymbol{\Psi} = \mathbf{A}^{-1}(\mathbf{q} - \mathbf{b}) . \tag{4.8}$$

Its perturbation is then given by:

$$\delta\boldsymbol{\Psi} = \mathbf{A}^{-1}\delta\mathbf{q} . \tag{4.9}$$

The Quasi-Geostrophic model is defined over a spheric surface, the metric of which is defined by the following inner product:

$$\langle \mathbf{a}, \mathbf{b} \rangle = \frac{1}{4\pi} \int_{-1}^{1} \int_{0}^{2\pi} \sum_{i=1}^{NL} a_i b_i^* d\nu d\lambda , \tag{4.10}$$

where $NL$ denotes the number of levels in the model, $\nu = \sin\varphi$ and the superscript $^*$ denotes the complex conjugate.

By expressing $\mathbf{a}$ and $\mathbf{b}$ with spherical harmonics, with $n$ being the total wavenumber, $m$ the zonal wavenumber and using the fact that $\mathbf{a}$ and $\mathbf{b}$ are real, Eq. (4.10) can be written as:

$$\begin{aligned}\langle \mathbf{a}, \mathbf{W}\mathbf{b} \rangle &= \sum_{l=1}^{NL} \sum_{n=0}^{NT} \sum_{m=-n}^{n} a_{l,m,n} b^*_{l,m,n} w_n \\ &= \sum_{l=1}^{NL} \sum_{n=0}^{NT} w_n \left[ a_{l,0,n,\Re} b_{l,0,n,\Re} + 2 \sum_{m=1}^{n} \sum_{\xi=\Re,\Im} a_{l,m,n,\xi} b_{l,m,n,\xi} \right] ,\end{aligned} \tag{4.11}$$

where $\Re$ and $\Im$ denote, respectively, the real and imaginary parts of the associated variable. $NT$ refers to the truncation and $NL$ to the number of levels considered in the spectral mode. For the three-level T21 triangular truncation spectral model considered in the present study, $NT = 21$ and $NL = 3$. $\mathbf{W}$ is the diagonal operator with weights $w_n$ taken to be dependent solely on the total wavenumber $n$. In the case $\mathbf{W} = \mathbf{I}$ we obtain the metric in (4.10), i.e. the sum over the $NL$ levels of the quadratic norm over the sphere. In the case $\mathbf{a} = \mathbf{b} = \mathbf{u}$, we take $\mathbf{N}(\mathbf{u}) = \langle \mathbf{u}, \mathbf{W}\mathbf{u} \rangle$.

For instance, in the Squared Streamfunction ($\mathbf{\Psi}^2$) metric $w_n = 1$, in the Kinetic Energy metric $w_n = n(n+1)$ and in the Enstrophy metric $w_n = [n(n+1)]^2$.

The operator $\mathbf{A}$ is self-adjoint with respect to the $\langle \cdot, \cdot \rangle$ metric, given that: 1) the Laplacian $\nabla^2$ is self-adjoint over the sphere, i.e. $\int f \nabla^2 g = \int g \nabla^2 f$ for all fields $f, g$ over the sphere; 2) the matrix $\mathbf{R}(3 \times 3)$ is symmetric. Since $\mathbf{A}$ is self-adjoint for the metric in (4.10), $\mathbf{A}^{-1}$ is self-adjoint as well.

## 4.3 Formulation on prediction errors for the QG model

We consider the state vector $\mathbf{x} = \Psi_{l,m,n,\xi} = \Psi_i$, where $i = (l, m, n, \xi)$ is an enumeration index spanning the whole spectrum, i.e. the whole domain of the quadruplet to which it corresponds. As in the previous section, $l = 1, \cdots, NL$ ; $n = 0, \cdots, NT$ ; $m = 0, \cdots, n$ ; $\xi = \Re, \Im$, leading to a state vector dimension $ND = 253 \times 2 \times 3 = 1518$.

In order to determine the Jacobian $\mathbf{J}$ as defined in (3.7), $\mathbf{J} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right)_{\mathbf{x}_N, \mu_N}$, one can invoke the prognostic equation for the perturbation in $\mathbf{q}$:

$$\begin{aligned} \frac{\partial \delta \mathbf{q}}{\partial t} &= \mathbf{J}_s(\delta \mathbf{q}, \mathbf{\Psi}) + \mathbf{J}_s(\mathbf{q}, \delta \mathbf{\Psi}) - \left[\mathbf{D}\delta\mathbf{\Psi} + c_H \nabla^8 \delta \mathbf{q} + \mathbf{E}\delta\mathbf{\Psi}\right] + \mathbf{p} \\ &= \frac{\partial \mathbf{G}(\mathbf{q})}{\partial \mathbf{q}} \delta \mathbf{q} + \mathbf{p}\,, \end{aligned} \tag{4.12}$$

where $\mathbf{p}$ is the model error term, given in (3.7).

Eq. (4.12) can equivalently be written as:

$$\frac{\partial \delta \mathbf{q}}{\partial t} = \mathbf{J}_s(\delta \mathbf{q}, \mathbf{\Psi}) + \mathbf{J}_s(\mathbf{q}, \mathbf{B}\delta \mathbf{q}) - \left[\mathbf{D}\mathbf{B}\delta\mathbf{q} + c_H \nabla^8 \delta \mathbf{q} + \mathbf{E}\mathbf{B}\delta\mathbf{q}\right] + \mathbf{p}\,, \tag{4.13}$$

where $\mathbf{B} = \mathbf{A}^{-1}$.

By applying the operator $\mathbf{B}$ to (4.12), the prognostic equation for the perturbation $\delta\mathbf{\Psi}$ is obtained:

$$\begin{aligned} \frac{\partial \delta \mathbf{\Psi}}{\partial t} &= \mathbf{B}\mathbf{J}_s(\mathbf{A}\delta \mathbf{\Psi}, \mathbf{\Psi}) + \mathbf{B}\mathbf{J}_s(\mathbf{q}, \delta \mathbf{\Psi}) - \left[\mathbf{B}\mathbf{D}\delta\mathbf{\Psi} + c_H \nabla^8 \delta \mathbf{\Psi} + \mathbf{B}\mathbf{E}\delta\mathbf{\Psi}\right] + \mathbf{B}\mathbf{p} \\ &= \frac{\partial \mathbf{F}(\mathbf{\Psi})}{\partial \mathbf{\Psi}} \delta \mathbf{\Psi} + \mathbf{B}\mathbf{p} = \mathbf{J}(\mathbf{\Psi})\delta\mathbf{\Psi} + \mathbf{B}\mathbf{p} = \mathbf{J}(\mathbf{\Psi})\delta\mathbf{\Psi} + \mathbf{\Phi}\delta\mu\,, \end{aligned} \tag{4.14}$$

where the tendency of the vector $\mathbf{\Psi}$ is given by $\mathbf{F}(\mathbf{\Psi}) = \mathbf{B}\mathbf{G}(\mathbf{q}) = \mathbf{B}\mathbf{G}(\mathbf{A}\mathbf{\Psi}+\mathbf{b}) = \partial\mathbf{\Psi}/\partial\mathbf{t}$.

The linear operator $\mathbf{J}$ is fully known when applied to the versors of a chosen orthonormalised basis. Therefore, the operator is in practice indistinguishable from a $ND \times ND$ matrix.

The Jacobian $\mathbf{J}$ can be obtained from (4.13) by taking $\delta\mathbf{q} = \mathbf{I}$ [or, equivalently, from (4.14) by taking $\delta\mathbf{\Psi} = \mathbf{I}$] and $\mathbf{p} = \mathbf{0}$ (no model error).

We consider an imperfect model where two parameters are perturbed in separate experiments: the forcing parameter [term $\mathbf{S}$ in Eq. (4.1)], and the diffusion time $\tau_{DIF} \propto c_H^{-1}$.

The forcing error is given by $\Delta\mathbf{S} = \mathbf{S}_{\text{perturbed}} - \mathbf{S}_{\text{unperturbed}} = \mathbf{S}_{\text{unperturbed}}\delta S$, where $\delta S$ is an adimensional error term. In a similar way, the error in diffusion time $\tau_{DIF}$ is given by $\Delta\tau_{DIF} = \tau_{DIF\,\text{perturbed}} - \tau_{DIF\,\text{unperturbed}} = \tau_{DIF\,\text{unperturbed}}\delta\tau_{DIF}$, where $\delta\tau_{DIF}$ is an adimensional error term.

The vector referring to the perturbation in the parameters, $\delta\mu$, is taken as a two-dimensional vector given by $\delta\mu = [\delta S, \delta\tau_{DIF}]^{\mathsf{T}}$ of adimensional perturbations.

The Jacobian with respect to the parameters that are to be perturbed is given by:

$$\mathbf{\Phi} = \left( \frac{\partial \mathbf{F}}{\partial \delta S} , \frac{\partial \mathbf{F}}{\partial \delta \tau_{DIF}} \right) = \left( \mathbf{B}\,\mathbf{S}_{\text{unperturbed}} , \mathbf{B}\,c_H \nabla^8 \xi_p \right) \tag{4.15}$$

calculated on the attractor of the perfect model, $\delta S = \delta\tau_{DIF} = 0$.

Given that in the equations of error evolution $\mathbf{J}$ and $\mathbf{\Phi}$ are always presented as operators, a computational routine has been prepared that determines the observable obtained from operating $\mathbf{J}$ or $\mathbf{\Phi}$ over a generic variable. The recursive application of that routine allows for higher-order powers of $\mathbf{J}$ and $\mathbf{\Phi}$ to be determined as well.

In our approach to the error dynamics, we shall need to determine time derivatives of the operators $\mathbf{J}$ and $\mathbf{\Phi}$ up to a certain order. For this purpose, we express

the second time derivative of $\delta\mathbf{\Psi}$ as:

$$\frac{\partial^2 \delta\mathbf{\Psi}}{\partial t^2} = \mathbf{J}^2\delta\mathbf{\Psi} + \mathbf{J}_1\delta\mathbf{\Psi} + \mathbf{J}\mathbf{\Phi}\delta\mu + \mathbf{\Phi}_1\delta\mu\,, \tag{4.16}$$

where the first time derivative of $\mathbf{J}$, denoted by $\mathbf{J}_1$, is such that:

$$\mathbf{J}_1\delta\mathbf{\Psi} = \mathbf{B}\mathbf{J}_s[\mathbf{A}\delta\mathbf{\Psi}, \mathbf{F}(\mathbf{\Psi})] + \mathbf{B}\mathbf{J}_s[\mathbf{G}(\mathbf{q}), \delta\mathbf{\Psi}]\,, \tag{4.17}$$

and the first time derivative of $\mathbf{\Phi}$, denoted by $\mathbf{\Phi}_1$, is given by:

$$\mathbf{\Phi}_1 = \left(0, \mathbf{B}c_H \nabla^8 \frac{\partial \xi_p}{\partial t}\right)\,, \tag{4.18}$$

where $\frac{\partial \xi_p}{\partial t} = \nabla^2\,\mathbf{F}(\mathbf{\Psi})$.

We then express the third time derivative of $\delta\mathbf{\Psi}$ as:

$$\frac{\partial^3 \delta\mathbf{\Psi}}{\partial t^3} = \mathbf{J}^3\delta\mathbf{\Psi} + (\mathbf{J}\mathbf{J}_1 + \mathbf{J}_1\mathbf{J})\,\delta\mathbf{\Psi} + \mathbf{J}_2\delta\mathbf{\Psi} + (\mathbf{J}_1\mathbf{\Phi} + \mathbf{J}\mathbf{\Phi}_1 + \mathbf{\Phi}_2)\,\delta\mu\,, \tag{4.19}$$

where the second time derivative of $\mathbf{J}$, denoted by $\mathbf{J}_2$, is such that:

$$\mathbf{J}_2\delta\mathbf{\Psi} = \mathbf{B}\mathbf{J}_s[\mathbf{A}\delta\mathbf{\Psi}, \mathbf{J}(\mathbf{\Psi})\mathbf{F}(\mathbf{\Psi})] + \mathbf{B}\mathbf{J}_s[\mathbf{A}\mathbf{J}(\mathbf{\Psi})\mathbf{B}\mathbf{G}(\mathbf{q}), \delta\mathbf{\Psi}], \tag{4.20}$$

and the second time derivative of $\mathbf{\Phi}$, denoted by $\mathbf{\Phi}_2$, is given by:

$$\mathbf{\Phi}_2 = \left(0, \mathbf{B}c_H \nabla^8 \frac{\partial^2 \xi_p}{\partial t^2}\right)\,, \tag{4.21}$$

where $\frac{\partial^2 \xi_p}{\partial t^2} = \nabla^2\,[\mathbf{J}\mathbf{F}(\mathbf{\Psi})]$.

For models of very large dimension $n$ as it is the case of the QG model, the operators $\mathbf{J}$ and $\mathbf{\Phi}$ are not obtained explicitly. Instead, they are obtained from computational algorithms involving operations in the spectral and spatial domain without ever making explicit $\mathbf{J}$ and $\mathbf{\Phi}$ as it the case of low order models (such as those in chapter 3). These operators are objects that are mathematically equivalent to matrices or applications in the domain $\mathbb{R}^{n\times n}$, where

$n$ is the dimension of phase space. Then, since we know how $\mathbf{J}$ and $\mathbf{\Phi}$ depend on state variables $\mathbf{x}$, the time derivatives of operators are simply obtained by the chain rule of the derivative of a composed function, in the form:

$$\mathbf{J}_1\,\mathbf{u} = \left(\frac{\partial \mathbf{J}}{\partial \mathbf{x}}\frac{d\mathbf{x}}{dt}\right)\mathbf{u} = \left[\frac{\partial \mathbf{J}}{\partial \mathbf{x}}\mathbf{f}'(\mathbf{x})\right]\mathbf{u}\,, \tag{4.22}$$

where $\mathbf{f}'(\mathbf{x})$ is the time derivative of the nonlinear model and $\mathbf{u}$ the error in the state variables (representing e.g. $\delta\mathbf{\Psi}$ in equations above). The same procedure holds for the time derivatives of $\mathbf{\Phi}$.

$\mathbf{J}_1$ is again an implicit operator, never made explicit in terms of a matrix. A similar approach can then followed for determining the second and higher-order derivatives. The applications of operators $\mathbf{J}, \mathbf{\Phi}$ and of their derivatives are thus swiftly computed for the QG model. Computing times for the overall analytical approach are discussed in section 4.7, after the actual experiments.

At this stage, we will be interested in the short term behaviour of the mean square error of either Potential Vorticity or the Streamfunction.

For this purpose, we implement the analytical approach derived in the previous chapter, more precisely sections 3.3.3.2 and 3.3.4 by considering the Mac-Laurin expansion of the mean square error in $\mathbf{\Psi}$ up to the third order in time, using Eq. (3.51), then to determine the corresponding Padé approximant of order [2:1], using Eq. (3.52).

The analytical approach is implemented, along with numerical tests, in the next sections, where further details are provided on the computation procedure.

## 4.4 Analytical vs. numerical implementation

In the present section, the formulation in Eq. (3.52) − with terms as in the previous section − is applied to the QG model, confronting the results obtained from the proposed analytical approach with those from numerical experiments.

For that purpose, several numerical experiments are carried out in order to assess the forecasting-time evolution of the mean square error in the Streamfunction, by considering the "Model minus Nature" approach and averaging over a sufficiently large sample of realisations of errors in the initial conditions, which in turn span the attractor so as to represent the average over the whole domain. The numerical integration scheme used in the present work is the first order Adams-Bashforth-Mouton (predictor-corrector) scheme, with a time-step of one hour. The results from such experiments are confronted with the outcome from the Padé approximations of the forecasting-time evolution of the mean square error, Eq. (3.52), so as to provide an insight on the quality of these approximations and reveal relevant features of the error behaviour.

The coefficients $c_k$ in Eq. (3.52) are given by Eq. (3.53) with $T_i = \overline{[\![U_i]\!]}/i!$, where $U_0$ is given by Eq. (3.42), $U_1$ by Eq. (3.44), $U_2$ by Eq. (3.46) and $U_3$ by Eq. (3.49). The terms $U_i$ depend on the $\mathbf{J}$, $\mathbf{\Phi}$ and their derivatives presented in section 4.3. In the averaging process leading to $\overline{[\![U_i]\!]}$, averages are firstly taken over an ensemble of initial conditions and then over the attractor. By considering the general case in which $U_i$ is a function of the initial error $\mathbf{u}_0$, of the error in the parameters $\delta\mu$ and of the state vector $\mathbf{x}$, we have:

$$\overline{[\![U_i]\!]} = \frac{1}{N_{IC}} \sum_{k=1}^{N_{IC}} \frac{1}{N_P} \sum_{l=1}^{N_P} \left[U_i \left(\mathbf{u}_{0l}, \mathbf{x}_k, \delta\mu\right)\right] \tag{4.23}$$

where $N_{IC}$ is the number of initial conditions considered over the attractor and $N_P$ the number of perturbation realisations in the initial conditions.

Note that, while permutation of these averages would be legitimate in the state-independent case, that would not hold in the state-dependent case.

In the experiments below both initial conditions and parameters are perturbed. The perturbations in the initial conditions are taken as state-independent in subsections 4.4.1 and 4.4.3, and state-dependent in subsection 4.4.2.

The perturbations are built from realisations of a standard Gaussian pseudo-random generator $w$. In the case of state-independent realisations, the $l^{\text{th}}$ realisation of the initial perturbation is solely a function of $w_l$:

$$\mathbf{u}_{0l} = \mathbf{u}_{0l}(w_l)\,, \tag{4.24}$$

whereas for the case in which perturbations depend on $\mathbf{x}_k$ (state-dependent), we have:

$$\mathbf{u}_{0l} = \mathbf{u}_{0l}(w_l, \mathbf{x}_k)\,, \tag{4.25}$$

where the algorithmic dependence on the state can eventually involve time integration of the system.

### 4.4.1 Uniformly distributed perturbations in the initial conditions and constant ones in the parameters

We begin by considering the case in which the perturbations in the initial conditions are introduced isotropically and homogeneously in the Streamfunction, with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$. Potential Vorticity is then determined [Eq. (4.5)] and its prognostic equation [Eq. (4.1)] integrated. Finally, from the resulting Potential Vorticity the final output in Streamfunction is obtained using Eq. (4.8).

As mentioned in the previous section, two parameters are perturbed in separate experiments: the forcing parameter $\mathbf{S}$ and the diffusion time $\tau_{DIF}$, by taking

their adimensional perturbations $\delta S$ and $\delta\tau_{DIF}$, respectively. In our tests, $\tau_{DIF}$ is taken to be 2 days.

The true distribution of the errors in the initial conditions is assumed to be known, therefore the analytical and numerical approaches recur to the same pseudo-random number generator, i.e. the error distribution used in both cases is the same. In the operational context, however, the true error distribution is not necessarily known, therefore the assumed distribution of errors in the initial conditions is not necessarily accurate – a problem not treated in the present study.

The results are presented, respectively, in Figs. 4.1 and 4.2.

Fig. 4.1 depicts the forecasting-time evolution of the mean quadratic error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the whole spectrum and levels, in the presence of uniformly distributed initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and with the addition of error in the adimensional forcing term $|\delta S| = 0.02$, with the Kinetic Energy metric over the sphere. The number of realisations considered in the averaging over the distribution of initial errors and the attractor is 5000.

In practice, 10 errors are taken for each of 500 initial conditions taken over the attractor (thus the total of 5000 realisations). These 500 initial conditions are obtained through an integration of 5000 days separated of 10 days each in order to obtain independent initial conditions. The 5000-day run is obtained after a relaxation time of 1000 days allowing the system to converge to the attractor.

An analogous test is portrayed in Fig. 4.2, but considering error in the diffusion time $|\delta\tau_{DIF}| \equiv |\mathrm{dTDIF}| = 0.2$ instead of that in the forcing term.

In both Figs. 4.1 and 4.2 the analytical approximation closely matches the numerical curve within the range of validity of the formulation. In fact, divergence

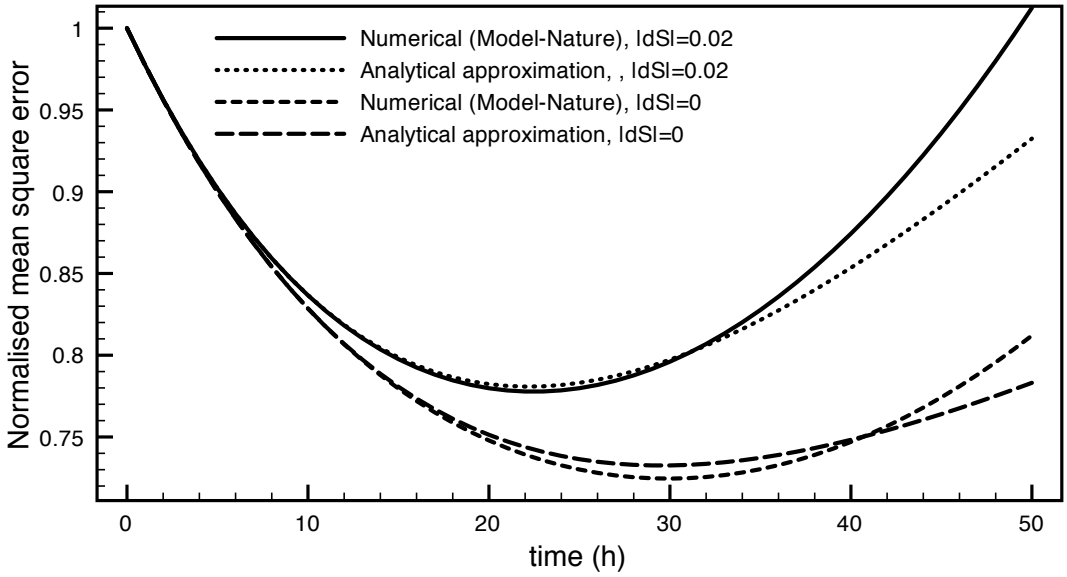



Figure 4.1: Forecasting-time evolution of the mean quadratic error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the whole spectrum and levels, in the presence of uniformly distributed initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and with the addition of error in the forcing term $|\delta S| = 0.02$, with the Kinetic Energy metric over the sphere. The number of realisations considered in the process is 5000.

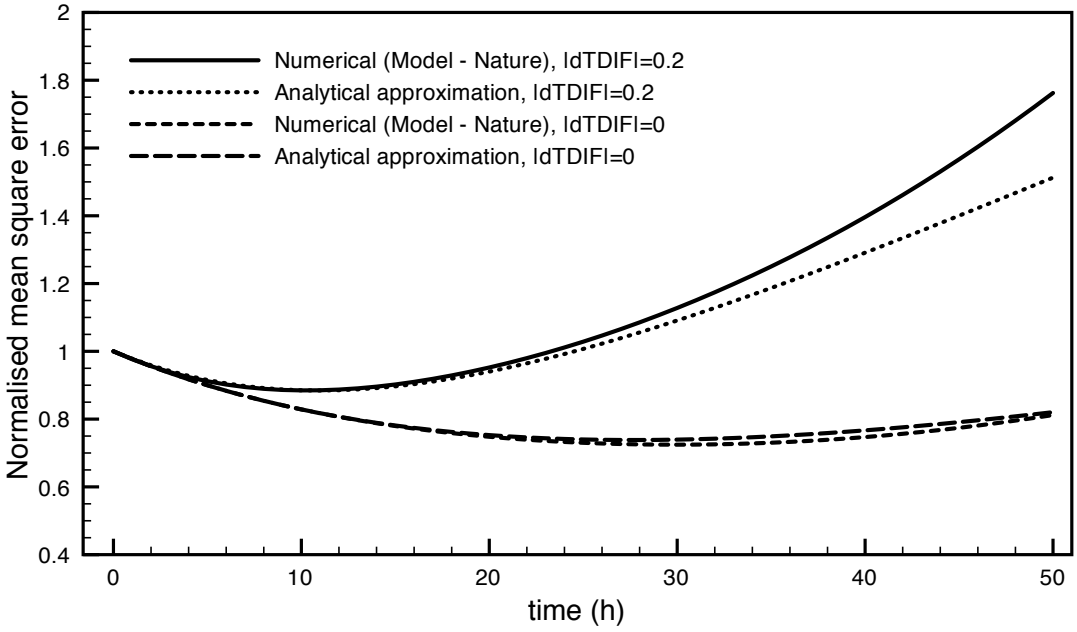



Figure 4.2: Same as Fig. 4.1, but considering error in the diffusion time $|\delta\tau_{DIF}| \equiv |\mathrm{dTDIF}| = 0.2$, instead of that in the forcing term.

only occurs beyond the limits of the short term scenario and after the minimum in mean square error is reached and accurately determined.

When only initial condition errors are present, the error evolution exhibits first a decrease, then to attain a minimum and finally moving on to a faster growth stage, in line with well-known error behaviour for this kind of system [see e.g. Fig. 3.6 or Nicolis, Perdigão, and Vannitsem (2009)].

It is known that systems with an initial dominance of negative Lyapunov exponents, a sign of stability, tend to exhibit an initial error decrease, followed by a rapid increase as the system aligns further with the unstable directions, with dominance of positive Lyapunov exponents. This explains the existence of the observed minimum in the mean square error of the horizontal velocity as seen in Figs. 4.1 and 4.2.

It is worth noting that when only parameter errors are present (not shown), the short term evolution of the mean quadratic error exhibits quadratic growth, also in line with well-known results, e.g. Vannitsem and Toth (2002); Nicolis (2003).

By combining initial condition and parameter errors, there is a displacement in the minimum of the mean square error of the horizontal velocity. In particular, it is anticipated and its depth weakened as compared to the case in which no parameter errors are present. This behaviour is expected, since adding parameter errors enhances the instability of the system, thus shortening the error decrease stage and anticipating its increase, leading in turn to a sooner and shallower minimum in the mean square error. These results are also consistent with those previously reported in chapter 3 (see e.g. Fig. 3.6). The sensitivity of the minimum of the mean square error to variations in parameter errors is more thoroughly addressed in subsection 4.5.

A look is now taken at the relative importance of initial and model errors in the course of time, along with the crossover times between mean square errors stemming from those contributions. Fig. 4.3 depicts the forecasting-time evolution of the mean quadratic error of the horizontal velocity (zonal and meridional), normalised relative to $\epsilon = 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and averaged over the whole spectrum and levels, with the Kinetic Energy metric over the sphere. The full line represents the case considering the presence of uniformly distributed initial condition errors with variance $\epsilon$ and no model error, whereas three model-imperfect scenarios are considered in the absence of errors in the initial conditions: $\delta S = 0.02$ (dotted line), $\delta S = 0.04$ (dashed line) and $\delta S = 0.06$ (dashed-dotted line). The number of realisations considered in the averaging over an ensemble of initial conditions and over the attractor is 5000.

Here, as in systems considered previously in subsections 3.4.1, 3.4.2, 3.4.3.1 and 3.4.3.2, crossovers occur after the time at which the mean square error (under the combined influence of initial and model errors, see e.g. Fig. 4.1) attains its minimum. Moreover, they exist only as long as the model error contribution is sufficiently large as to eventually outweigh that of the initial conditions. Still in line with results for previous systems, the time at which it occurs, $t_c$, is anticipated with the increase in the considered parameter errors.

The overall behaviour observed thus far for the error in the Quasi-Geostrophic model considered here comes in agreement with features brought out in low-order systems considered in the previous chapter and is naturally interesting to be seen here as well. Our formulation thus allows for the generic features seen in low-order systems to be brought out as well for a system bearing a higher degree of complexity.

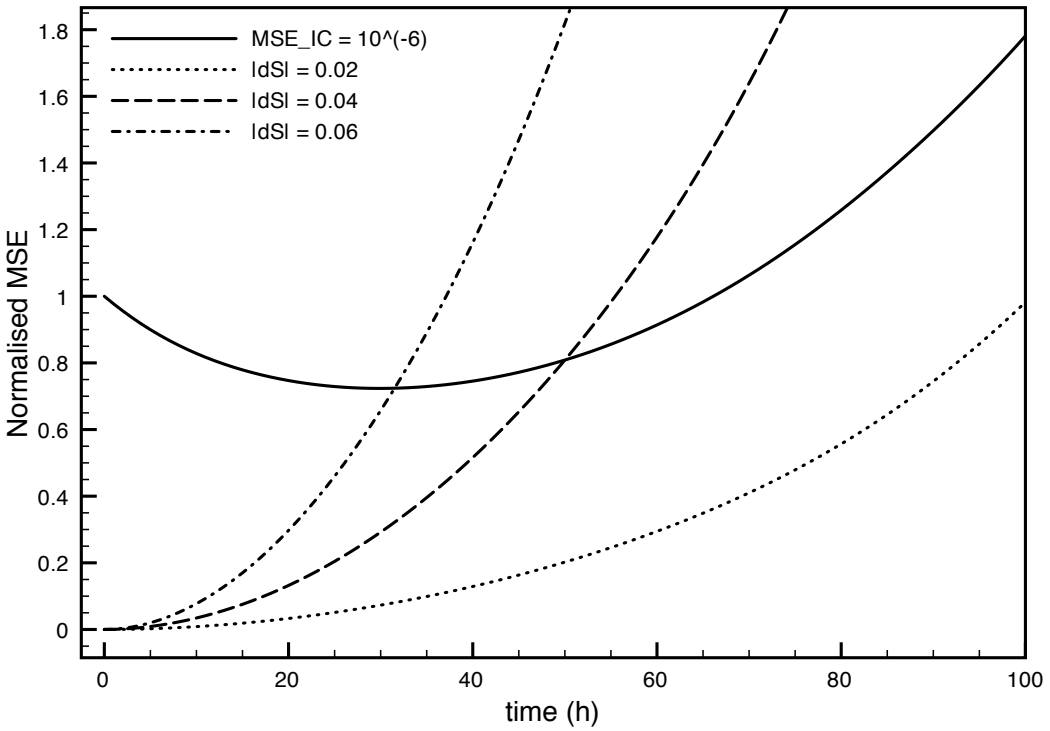



Figure 4.3: Forecasting-time evolution of the mean quadratic error of the horizontal velocity (zonal and meridional), normalised relative to $\epsilon = 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and averaged over the whole spectrum and levels, with the Kinetic Energy metric over the sphere. The full line represents the case considering the presence of uniformly distributed initial condition errors with variance $\epsilon$ and no model error, whereas three model-imperfect scenarios are considered in the absence of errors in the initial conditions: $\delta S = 0.02$ (dotted line), $\delta S = 0.04$ (dashed line) and $\delta S = 0.06$ (dashed-dotted line). The number of realisations considered in the averaging over an ensemble of initial conditions and over the attractor is 5000.

### 4.4.2 Anisotropic perturbations in the initial conditions and constant ones in the parameters

We consider now the case in which the ensemble of initial conditions is generated with the *breeding* method, which simulates the development of growing errors in an analysis cycle (Toth and Kalnay, 1997).

Previous studies on forecasting error growth on the QG model in this context include those of Pires (1996), Vannitsem and Nicolis (1997), Swanson et al. (1999) and Swanson et al. (2000).

The errors in initial conditions obtained from 4D variational data assimilation (4D-Var) are mostly projected onto the space spanned by breeding modes (Pires, 1996; Swanson et al., 1999). This justifies why we use such procedure to mimic data assimilation errors.

For the purpose of this section, a difference field is generated by Monte-Carlo at a time $t = -\tau$ prior to the forecasting stage. Then, it is led to evolve over a time period corresponding to an assimilation stage (*breeding time*, BT $= \tau$), during which it scaled down at time intervals (scaling periods) of 12 h. In our tests we consider BT $= 0$ h (spatial white noise errors), 24 h and 72 h. The resulting errors are then normalised to the Kinetic Energy of $10^{-6}\mathrm{m}^2\mathrm{s}^{-2}$. The normalisation factor is of the order of both the generated BGMs and the errors generated by Monte-Carlo, so that the experiments be comparable. Finally, the resulting normalised errors are fed as initial perturbations [$\mathbf{u}_{0l}$ from Eq. (4.23)] in the actual forecast, at $t = 0$.

Even though the states $\mathbf{x}$ at $t = -\tau$ are uniformly distributed, at $t = 0$ they are given by a propagator $\mathrm{P_{BT}}[\mathbf{x}(-\mathrm{BT})]$ representing the breeding process, which introduces anisotropy in the state distribution. The initial errors considered in the forecast will thus no longer be uniformly distributed; instead, they will

be gradually projected onto the unstable directions as the breeding time BT is increased.

As for parameter errors, they are generated as in the previous tests. The effects of the different BT on the error dynamics are presented in Fig. 4.4, for both numerical and analytical cases.

Given the qualitative redundancy in outcomes and conclusions from experiments concerning perturbations in the forcing and diffusion parameters, only results from experiments with the former are shown.

Fig. 4.4 depicts the forecasting-time evolution of the mean square error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the whole spectrum and levels, in the presence of uniformly distributed initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ subsequently subjected to breeding, with the addition of error in the forcing term $|\delta S| = 0.02$ and with the Kinetic Energy metric over the sphere. The number of realisations considered in the averaging over the distribution of initial errors and the attractor is 5000, corresponding to 10 errors taken for each of 500 initial conditions taken over the attractor. Three cases are considered: without breeding (BT=0) and with breeding times BT of 24 h and 72 h.

The analytical approximation proposed in this thesis is fairly accurate in the short term and even beyond, capturing the key error behaviour. This reveals the validity of the proposed formulation even for anisotropic distributions of initial errors. Moreover, the dynamical behaviour is consistent with what would be expected from the application of the breeding method to generate initial perturbations. As the breeding time increases, the system does, in fact, align itself further with the unstable directions, a behaviour that comes in agreement with such references as Pires (1996) and Swanson et al. (1999). This can be seen by

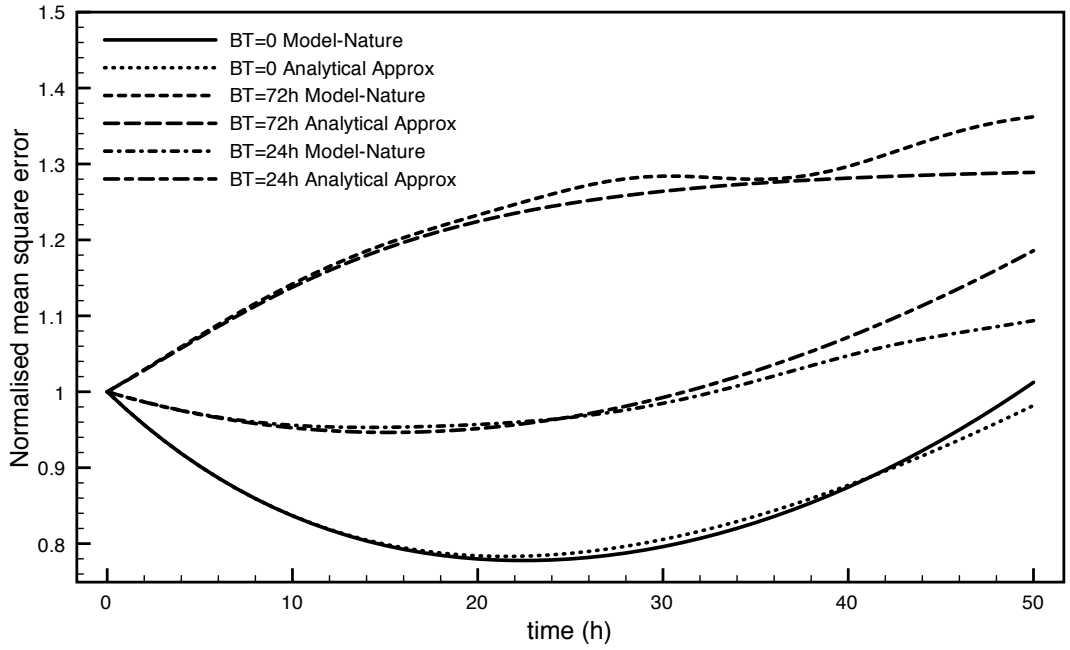



Figure 4.4: Forecasting-time evolution of the mean square error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the whole spectrum and levels, in the presence of uniformly distributed initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ subsequently subjected to breeding, with the addition of error in the forcing term $|\delta S| = 0.02$ and with the Kinetic Energy metric over the sphere. The number of realisations considered in the process is 5000. Three cases are considered: without breeding (BT=0) and with breeding times BT of 24 h and 72 h.

noting that the minimum of the mean square error is anticipated and weakened for greater breeding times, then to finally vanish altogether.

Experiments with 10000 realisations have been performed that showed only a very slight, barely noticeable improvement with respect to the results obtained for 5000 realisations (not shown due to redundancy of results). In fact, the relative difference of the respective mean square errors $E^2_{10^4}$ and $E^2_{5\cdot 10^3}$ lies within:

$$0.5\% \leq \frac{E^2_{10^4} - E^2_{5\cdot 10^3}}{E^2_{10^4}} \leq 2\% \,, \tag{4.26}$$

whereas the relative difference between numerically and analytically determined mean square errors is larger:

$$\frac{E^2_{\text{Numeric}} - E^2_{\text{Analytic}}}{E^2_{\text{Numeric}}} > 2\% \,. \tag{4.27}$$

Therefore, the comparison between numerically and analytically determined mean square errors is legitimate even with only 5000 realisations. The fact that the dimension of the attractor is much smaller than that of the whole phase space helps understand how a seemingly low number of realisations is enough for numerical convergence to occur.

### 4.4.3 Perturbations introduced at a particular scale of the motion

In the previous experiments, averages have been taken over the whole spectrum of 21 total wavenumbers $n$ considered in the spectral model. Now the formulation is tested for individual wavenumbers and error features are brought out for different scales.

Fig. 4.5 depicts the forecasting-time evolution of the mean square error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the three levels, in the presence of uniformly distributed

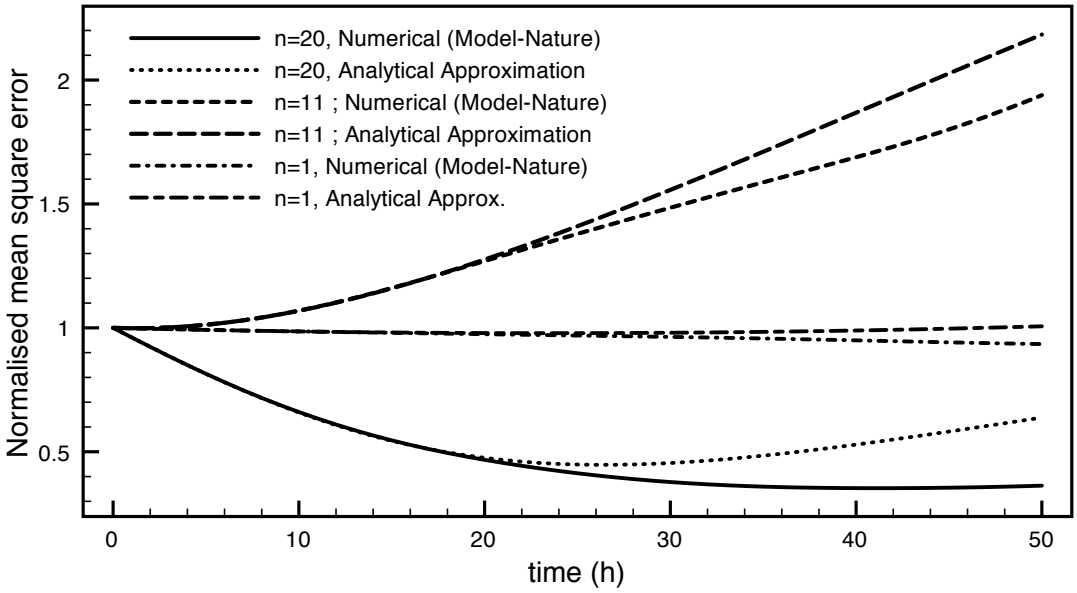



Figure 4.5: Forecasting-time evolution of the mean square error of the horizontal velocity (zonal and meridional), normalised relative to its initial value and averaged over the three levels, in the presence of uniformly distributed initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and with the addition of error in the forcing term $|\delta S| = 0.02$, with the Kinetic Energy metric over the sphere. The number of realisations considered in the process is 5000. Three total wavenumbers representing small ($n = 20$), intermediate ($n = 11$) and large scales ($n = 1$) are considered.

initial condition errors with variance $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ and with the addition of error in the forcing term $|\delta S| = 0.02$, with the Kinetic Energy metric over the sphere. The number of realisations considered in the averaging over the distribution of initial errors and the attractor is 5000. Three total wavenumbers representing small ($n = 20$), intermediate ($n = 11$) and large scales ($n = 1$) are considered.

As in previous cases, the mean square error is reasonably well represented by the analytical approximation within its range of validity, i.e. in the short term. Divergence occurs afterwards.

The error growth is stronger in the intermediate scales. This is consistent

with knowledge that the intermediate scales are the most unstable ones, unlike the more stable smaller and larger scales (Vannitsem and Nicolis, 1997).

## 4.5 Sensitivity of the M.S.E. minimum to parameter error variations

### 4.5.1 Perturbation in the initial condition and in the forcing term

In order to better analyse the displacement of the minimum of the mean square error for different forcing errors, let us observe Figs. 4.6 and 4.7.

Fig. 4.6 shows the time at which the mean square error of the horizontal velocity (zonal and meridional) is minimum, for different values of the error in the forcing term, in a Quasi-Geostrophic system [Eq. (4.1)]. The triangles refer to the case in which there is a systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions of the horizontal velocity. The squares refer to the case in which the initial condition errors are random, uniformly distributed as in Fig. 4.1. The circles refer to the case in which both systematic and random initial condition errors are considered. The number of realisations is 5000 and the metric used is that of Kinetic Energy.

Fig. 4.7 depicts the normalised minima of the mean square error of the horizontal velocity (zonal and meridional) for different forcing errors, as in Fig. 4.6, in the following cases: (a) systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions; (b) random, uniformly distributed initial condition errors as in Fig 4.6; (c) both systematic and random initial condition errors are considered. The normalisation factor is $1 \times 10^{-5}$.

From the referred figures it can be seen that, as the absolute value of the

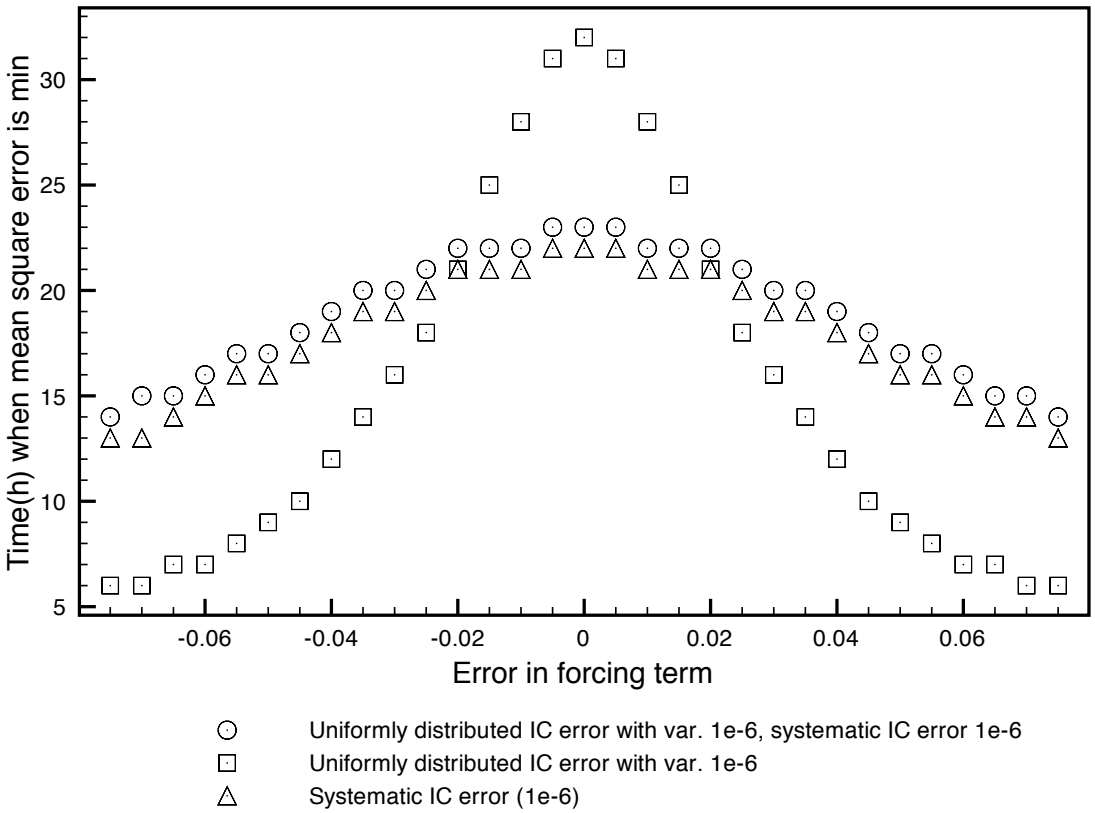



Figure 4.6: Time at which the mean square error of the horizontal velocity (zonal and meridional) is minimum, for different values of the error in the forcing term, in a Quasi-Geostrophic model [Eq. (4.1)]. The triangles refer to the case in which there is a systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions. The squares refer to the case in which the initial condition errors are random, uniformly distributed as in Fig. 4.1. The circles refer to the case in which both systematic and random initial condition errors are considered. The number of realisations and the metric used for averaging are as in Fig 4.1.

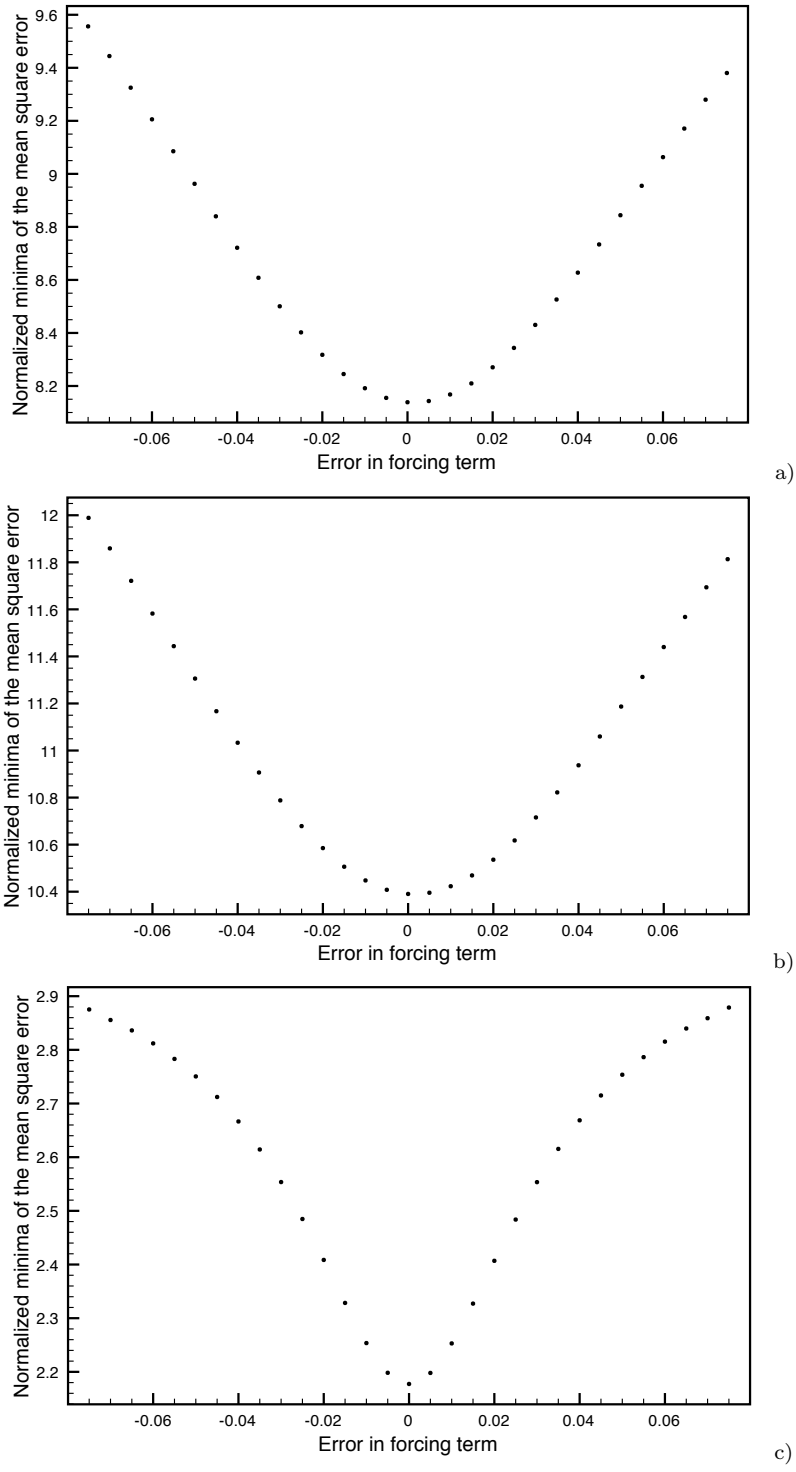



Figure 4.7: Normalised minima of the mean square error of the horizontal velocity (zonal and meridional) for different forcing errors, as in Fig. 4.6, in the following cases: (a) systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions; (b) random, uniformly distributed initial condition errors as in Fig 4.1; (c) both systematic and random initial condition errors are considered. The normalisation factor is $1 \times 10^{-5}$.

error in the forcing term increases, the time at which the mean square error of the horizontal velocity is minimum decreases, while the actual value of that minimum increases. The adding of forcing error, be it positive or negative, contributes thus to the shift of that minimum to higher values, while anticipating it in time.

It is also worth noting that, for forcing term errors $|\delta S| < 0.02$, the time at which the minimum in the mean square error occurs is higher when considering systematic errors in the initial conditions than when they are combined with random initial condition errors and than when those systematic errors are not present at all. The converse happens when $|\delta S| > 0.02$. Globally, the time at which the minimum occurs decreases with increasing $|\delta S|$ at a faster rate when only random initial condition errors are present than when only systematic initial condition errors are present. When both are present, the rate remains roughly the same as that of the case in which only systematic error is present, given its relative dominance.

As for the value of the minimum itself, it increases with $|\delta S|$, though with different rates depending on whether there is systematic error in the initial conditions. The presence of such systematic error leads to an increased rate for small $|\delta S|$, then followed by a decreased rate for higher $|\delta S|$. The distribution of the minima for different $|\delta S|$ thus becomes "spikier" when systematic initial condition errors are present.

By adding random initial condition errors to the case in which systematic errors are present the time at which the minimum of the mean square error occurs is slightly later and the minimum has a higher value. The qualitative shape of the plot is nevertheless similar. The most significant difference occurs depending on whether systematic error is present or not given a certain random initial condition error.

A word must be given on the slight asymmetry that can be observed in the plots when systematic initial condition errors are present, while noting the fact that such asymmetry is not present otherwise. A positive bias in the initial condition shifts the minimum of the mean square error to values that are slightly larger in the case of negative $\delta S$ than in the case of positive $\delta S$, while also anticipating the minimum slightly *less* when $\delta S < 0$ than when $\delta S > 0$, particularly for larger values of $|\delta S|$.

Negative biases in the initial conditions have the symmetric effect, i.e. yield the same results for opposite signs of $\delta S$ (not shown). Stronger biases, either positive or negative, enhance their respective behaviours (not shown).

### 4.5.2 Perturbation in the initial condition and in the diffusive time scale

In this subsection, aside from initial condition errors, we consider errors in the diffusive time scale $\tau_{DIF}$.

Fig. 4.8 portrays the time at which the mean square error of the horizontal velocity (zonal and meridional) is minimum, for different values of the error in the diffusive term $\tau_{DIF}$ (days). The triangles refer to the case in which there is a systematic error of $1 \times 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions of the horizontal velocity. The squares refer to the case in which the initial condition errors are random, uniformly distributed as in Fig 4.1. The circles refer to the case in which both systematic and random initial condition errors are considered. The number of realisations and the metric used for averaging are as in Fig 4.6.

Fig. 4.9 shows the normalised minima of the mean square error of the horizontal velocity (zonal and meridional) for different errors in the diffusive term $\tau_{DIF}$, as in Fig. 4.8, in the following cases: (a) systematic error of $1 \times 10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ in the

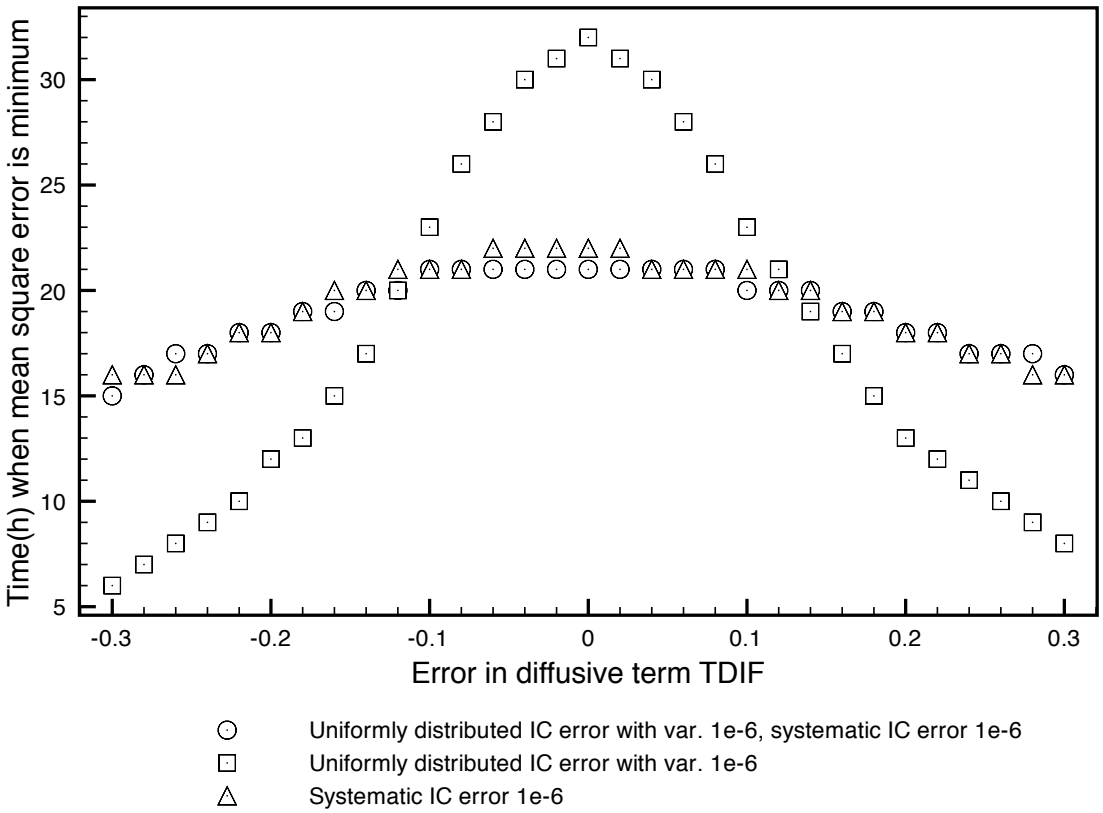



Figure 4.8: Time at which the mean square error of the horizontal velocity (zonal and meridional) is minimum, for different values of the error in the diffusive term $\tau_{DIF}$ in a Quasi-Geostrophic model [Eq. (4.1)]. The triangles refer to the case in which there is a systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions. The squares refer to the case in which the initial condition errors are random, uniformly distributed as in Fig 4.1. The circles refer to the case in which both systematic and random initial condition errors are considered. The number of realisations and the metric used for averaging are as in Fig 4.1.

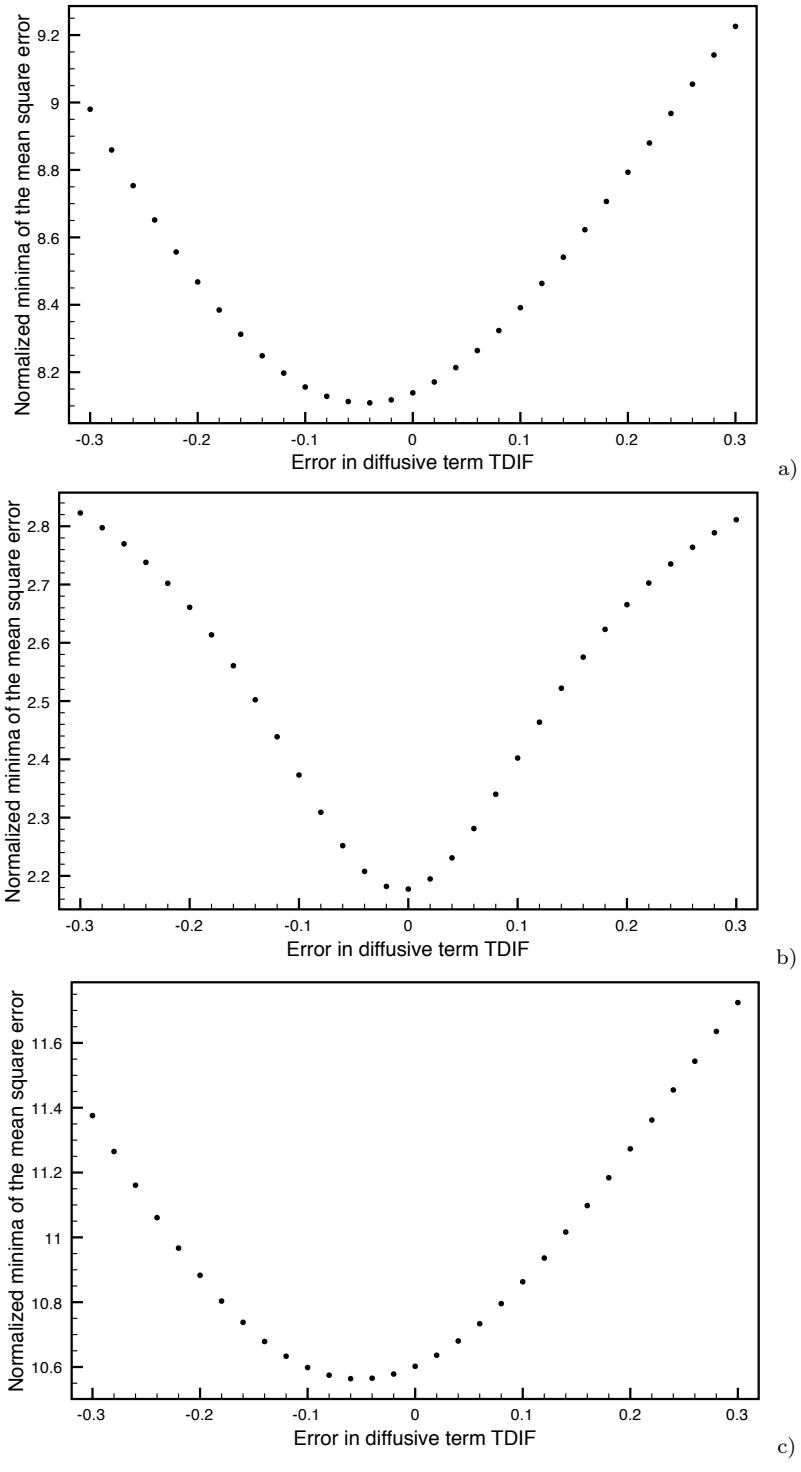



Figure 4.9: Normalised minima of the mean square error of the horizontal velocity (zonal and meridional) for different errors in the diffusive term $\tau_{DIF}$ (days), as in Fig. 4.8, in the following cases: (a) systematic error of $1 \times 10^{-6} \mathrm{m}^2\,\mathrm{s}^{-2}$ in the initial conditions; (b) random, uniformly distributed initial condition errors as in Fig 4.1; (c) both systematic and random initial condition errors are considered. The normalisation factor is $1 \times 10^{-5}$.

initial conditions; (b) random, uniformly distributed initial condition errors as in Fig 4.1; (c) both systematic and random initial condition errors are considered. The normalisation factor is $1 \times 10^{-5}$.

By looking at these figures one can identify a behaviour rather similar to that already seen in the forcing error case, namely the shift towards higher and anticipated values of the minima in the mean quadratic error when parameter errors are added. Still as in the forcing error case, the distribution of minima across the spectrum of $\delta\tau_{DIF}$ values when only random initial condition errors are present is "spikier" than when systematic initial errors are present. The value of $\delta\tau_{DIF}$ below which the minima in the random initial condition error case occur later than those in the biased error case is approximately 0.12. Above that value the situation is reversed.

A slight asymmetry is seen between the negative and positive $\delta\tau_{DIF}$ sides of the plots, even in the unbiased initial condition error case, when only random initial errors are considered. That asymmetry is strongly enhanced with the addition of systematic initial condition error, along with a shift of the whole plot to the negative side of the $\delta\tau_{DIF}$ spectrum, more precisely about 0.04 days to the right. The opposite behaviour occurs when the bias in the initial condition error is negative (not shown).

## 4.6 Probabilistic approach

Up to now, the behaviour of the square error has been considered on its mean. However, it is also of interest to study further elements of the probability distribution of the square error norm over the sphere and the model levels, namely its structure and transient evolution. For this purpose, statistical distributions of the square error in horizontal velocity (zonal and meridional) are determined

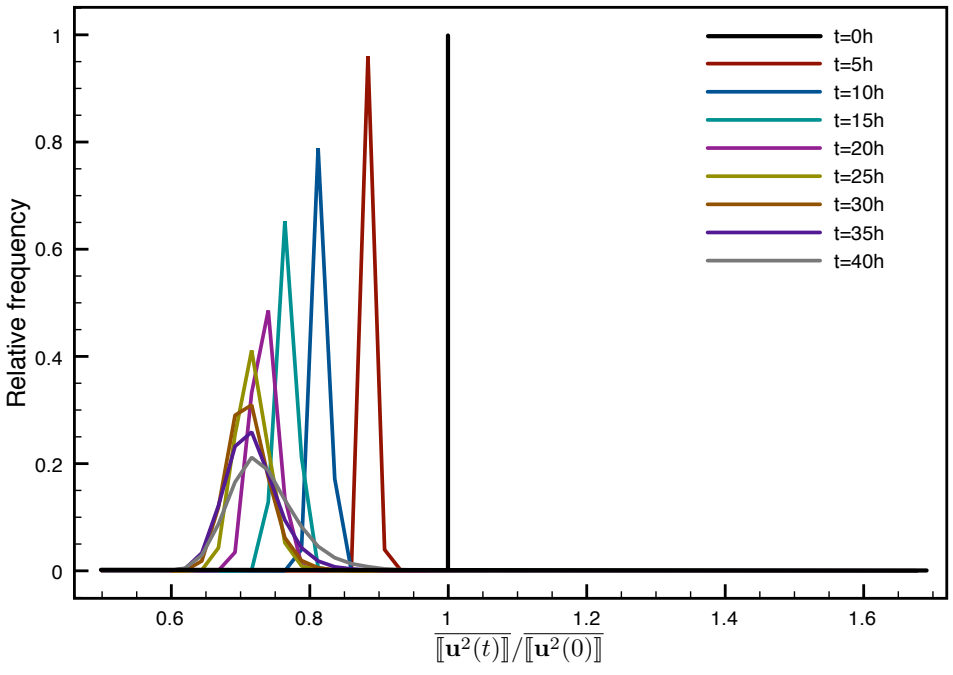
in the absence and in the presence of model error (Fig. 4.10). In both cases the initial errors are sampled from an unbiased uniform distribution with variance equal to $10^{-6}\mathrm{m}^2\,\mathrm{s}^{-2}$ in horizontal velocity (zonal and meridional).

The square error distributions are normalised with respect to the one at $t = 0$, in such a way that the distribution of the relative square error norm, $N_R = \overline{[\![\mathbf{u}^2(t)]\!]}/\overline{[\![\mathbf{u}^2(0)]\!]}$ is shown for each considered $t$: every 5 h from 0 to 40 h. That way, $N_R(0) = 1$. The total number of realisations is now 20000 (2000 initial conditions over 10 different positions on the attractor).

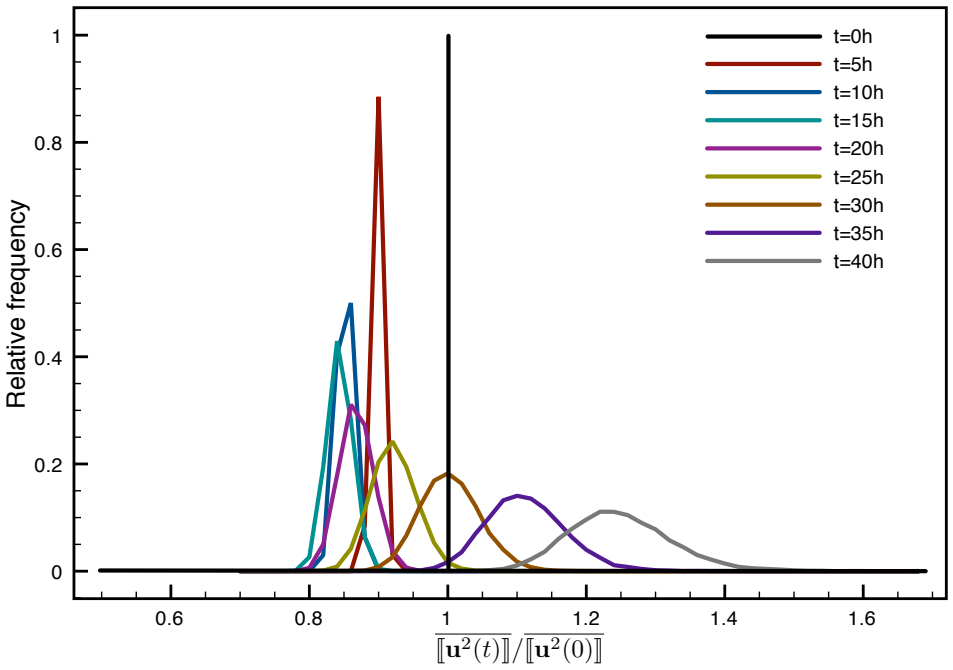
In both cases the mean and mode of the distributions exhibit, as expected, a decrease until a certain time of minimum $t_m$, naturally followed by an increase. In the case featuring systematic error in the parameters (Fig. 4.10b), $t_m$ is lower, i.e. the minimum occurs earlier with respect to the null parameter error case, Fig. 4.10a.

The behaviour seen in the mean of the distributions comes in agreement with previous considerations regarding the mean square error evolution (chapter 3 and previous sections of this chapter). In fact, by comparing, for each $t > 0$, the mean square errors with and without model error, one can recuperate such generic error features as the anticipation of the minimum and the relative increase in total error in the presence of both initial and model errors as compared to the case in which solely initial errors are present − which is consistent with e.g. Fig. 4.1.

As for the distributions themselves, their spread or standard deviation seems to increase for all computed $t$, as noted from visual inspection. This shows that there are errors growing much faster and others much slower than the average (or even decreasing). This is in accordance with the fact that the same initial square error variance may be projected either in the fastest modes leading to the largest square errors or in the most stable modes over which the square error decreases.




a)




b)

Figure 4.10: Short to intermediate time statistical distribution of the mean square error in horizontal velocity (zonal and meridional), normalised with respect to that of its initial distribution, in the absence (a) and in the presence (b) of model error, $|\delta S| = 0.04$. The number of realisations considered in the averaging over the distribution of initial errors and the attractor is 20000.

Moreover, as the effect of model error increases in time, so does the spread of the square error distribution as compared to the case in which no model errors are present, even as the normalised mean square error decreases initially. This spread increase is accompanied with the tendency to develop a modest tail in the direction of large error values.

The behavioural features hereby presented come in agreement with results seen in low order systems (see e.g. Fig. 3.8, chapter 3). As noted then, the behaviour observed in the presence of model error may reflect its role in delocalising the system in phase space.

## 4.7 Discussion and Conclusions

The experiments conducted in the present work have revealed the consistency of the proposed analytical formulation in assessing the error dynamics of a system with a greater degree of complexity, namely a spatially extended spectral model such as the Quasi-Geostrophic model, even in the presence of anisotropies in the distribution of initial errors (when considering breeding). It can thus be concluded that the dynamics of prediction errors in the short to intermediate time regimes can be reasonably well evaluated in an analytical way without the need to actually run the numerical model.

It has been seen that generic error features revelated here for the Quasi-Geostrophic model are consistent with those observed in low-order systems. Furthermore, the generalised formulation is seen to accurately reveal and depict additional features coming from the higher complexity of the Quasi-Geostrophic model relative to lower-order models.

It is interesting to note the economy in average computing time of the analytical approximation ($ct_A$) when compared with the average time needed to compute

the numerical (model minus nature) solution ($ct_N$). Extensive tests have led to a relation of $ct_A = 0.054\, ct_N$, thus indicating that the analytical approximation takes about 5 % of the time needed to compute the numerical solution.

For instance, tests performed in an ordinary PC with 1.6 GHz of CPU speed, 3 MB of RAM and 4 MB of cache yielded the following average computing times for the evolution of every 100 realisations of the square error leading to the pictures in this chapter: 60.2 s for model minus nature; 3.3 s for the analytical approximations (Taylor and Padé expansions); $\sim$ 64 s for the whole program combined (numerical and analytical solutions).

The computing speed aspect is rather relevant as more complex systems, used in operational settings, could benefit from such an economy in quickly assessing the error dynamics of the model in use without having to go through an extensive integration of the numerical model itself.

The analytical approximation used in the present chapter, while far more economical than the numerical integration approach (spares about 95 % of the computing time), is not as economical as the simpler, more restrictive method by Nicolis, Perdigão, and Vannitsem (2009). In fact, the former generates surrogates of the errors in the initial conditions along the assumed, eventually inexact, probability distribution of the errors, whereas the latter (Nicolis, Perdigão, and Vannitsem, 2009) does not require generation of surrogates as long as the covariance matrix of the initial condition errors is known, which is easy in particular if the errors have an isotropic distribution.

One key generalisation of our method with respect to that of Nicolis, Perdigão, and Vannitsem (2009) as is that ours includes the case in which the errors are state-dependent. However, if the covariance matrix of the error is known, a simpler yet less general extension of the method in Nicolis, Perdigão, and Vannitsem

(2009) can be applied in which additional terms are considered relating the attractor dynamics with the covariance structure of the errors.

That way, there will be terms consisting on means over the attractor of products $[\![F(x)\,C(x)]\!]$, with $F(x)$ representing the attractor dynamics and $C(x)$ that of the error covariance structure. This product is then decomposable as a product of means, $[\![C(x)]\!]\,[\![F(x)]\!]$, and a covariance term. This term can be decisive in the square error curve (as the case in subsection 4.4.2), leading to a completely different behaviour when comparing: a) the case in which the initial condition errors are indifferently projected onto stable or unstable manifolds (as when they are generated by Monte Carlo yielding a Gaussian or uniform error distribution); with b) the case in which those errors tend to be more projected onto certain directions (e.g. when breeding is considered, favouring the error projection onto the unstable directions).

By considering, for instance, the case in which errors would be state-dependent and projected solely onto the stable manifold, the square error would decrease towards zero by definition of stable manifold. This is a case in which the generic conclusions from Nicolis, Perdigão, and Vannitsem (2009) would no longer hold, as the error behaviour would be completely different. Another case in which the formulation derived by Nicolis, Perdigão, and Vannitsem (2009) would no longer hold can be considered by constructing, by variational methods (e.g. 4DVAR), perturbations in the initial conditions that produce error evolution curves that are completely atypical (such as having the square error grow at an early stage then to decrease).

In the case of errors obtained by data assimilation there are formulas of the covariance matrix of analysis errors. In particular for the 4DVAR case performed in a perfect model scenario, Pires et al. (1996) proposes that matrix. In general,

that covariance matrix is given by $C = H_J^{-1}\,\sigma_{obs}^2$, where $H_J^{-1}$ is the inverse of the Hessian of 4DVAR's cost function. That formula could be applied in the analytical expressions for the error derived in the thesis and from there an evaluation of the state-dependent scenario could be performed.

All in all, the analytical formulation provides an efficient, relatively quick way of analysing the error evolution without the need for extensive model runs, and is valid in situations with key aspects of realistic relevance for numerical modelling such as having state-dependent errors.

The results are globally encouraging, inspiring further developments in both theory and applications, as the discussion held in the previous paragraphs suggests. A long way is thus ahead along this line of research.

# Chapter 5

# Beyond short-term – On a simple error model

## 5.1 Introduction

The formulation presented in previous chapters has succeeded in tackling the error behaviour in the small perturbation, short-term scenario. It has done so in a general, systematic and fundamental way by taking into account both initial condition and model-related errors and connecting them to intrinsic properties of the underlying system.

Our work would not be complete without an attempt to explore the long-term, asymptotic error behaviour.

Given the limitations of a general, model-independent formulation, we hereby present a descriptive analytical model not aiming at universality, but rather at a particular case study. In this chapter, that case will be the Quasi-Geostrophic model from Marshall and Molteni (1993).

We begin by considering the short-term model proposed by Teixeira et al. (2007), where a simple model of error growth due to truncation (time-step) is constructed by considering the relative error growth rates in the stable and unstable directions.

Our purpose is to generalise that model in such a way as to be able to contemplate the longer-term error evolution until saturation.

For that purpose, a simple model of error evolution is introduced (section 5.2), by considering the relative error growth rates in the stable and unstable directions, with the addition, in the latter case, of a nonlinear saturation term. The application of the proposed formulation to a Quasi-Geostrophic model is then performed and analysed (section 5.3), followed by discussion and conclusions (section 5.4).

As in preceding chapters, the error evolution model will be compared to the "model minus nature" numerical simulations. The difference is that in the present chapter the analytical model focuses on the time step sensitivity.

## 5.2 Error growth model

A simple model of error growth can be constructed by considering the relative error growth rates in the stable ($\epsilon_s$) and unstable ($\epsilon_u$) directions (Teixeira et al., 2007):

$$\frac{d\epsilon_s}{dt} = \alpha\epsilon_{\Delta t} - \sigma_d\epsilon_s + a\epsilon_u \tag{5.1a}$$

$$\frac{d\epsilon_u}{dt} = \beta\epsilon_{\Delta t} + \lambda_{\mathrm{I}}\epsilon_u + b\epsilon_s \tag{5.1b}$$

Here, $\epsilon_s$ is the norm of the error in the direction of the stable modes, $\epsilon_u$ is the norm of the error in the direction of the unstable modes, $\sigma_d$ is the rate of decay of stable modes, $\lambda_{\mathrm{I}}$ is the sensitivity to the initial conditions, $\epsilon_{\Delta t}$ is the typical one-step integration (truncation) error given the time-step $\Delta t$, $\alpha$ and $\beta$ are the relative projections of the one-step integration error onto the stable and unstable directions, respectively, and $a$ and $b$ are coupling constants.

As noted in Teixeira et al. (2007), the principle behind this approach derives from the generalised Hartman-Grobman theorem (Pugh, 1969; Pugh, 1987), which implies that there is a nonlinear change of coordinates so that the dynamics of the nonlinear system is equivalent to that of a linear system with the same positive and negative eigenvalues as those of the linearised system. This essentially means that the dynamics of the tangent linear model is qualitatively the same as that of the nonlinear model in the short-term.

As seen and discussed in Teixeira et al. (2007), this error growth model successfully represents the short-term error behaviour of the QG model.

In order to represent not only the initial short-term behaviour but also the longer-term one leading to nonlinear saturation, the following model is hereby proposed:

$$\frac{d\epsilon_s}{dt} = -\sigma_d \epsilon_s + \alpha \epsilon_{\Delta t} \tag{5.2a}$$

$$\frac{d\epsilon_u}{dt} = \lambda_{\mathrm{I}} \epsilon_u + \beta \epsilon_{\Delta t} - \frac{\eta}{\lambda_{\mathrm{I}}^2} \epsilon_u^2 \,, \tag{5.2b}$$

where $\eta$ is a saturation coefficient. Here, the coupling between stable and unstable modes is assumed to be relatively weak, i.e. $|a\, b| < |\sigma_d \lambda_{\mathrm{I}}|$, therefore the coupling terms in (5.1) are now neglected. The adequateness of this assumption was tests in preliminary experiments and is patent in the results from the experiments below.

The physically relevant solutions of (5.2a) and (5.2b) are analytically derived to be:

$$\epsilon_s(t) = \left( \epsilon_{s0} - \frac{\alpha\, \epsilon_{\Delta t}}{\sigma_d} \right) \exp(-\sigma_d t) + \frac{\alpha\, \epsilon_{\Delta t}}{\sigma_d} \tag{5.3a}$$

$$\epsilon_u(t) = \frac{\lambda_{\mathrm{I}}^2}{\eta} \cdot \frac{K \psi_- \exp\left[-t(\psi_+ - \psi_-)\right] + \psi_+}{K \exp\left[-t(\psi_+ - \psi_-)\right] + 1} \,, \tag{5.3b}$$

where

$$\psi_{\pm} = \frac{\lambda_{\mathrm{I}}}{2}\left(1 \pm \sqrt{1 + 4\beta\epsilon_{\Delta t}\frac{\eta}{\lambda_{\mathrm{I}}^2}}\right) \tag{5.4}$$

correspond to the two steady-state solutions, and

$$K = \frac{\psi_+ + \epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}^2}}{\epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}^2} - \psi_-} \tag{5.5}$$

is the integrating constant of (5.3b) for the initial value condition $\epsilon_u(0) = \epsilon_{u0}$, valid for the set of possible parameter values in the present context, i.e. those taken positive.

The governing equation for the error norm in the direction of the unstable modes [Eq. (5.2b)] has the form of a logistic equation with harvesting, like the error growth equation proposed by Savijarvi (1995):

$$\frac{de}{dt} = (a\,e + s)\left(1 - \frac{e}{e_i}\right) \tag{5.6}$$

where $e$ is the rms (root mean square) error, $a$ is the growth rate of the errors due to initial conditions (corresponding to $\lambda_I$ in our model), $e_i$ is a saturation value corresponding to the rms difference between randomly picked flow states and $s$ is a model error term. Eq. (5.6) is in turn based on previous works from Lorenz (1982), who had introduced that quadratic law without $s$; and Dalcher and Kalnay (1987), who adapted Lorenz's (1982) law for the the error variance and, inspired by Leith (1982), introduced the model error term $s$. In both Eqs. (5.6) and (5.2b) the "harvesting" term corresponds to the model error forcing term, which in our case accounts for the time-step sensitivity.

The novelty of our error growth model with respect to Savijarvi (1995) and predecessors lies on the inclusion of the prognostic equation for the error in the

stable modes [Eq. (5.2a) and Teixeira et al. (2007)], without which the short-to-intermediate term behaviour would likely not be properly accounted for, as we shall see in the experiments below.

It is worth taking a closer look at the approach to the error in the unstable directions, for which purpose we focus on what Eq. (5.3b) can tell us about the error behaviour:

Within the domain of differentiability of $\epsilon_u(t)$, its first two time derivatives are given by:

$$\epsilon'_u \equiv \frac{d\epsilon_u(t)}{dt} = \frac{\lambda_{\mathrm{I}}^2}{\eta} \cdot \frac{K(\psi_+ - \psi_-)^2 \exp\left[-t(\psi_+ - \psi_-)\right]}{\left\{K \exp\left[-t(\psi_+ - \psi_-)\right] + 1\right\}^2} \tag{5.7}$$

$$\begin{aligned} \epsilon''_u(t) \equiv \frac{d^2\epsilon_u(t)}{dt^2} =& \frac{\lambda_{\mathrm{I}}^2}{\eta} \cdot \frac{K(\psi_+ - \psi_-)^3 \exp\left[-t(\psi_+ - \psi_-)\right]}{\left\{K \exp\left[-t(\psi_+ - \psi_-)\right] + 1\right\}^3} \\ & \cdot \left\{K \exp\left[-t(\psi_+ - \psi_-)\right] - 1\right\}. \end{aligned} \tag{5.8}$$

Within the domain of differentiability of $\log[\epsilon_u(t)]$, its first two time derivatives are given, as functions of $\epsilon_u(t)$ and its time-derivatives, by:

$$\frac{d}{dt}[\log \epsilon_u(t)] = \frac{\epsilon'_u(t)}{\epsilon_u(t)} \tag{5.9}$$

$$\frac{d^2}{dt^2}[\log \epsilon_u(t)] = \frac{\epsilon''_u(t)\epsilon_u(t) - [\epsilon'_u(t)]^2}{[\epsilon_u(t)]^2}. \tag{5.10}$$

For our class of parameters the first time derivative of both $\epsilon_u(t)$ and $\log[\epsilon_u(t)]$ is always non-negative and tends towards zero as $t$ tends towards infinity. Therefore, one can expect the error in the unstable directions to grow until reaching an asymptotic, saturation level.

As far as the second time derivative of $\epsilon_u(t)$ is concerned,

$$\begin{aligned} \epsilon''_u(t) = 0 \Leftrightarrow t &= \frac{\log\left[\left(\psi_+ - \epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}^2}\right)\left(\psi_- - \epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}^2}\right)\right]}{\psi_+ - \psi_-} \\ \Leftrightarrow t &= \frac{\log\left(\beta\epsilon_{\Delta t}\eta - \epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}} + \epsilon_{u0}^2\frac{\eta^2}{\lambda_{\mathrm{I}}^4}\right)}{\lambda_{\mathrm{I}}\sqrt{1 + 4\beta\epsilon_{\Delta t}\frac{\eta}{\lambda_{\mathrm{I}}^2}}}. \end{aligned} \tag{5.11}$$

For $\epsilon_u''(t)$ to have zeros,

$$\begin{aligned} &\beta\epsilon_{\Delta t}\eta - \epsilon_{u0}\frac{\eta}{\lambda_{\mathrm{I}}} + \epsilon_{u0}^2\frac{\eta^2}{\lambda_{\mathrm{I}}^4} > 0 \\ \Leftrightarrow &\left(\epsilon_{u0} > \frac{\lambda_{\mathrm{I}}^2}{\eta}\psi_+\right) \vee \left(\epsilon_{u0} < \frac{\lambda_{\mathrm{I}}^2}{\eta}\psi_-\right) . \end{aligned} \tag{5.12}$$

The solutions in (5.12) not corresponding to extrema in the error evolution curve will thus be inflection points.

It can be verified straightforwardly that, in particular for $\epsilon_{u0} = 0$, $\epsilon_u''(t)$ could have no zeros for our class of parameters. In fact, since $\beta\epsilon_{\Delta t}\eta > 0$, we would have the log of a negative quantity for $\epsilon_{u0} = 0$. Therefore, in this particular case, $\epsilon_u''(t)$ is always negative. This in turn means that the growth rate of the error in the unstable directions will always decrease.

As for the second derivative of $\log[\epsilon_u(t)]$, it is always negative for our class of parameters. In fact,

$$\frac{d^2}{dt^2}[\log \epsilon_u(t)] < 0 \Leftrightarrow K \exp\left[-t(\psi_+ - \psi_-)\right] > -1 , \tag{5.13}$$

which is a universal condition, since both $K$ and the exponential are positive. The growth rate of the logarithm of the error in the unstable directions will thus decrease as well.

Let us briefly analyse the behaviour of $\epsilon_u$ relative to the truncation error. For our class of parameters, $\frac{\partial \epsilon_u}{\partial \epsilon_{\Delta t}} > 0$ and $\frac{d}{dt}\frac{\partial \epsilon_u}{\partial \epsilon_{\Delta t}} < 0$ until saturation, where both conditions tend to zero. The same holds for the logarithmic derivatives. Therefore, for every $t$ before saturation, $\epsilon_u$ grows with the truncation error. Still, the ratio between two generic $\epsilon_{uA}$ and $\epsilon_{uB}$ (or their logarithms), respectively with truncation errors $\epsilon_{DA}$ and $\epsilon_{DB}$ where $\epsilon_{DA} > \epsilon_{DB}$, monotonously decreases until reaching zero.

These considerations are confronted with case study results in the following section.

## 5.3 A Case study: the Quasi-Geostrophic model

In the present section a case study is considered that will allow for the illustration of error growth features brought up in the analysis of the error model presented in the previous section.

The time evolution of the base 10 logarithm of the $\mathcal{L}^2$ norm error in the non-dimensional Streamfunction state vector of the QG model based on the difference between results with time steps of 180, 80, 40, 20, 10, 5 and 2.5 minutes and those obtained with the control time-step of 1.25 minutes, is shown in Fig. 5.1 [based on Teixeira et al. (2007)]. The nondimensional Streamfunction is obtained by normalisation of the Streamfunction with respect the square of the radius of the Earth ($6.371 \times 10^6$ m) times twice the rotation rate of the Earth ($7.292 \times 10^{-5}$ s$^{-1}$). No initial condition errors are taken in those cases. The figure also shows two examples of errors caused by initial conditions, in order to allow for a comparison between truncation error and initial-condition error, and a line showing growth proportional to time $t$.

In order to evaluate the applicability of our proposed model of error growth to the referred QG model case, an attempt to reproduce the outcome represented in Fig. 5.1 is performed by using our proposed model of error growth and presented in Fig. 5.2. The chosen parameters are $\lambda_{\mathrm{I}} = 0.5$, $\eta = 4$, $\alpha = 0.2$, $\beta = 0.01$, $\sigma_d = 0.01$, $\epsilon_{\Delta t} = 10^{-12}(\Delta t)^4$, $\epsilon_{u0} = 0$, $\epsilon_{s0} = 0$, where $\Delta t$, the time steps considered (in minutes), are the same as those in Fig. 5.1.

The parameters were chosen in such a way as to allow for an accurate qualitative reproduction of the results shown in Fig. 5.1, following the procedure from Teixeira et al. (2007): $\lambda_{\mathrm{I}}$ can be estimated from the exponentially growing part of the solution and corresponds approximately to the sustained error growth rates

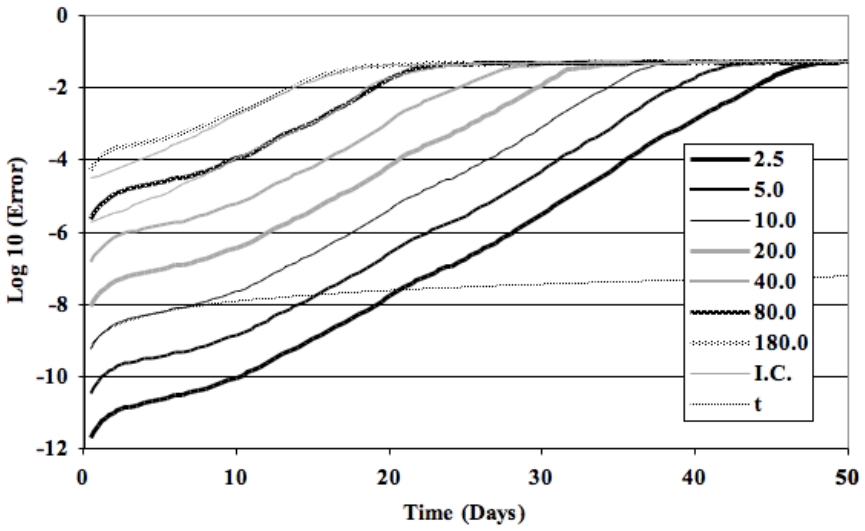



Figure 5.1: Time evolution of the base 10 logarithm of the $\mathcal{L}^2$ norm error in the non-dimensional Streamfunction state vector of the QG model based on the difference between numerical results with time steps $\Delta t$ of 180, 80, 40, 20, 10, 5 and 2.5 minutes and those obtained with the control time-step of 1.25 minutes. Also shown are errors corresponding to two simulations with perturbed initial conditions of different magnitudes that use the control time step ("I.C.", thin solid gray lines). Based on Teixeira et al. (2007).

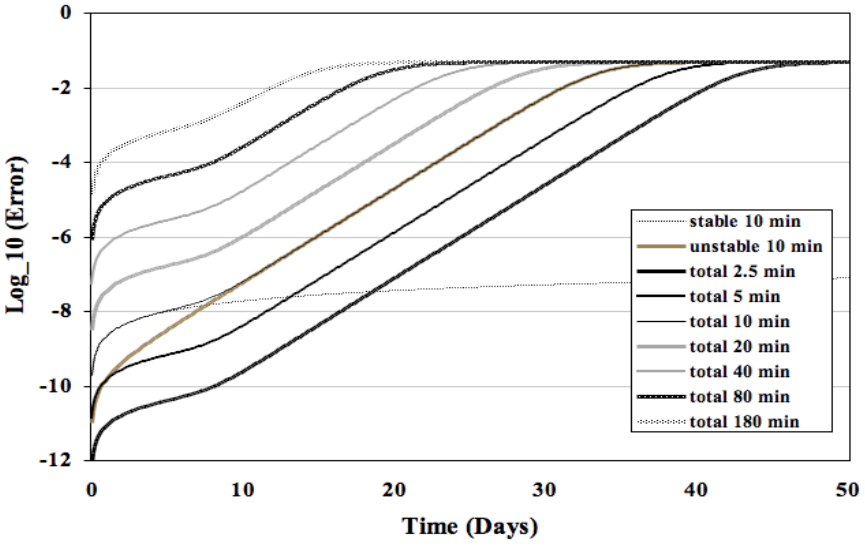



Figure 5.2: Representation of Fig. 5.1 using our proposed error growth model. The chosen parameters are $\lambda_{\mathrm{I}} = 0.5$, $\eta = 4$, $\alpha = 0.2$, $\beta = 0.01$, $\sigma_d = 0.01$, $\epsilon_{\Delta t} = 10^{-12}(\Delta t)^4$, $\epsilon_{u0} = 0$, $\epsilon_{s0} = 0$, where $\Delta t$, the time steps considered, are the same as those in Fig. 5.1.

in the QG model; $\alpha$ can be estimated from the linear short-time solution of the stable error; $\sigma_d$ can be obtained, as a function of $\alpha$, from the long-time stable error solution; and $\beta$ can be obtained from the equation $\epsilon_u(t) = \epsilon_s(t)$, which leads (assuming a linear approximation for the stable error) to $\beta = f(\alpha, \lambda_{\mathrm{I}}, \tau)$, where $\tau$ is the approximate time when the plateau period is over. The quartic dependence of $\epsilon_{\Delta t}$ on $t$ comes from the QG integration scheme being of fourth order. As for the new saturation term, $\eta$, its value is chosen so that the saturation level of the simulated test (Fig. 5.2) matches that of the numerical experiment (Fig. 5.1).

By having chosen $\alpha$ to be considerably larger than $\beta$ we have ensured the observed occurrence of the plateau in the growth rate between the rapid initial error growth (as the solution moves onto the attractor) and the stage at which the unstable error growth becomes dominant. Had $\beta$ been dominant, the unstable component would have dominated with no inflection whatsoever. The plateau does, in fact, appear only when stable modes are considered and relevant enough in the intermediate times. Error models based solely on logistic equations are unable to represent that kind of intermediate term behaviour.

By comparing Figs. 5.1 and 5.2, a clear agreement between them is immediately noticeable, not only on the short-term but rather all the way up to saturation. In fact, both reveal that, in general, the initial condition error grows exponentially as expected, whereas the truncation error exhibits an initially rapid, super-exponential growth during the first 2 days, followed by a period (between about days 2 and 10) of relatively slow growth. This in turn is followed by exponential growth until reaching saturation.

It is worth noticing that, at the saturation level, differences between asymptotic values of the errors for different $\epsilon_{\Delta t}$ are barely noticeable, in such a way that the corresponding curves apparently overlap. This can be understood by noting

that: *a)* $4\beta\epsilon_{\Delta t}\eta/\lambda_{\mathrm{I}}^2 \ll 1$, which leads to $\epsilon_u$ tending towards approximately $\lambda_{\mathrm{I}}^3/\eta$; *b)* $\epsilon_{\Delta t}\alpha \ll \sigma_d$ in $\epsilon_s$, which makes its square negligible when added to that of $\epsilon_u$ as $t \to \infty$. The resulting significant terms in the $\mathcal{L}^2$ norm (and its logarithm by that matter) are thus practically independent from $\epsilon_{\Delta t}$.

As it would be expected for the sake of consistency, although a saturation term has been added, the short-term error evolution modelled with the unstable modes determined by (5.2b) is similar to that of the model proposed by Teixeira et al. (2007). This can be clearly understood by noting that (5.2b) tends to its counterpart in Teixeira et al. (2007) for short-term time scales.

A comparison between the numerical solution and both our proposed error growth model [Eqs. (5.3a, 5.3b)] and that of Savijarvi (1995) is portrayed in Fig. 5.3. The circles represent the time evolution of the base 10 logarithm of the $\mathcal{L}^2$ norm error in the non-dimensional Streamfunction state vector of the QG model based on the difference between numerical results with time step $\Delta t$ of 10 minutes and those obtained with the control time-step of 1.25 minutes. The solid line depicts the representation of the numerical solution using our proposed error growth model [Eqs. (5.3a, 5.3b)], with parameters as in Fig. (5.2). The crosses depict the best fit of the Savijarvi (1995) model [Eq. (5.6)] with respect to the numerical solution (crosses). The parameters corresponding to the best fit are $a = 0.58$, $e_i = 0.055$ and $s = 2 \times 10^{-11}$, and the variance of residuals (reduced $\mathcal{X}^2$) is $3.67 \times 10^{-6}$. In all three cases the error in the initial conditions is set to zero, but the initial value is not represented graphically, as all three logarithmic curves would start at $-\infty$.

By observing Fig. 5.3 a significant overlap is seen for most of the time when comparing our proposed error model [Eqs. (5.3a, 5.3b)] and the numerical solution from the early stages of error evolution until saturation, as had been already

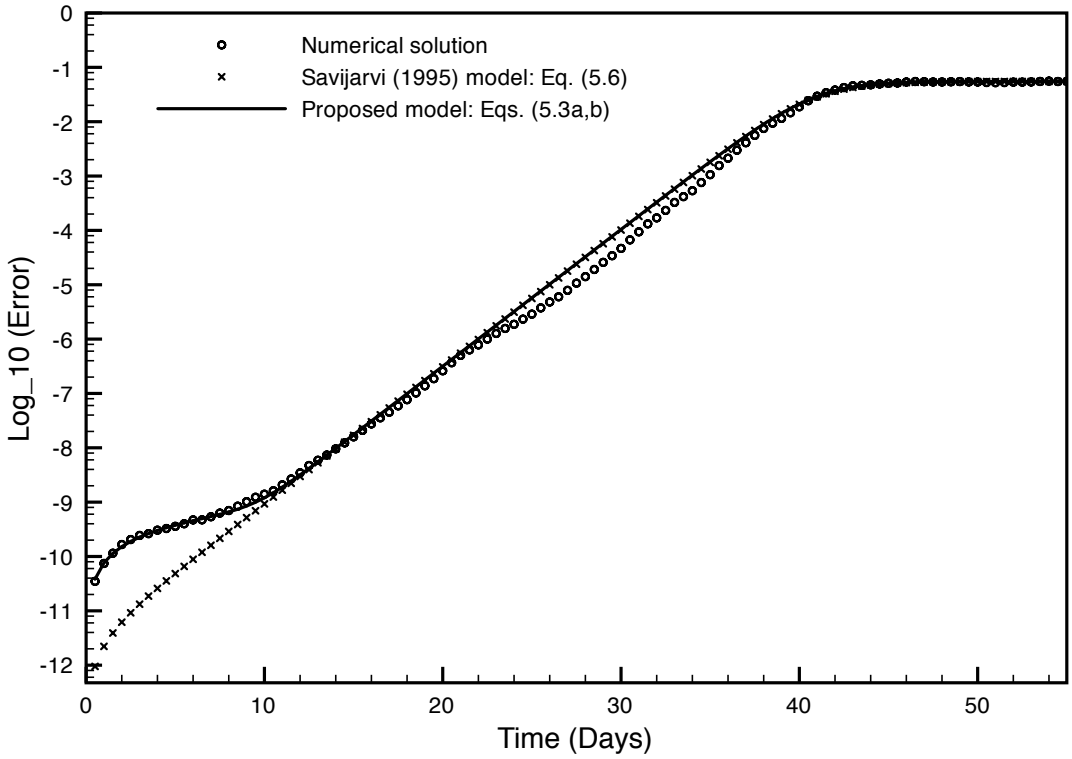



Figure 5.3: Time evolution of the base 10 logarithm of the $\mathcal{L}^2$ norm error in the non-dimensional Streamfunction state vector of the QG model based on the difference between numerical results with time step $\Delta t$ of 10 minutes and those obtained with the control time-step of 1.25 minutes (numerical solution, represented in circles). Representation of the numerical solution using our proposed error growth model [Eqs. (5.3a), (5.3b)] (solid line), with parameters as in Fig. 5.2. Best fit of the Savijarvi (1995) model [Eq. (5.6)] with respect to the numerical solution (crosses).

suggested by inspecting Figs. 5.1 and 5.2.

However, that does not happen when considering the Savijarvi (1995) model [Eq. (5.6)]. In fact, it only overlaps the numerical solution in the longer term when the unstable modes are dominant, while failing to capture the short-term behaviour of the error, where stable modes are dominant. The error model proposed by Savijarvi (1995) captures the error in the unstable directions, as does Eq. (5.2b), to which it is formally equivalent. For an overall assessment on the error evolution when both stable and unstable modes are relevant, Fig. 5.3 shows that our proposed error model is a good candidate for representing the numerical solution.

A complementary evaluation can be performed by comparing the referred actual and simulated QG cases (numerical solution vs. our proposed model) as far as the ratios of the logarithms of $\epsilon_u(t)$ with different time steps A and B are concerned, in the following form:

$$\text{Ratio}(t) = \frac{\log_{10}(\text{Error}_B(t))}{\log_{10}(\text{Error}_A(t))}\,, \tag{5.14}$$

where Error $\equiv \epsilon_u(t)$.

The corresponding results for the actual and simulated cases are plotted respectively in Figs. 5.4 and 5.5, for the following time step ratios (B vs. A), where B = 2 A: 80 min vs. 40 min, 40 min vs. 20 min, 20 min vs 10 min, 10 min vs. 5 min and 5 min vs. 2.5 min.

An agreement is qualitatively apparent between the ratios as shown in Figs. 5.4 and 5.5, despite minor numerical fluctuations in Fig. 5.4. This further signals the applicability of the proposed error model.

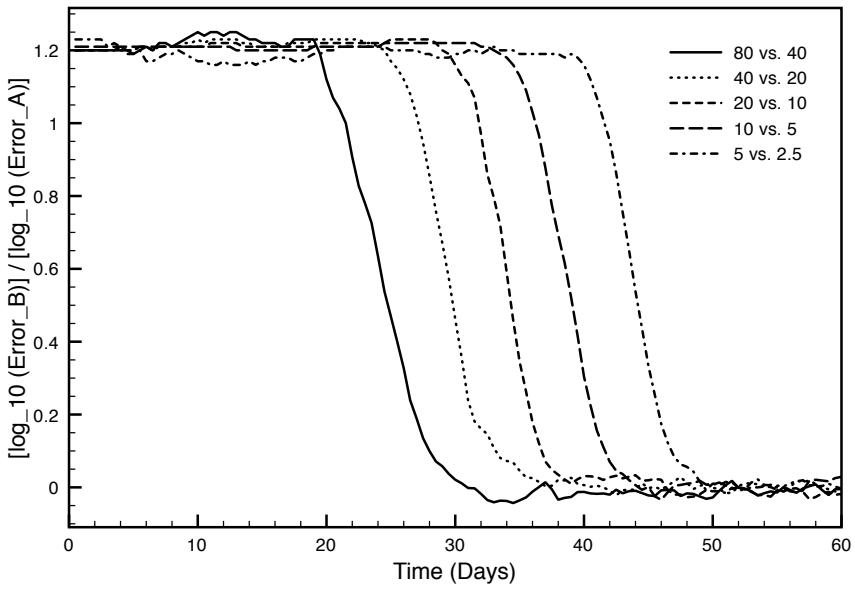



Figure 5.4: QG actual model run case as in Fig. 5.1: Time-evolution of the ratios between the logarithms of the simulated error in the unstable directions, for the following time steps (B vs. A): 80 min vs. 40 min, 40 min vs. 20 min, 20 min vs. 10 min, 10 min vs. 5 min, 5 min vs. 2.5 min.

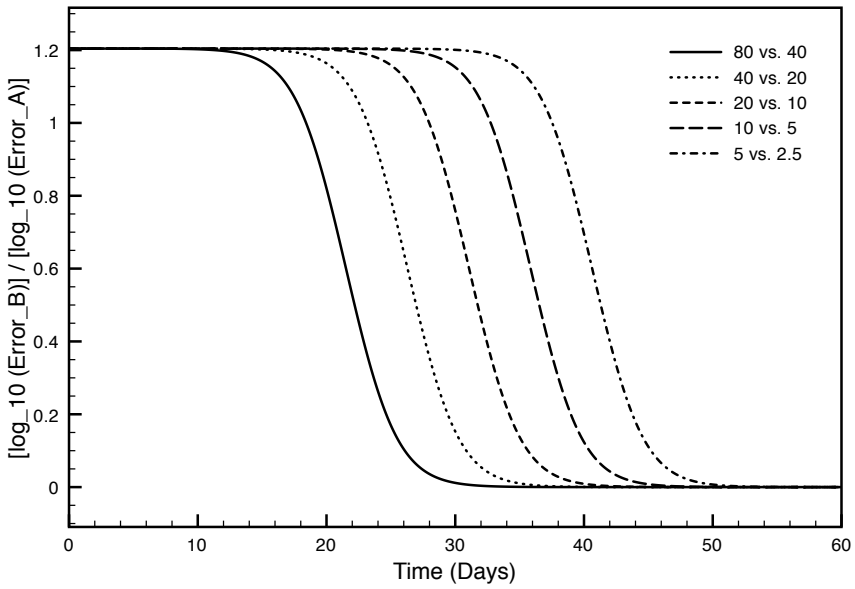



Figure 5.5: QG case as simulated with the proposed error growth model as in Fig. 5.2: Time-evolution of the ratios between the logarithms of the simulated error in the unstable directions, for the following time steps (B vs. A): 80 min vs. 40 min, 40 min vs. 20 min, 20 min vs. 10 min, 10 min vs. 5 min, 5 min vs. 2.5 min.

From inspection of Figs. 5.4 and 5.5, three stages are distinctive in the ratio evolution. The first corresponds to a constant ratio of $\sim 1.2$ and holds as long as neither A nor B cases are saturating. The second stage corresponds to the time interval at which the error in case B is already saturating whereas A is still not doing so, which leads to a decreasing ratio. Finally, the third corresponds to the case in which the error in both A and B cases have saturated to practically the same asymptotic value, hence their null ratio.

Given that the time-step ratio between B and A has been taken as 2 for all cases, the ratio evolution curves are parallel for all cases. The higher A and B are, the sooner saturation occurs, therefore the sooner the second and third regimes of ratio evolution occur.

The behaviour exhibited by the referred ratio is consistent with the error behaviour seen in Figs. 5.1, 5.2 and noted in the analytical inspection (section 5.2).

## 5.4 Discussion and Conclusions

A clear resemblance between the output from numerical (model minus nature) and simulated (our model) tests can been seen by comparative analysis. Furthermore, the ratios between realisations of the unstable modes for different time steps agree when the actual error behaviour from QG model runs and simulated error behaviour from the error growth model are compared.

This suggests that the proposed error model is adequate to the representation of the considered QG model case, not only in the short-term but also in the longer term until saturation. In fact, the results successfully reveal the initial rapid super-exponential growth, followed by an interval of slower growth rate

("plateau"), which in turn is followed by exponential growth until saturation is achieved.

With respect to previous efforts, our model has thus joined the ability of Teixeira et al. (2007) to describe the short-to-intermediate time error behaviour, notably the intermediate-time "plateau" unrepresentable by the approach in Savijarvi (1995) (and predecessors) alone, with that of the latter to describe the error behaviour until saturation unrepresentable by the formulation in the former.

Unlike the approach to the error dynamics from the previous chapters, the error model in the current chapter allows for longer-term error behaviour to be assessed in the QG model. Nevertheless, while these results are encouraging, one must note that they are not necessarily model-independent as the general, systematic approach derived in chapter 3 and applied there and in chapter 4. The possible developments stemming from this could probably pave the way for yet another thesis.

# Chapter 6

# Conclusions

The preceding chapters have featured their own conclusions, therefore in this concluding chapter a summary of the overall conclusions is presented.

The core purpose of this thesis was to address pertinent challenges underneath the estimation of the state of the system in two main contexts: on one hand, that of atmospheric downscaling; on the other hand, that of atmospheric predictability.

For that purpose, novel methods in nonlinear statistics and dynamics have been studied, developed and implemented in the aforementioned contexts.

As far as atmospheric downscaling is concerned, nonlinear statistical features have been assessed within the statistical response of the monthly winter (December to February) precipitation to the North-Atlantic Oscillation over the North Atlantic European Region. For that purpose, two major methodologies have been developed and implemented.

On one hand, a diagnostic measure has been built in order to measure the asymmetric part of an estimated variable's response to its predictor, a measure undetected by linear correlation. As a practical application, that variable was chosen to be the precipitation and its predictor the NAO regime (NAO+ and

NAO-). The asymmetric features have then been used to define an asymmetry-based measure of non-Gaussianity.

On the other hand, an information-theoretical assessment on non-Gaussianity has also been performed and a corresponding measure of information-theoretical correlation − also transcending the limited scope of linear correlation − defined and applied to the aforementioned downscaling application.

As main results, the proposed estimators for asymmetry and non-Gaussianity have been proven to be consistent in their domain of validity. The statistical response of monthly precipitation to NAO has been seen to be asymmetric and non-Gaussian. New relevant features have been brought out as a result of the application of the proposed nonlinear statistical methods − more detailed conclusions are presented in the final section of chapter 2.

As far as atmospheric predictability is concerned, a systematic formalism has been derived for the dynamics of prediction errors under the combined influence of initial-condition and model-related errors. Its analytical results have been confronted with those from numerical experiments and some generic features for the error dynamics brought out and connected with intrinsic system properties.

Analytical and numerical results have been seen to agree within the domain of validity of the analytical formulation: that of small perturbations in the short-to-intermediate time regime.

The error evolution when only initial condition errors or parameter errors are present has been found to agree with well-known results, another sign of consistency of the method.

By adding parameter errors to systems with initial condition errors evolving to a short-term minimum in the mean square error, a displacement in that minimum

has been seen. In particular, it is anticipated and its depth weakened as compared to the case in which no parameter errors are present. This behaviour has been seen in low-order and higher-order (intermediate complexity) systems alike and can be understood bearing in mind that adding parameter errors increases the instability of the system and thus its predictive uncertainty.

As far as the whole distribution of the mean square error is concerned, the distributions for the case in which only initial errors are present were seen to be sharper, narrower and with a lower-valued support set than the distributions for the case in which both initial and model-related errors are present. It has been determined that, as the effect of model error increases in time, so does the broadness of the square error distribution as compared to the case in which no model errors are present, even as the mean square error decreases initially. The changes in the error distribution as model error is added have been seen to reflect its role in delocalising the system in phase space.

All in all, the proposed formulation for assessing the error dynamics has served its purposes: enabling an evaluation of the dynamics of prediction errors without the need to actually run the numerical model under analysis; revealing new generic model-independent features of the error dynamics, under the combined influence of initial condition and model-related errors, not only for low-order but also to systems exhibiting a higher order of complexity.

By having addressed several complex problems, this thesis was not meant at a grand closure in all fronts, but rather at providing an innovative contribution, however modest it may seem, in each of them. That has been accomplished, paving the way for further developments to be made in both statistical downscaling and dynamical predictability, using the tools, methodologies and results from the thesis.

# Appendix A

# Derivation of Eq. (2.11)

We begin by recalling Eq. (2.8):

$$c(X,Y) = c_M + \frac{c_+ \sigma_{X_+} \sigma_{Y_+}}{2} + \frac{c_- \sigma_{X_-} \sigma_{Y_-}}{2}, \tag{A.1}$$

where

$$c_M = E_{Y_+} E_{X_+} = E_{Y_-} E_{X_-} \,. \tag{A.2}$$

By considering the case in which $X$ and $Y$ are jointly Gaussian, the relations in Eq. (2.10) hold:

$$\begin{aligned} E_{X_+} &= -E_{X_-} = \sqrt{\alpha} \\ E_{Y_+} &= -E_{Y_-} = \mathrm{sgn}(c)\sqrt{\alpha} \\ \sigma^2_{X_+} &= \sigma^2_{X_-} = 1 - \alpha \\ \sigma^2_{Y_+} &= \sigma^2_{Y_-} = 1 - \alpha\, c^2 \\ c_M &= \alpha\, c\,, \end{aligned} \tag{A.3}$$

where $\alpha = \frac{2}{\pi}$ and $c \equiv c(X,Y)$. These relations proceed from the joint Gaussianity of $(X,Y)$ and consequent marginal Gaussianity of $X$ and $Y$, by integrating the corresponding probability densities in the subdomains above (subscript "+") and below (subscript "−") the median of $X$.

## A. DERIVATION OF EQ. (2.11)

We now note that, in the jointly Gaussian case, $c_+ = c_-$. This being said, and taking (A.3) into consideration, Eq. (A.1) becomes:

$$\begin{aligned} c &= \alpha c + c_+ \sqrt{2(1-\alpha)}\sqrt{1-\alpha c^2} \\ &= \alpha c + c_- \sqrt{2(1-\alpha)}\sqrt{1-\alpha c^2}\,. \end{aligned} \tag{A.4}$$

Therefore, $c_+$, $c_-$ can be written in the following form:

$$c_+ = c_- = c\sqrt{\frac{1-\alpha}{1-\alpha c^2}}\,. \tag{A.5}$$

We have thus come to Eq. (2.11).

# Appendix B

# Derivation of Eq. (2.47)

The Gaussian Mutual Information, $I_g(X,Y)$, is the MI between variables the joint distribution of which is Gaussian with the same mean and covariance matrix as that of the joint distribution of $(X,Y)$.

If $X$ and $Y$ are jointly Gaussian, $I_g(X,Y) = I(X,Y)$. If not, then there is an extra of information to account for the non-Gaussian features of their joint distribution: $I_{ng}(X,Y)$, presented in Sec. 2.3.6 and given by:

$$I_{ng}(X,Y) = I(X,Y) - I_g(X,Y)\,. \tag{B.1}$$

The contribution of non-Gaussianity to MI is thus defined as the difference between the actual Mutual Information $I(X,Y)$ between the variables $X$ and $Y$ and the Mutual Information $I_g(X,Y)$ that $X$ and $Y$ would have if they were jointy Gaussian.

We now recall (2.40) now in Differential Entropy form:

$$I(X,Y) = h(X) + h(Y) - h(X,Y)\,, \tag{B.2}$$

and apply it to $I_g(X,Y)$, thus yielding the following expression:

$$I_g(X,Y) = h(X_G) + h(Y_G) - h[(X,Y)_G]\,. \tag{B.3}$$

## B. DERIVATION OF EQ. (2.47)

Note that if the joint distribution of $(X, Y)$ is Gaussian, then the marginal distributions of $X$ and $Y$ are also normal, hence the subscript $G$ in $X$ and $Y$ in (B.3). We remind, however, that while joint Gaussianity implies marginal Gaussianity, the converse may not hold, i.e. marginal Gaussianity does not imply a joint one.

By subtracting (B.3) from (B.2) onto (B.1), we are led to:

$$I_{ng}(X,Y) = h(X) - h(X_G) + h(Y) - h(Y_G) - h(X,Y) + [h(X,Y)_G]\,. \tag{B.4}$$

By recalling the definition of Negentropy (2.28) in Sec. 2.3.3, we have:

$$\mathsf{J}(X,Y) = h[(X,Y)_G] - h(X,Y) \tag{B.5}$$

$$\mathsf{J}(X) = h(X_G) - h(X) \tag{B.6}$$

$$\mathsf{J}(Y) = h(Y_G) - h(Y)\,. \tag{B.7}$$

Consequently, $I_{ng}$ becomes:

$$I_{ng}(X,Y) = \mathsf{J}(X,Y) - \mathsf{J}(X) - \mathsf{J}(Y)\,, \tag{B.8a}$$

which is equivalent to

$$I_{ng}(X,Y) = \mathsf{J}(X,Y_r) - \mathsf{J}(X) - \mathsf{J}(Y)\,, \tag{B.8b}$$

since, as noted in Sec. 2.3.6, Negentropy $\mathsf{J}(X,Y)$ is invariant under a two-dimensional linear homeomorphism of $(X,Y)$, thus being, without loss of generality, equal to the Negentropy between the uncorrelated variables $X$ and the prediction residue $Y_r$ (2.14).

We have thus reached (2.47).

# Appendix C

# Derivation of Eqs. (3.54)-(3.55)

We hereby derive Eqs. (3.55)-(3.56) as in Nicolis, Perdigão, and Vannitsem (2009).

To obtain an explicit form starting from Eq. (3.54) one needs to evaluate the first three time derivatives of $\mathbf{u}$ at $t = 0$, when the system is on its reference attractor. The first derivative is available from Eq. (3.6). Projecting along phase space dimension $i$ one obtains immediately the expansion coefficient $A_i$ in the form given by Eq. (3.56a).

To obtain the second derivative we differentiate Eq. (3.6) once with respect to time:

$$\frac{d^2\mathbf{u}}{dt^2} = \mathbf{J} \cdot \frac{d\mathbf{u}}{dt} + \frac{d\mathbf{J}}{dt} \cdot \mathbf{u} + \frac{d\mathbf{\Phi}}{dt} \cdot \delta\mu \,, \tag{C.1}$$

or, substituting $d\mathbf{u}/dt$ from Eq. (3.6),

$$\frac{d^2\mathbf{u}}{dt^2} + \mathbf{J} \cdot (\mathbf{J} \cdot \mathbf{u}) + \frac{d\mathbf{J}}{dt} \cdot \mathbf{u} + \left( \mathbf{J} \cdot \mathbf{\Phi} + \frac{d\mathbf{\Phi}}{dt} \right) \cdot \delta\mu \,. \tag{C.2}$$

Projecting this relation along phase space direction $i$ yields the expansion coefficient $B_i$ in the form given by Eq. (3.56b).

Finally, to obtain the third derivative we differentiate Eq. (C.2) once with

respect to time substituting again $d\mathbf{u}/dt$ from Eq. (3.6),

$$\begin{aligned}\frac{d^3\mathbf{u}}{dt^3} =& \frac{d\mathbf{J}}{dt}\cdot(\mathbf{J}\cdot\mathbf{u}) + \mathbf{J}\cdot\left(\frac{d\mathbf{J}}{dt}\cdot\mathbf{u}\right) + \mathbf{J}\cdot[\mathbf{J}\cdot(\mathbf{J}\cdot\mathbf{u}+\mathbf{\Phi}\delta\mu)] \\ &+ \frac{d^2J}{dt^2}\cdot\mathbf{u} + \frac{d\mathbf{J}}{dt}\cdot(\mathbf{J}\cdot\mathbf{u}+\mathbf{\Phi}\cdot\delta\mu) \\ &+ \left(\frac{d\mathbf{J}}{dt}\cdot\mathbf{\Phi} + \mathbf{J}\cdot\frac{d\mathbf{\Phi}}{dt} + \frac{d^2\mathbf{\Phi}}{dt^2}\right)\cdot\delta\mu\end{aligned} \tag{C.3}$$

or, grouping similar terms together,

$$\begin{aligned}\frac{d^3\mathbf{u}}{dt^3} =& \mathbf{J}\cdot[\mathbf{J}\cdot(\mathbf{J}\cdot\mathbf{u})] + 2\frac{d\mathbf{J}}{dt}\cdot(\mathbf{J}\cdot\mathbf{u}) + \mathbf{J}\cdot\left(\frac{d\mathbf{J}}{dt}\cdot\mathbf{u}\right) \\ &+ \frac{d^2\mathbf{J}}{dt^2}\cdot\mathbf{u} + \left(\mathbf{J}\cdot(\mathbf{J}\cdot\mathbf{\Phi}) + 2\frac{d\mathbf{J}}{dt}\cdot\mathbf{\Phi} + \mathbf{J}\cdot\frac{d\mathbf{\Phi}}{dt} + \frac{d^2\mathbf{\Phi}}{dt^2}\right)\cdot\delta\mu\,.\end{aligned} \tag{C.4}$$

Projecting this relation along phase space direction $i$ keeping in mind that the dots imply summation over intermediate indices yields the expansion coefficient $C_i$ in the form given by Eq. (3.56c).

# Appendix D

# Derivation of Eq. (3.60)

We hereby derive Eq. (3.62) as in Nicolis, Perdigão, and Vannitsem (2009).

For that purpose, we begin by deriving explicit expressions for the coefficients of $t$, $t^2$ and $t^3$ terms in Eq. (3.61). $\underline{t\text{-terms}}$

Using Eq. (3.56a) we obtain

$$2\sum_i A_i u_i = 2\left[\sum_{ij}\left(J_{ij}u_i u_j + \Phi_{ij}u_i\delta\mu_j\right)\right] . \tag{D.1}$$

$\underline{t^2\text{-terms}}$

The starting point are Eqs. (3.56a) and (3.56b). We expand the square of $A_i$ and add $B_i u_i$, grouping together terms that are quadratic in $u$, quadratic in $\Phi\delta\mu$, linear in $u$ and $\Phi\delta\mu$ and, finally, terms containing time derivatives of phase space functions. We obtain in this way,

$$\begin{aligned}\sum_i \left(A_i^2 + B_i u_i\right) = &\sum_{ijk}\left(J_{ij}J_{ik}u_j u_k + J_{ij}J_{jk}u_i u_k\right)\\ &+\sum_{ij}\Phi_{ij}\Phi_{ik}\delta\mu_j\delta\mu_k\\ &+\sum_{ijk}\left(2J_{ij}\Phi_{ik}u_j + J_{ij}\Phi_{jk}u_i\right)\delta\mu_k\\ &+\sum_{ij}\left(\frac{dJ_{ij}}{dt}u_i u_j + \frac{d\Phi_{ij}}{dt}u_i\delta\mu_j\right) .\end{aligned} \tag{D.2}$$

## D. DERIVATION OF EQ. (3.60)

$t^3$-terms

We first evaluate the $A_iB_i$ part of the coefficient using Eqs. (3.56a)-(3.56b). We carry out explicitly the multiplication of expression (3.56a) by (3.56b), grouping together terms as above. We also take special care to combine whenever possible terms involving time derivatives in such a way as to obtain expression where the time derivative acts on the full phase space function multiplying initial errors. This yields the expression:

$$\begin{aligned}\sum_i A_iB_i =& \sum_{ijkl} \left(J_{ij}J_{jk}J_{il}u_ku_l + J_{ij}\Phi_{ik}\Phi_{jl}\delta\mu_k\delta\mu_l\right)\\ &+ \sum_{ijkl} \left(J_{ij}J_{il}\Phi_{jk}u_l + J_{ij}J_{jl}\Phi_{ik}u_l\right)\delta\mu_k\\ &+ \sum_{ijl} J_{il}\frac{dJ_{ij}}{dt}u_ju_l\\ &+ \frac{1}{2}\sum_{ijk}\left[\frac{d}{dt}(\Phi_{ij}\Phi_{ik})\delta\mu_j\delta\mu_k + \frac{d}{dt}\left(J_{ij}\Phi_{ik}\right)u_j\delta\mu_k\right].\end{aligned} \tag{D.3}$$

Turning next to the $C_iu_i$ part of the coefficient of the $t^3$ term in Eq. (3.61) we obtain, using Eq. (3.56c) and grouping terms in the same way as above,

$$\begin{aligned}\sum_i C_iu_i =& \sum_{ijkl} J_{ij}J_{jk}J_{kl}u_iu_l\\ &+ \sum_{ijkl} J_{ij}J_{jk}u_i\Phi_{kl}\delta\mu_l\\ &+ \sum_{ijk}\left(\frac{dJ_{ij}}{dt}u_i\Phi_{ik}\delta\mu_k + \frac{dJ_{ij}}{dt}J_{jk}u_iu_k\right)\\ &+ \sum_{ijk}\frac{d}{dt}\left(J_{ij}J_{jk}\right)u_iu_k\\ &+ \sum_{ij}\frac{d^2J_{ij}}{dt^2}u_iu_j\\ &+ \sum_k\left[\sum_i\frac{d^2\Phi_{ik}}{dt^2} + \sum_{ij}\frac{d}{dt}\left(J_{ij}\Phi_{jk}\right)u_i\right]\delta\mu_k\,.\end{aligned} \tag{D.4}$$

Summing (D.3) and (D.4) divided by 3 yields the explicit expression of the coefficient of the $t^3$ term in Eq. (3.61).

The next step is to average expressions (D.1)-(D.4) over the invariant density of the reference attractor and the distribution of initial errors. The first operation alone will eliminate a number of terms, expressed entirely in terms of time derivatives of phase space functions. The terms concerned by this elimination are thus: the last two terms in (D.2); the last two terms in (D.3); and the last three terms in (D.4). The reason is that in an ergodic system phase space averages and long time averages along a typical trajectory are equal,

$$\left[\!\left[\frac{dg}{dt}\right]\!\right] = \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\frac{dg(t)}{dt} = \lim_{T\to\infty}\frac{1}{T}\left[g(T)-g(0)\right] , \tag{D.5}$$

where the brackets $[\![\cdot]\!]$ stand for the phase space average. Now for any bounded function $g$ (the kind of function one deals with in physical systems), $g(T)-g(0)$ is finite and hence the term in the right hand side of the last equality of Eq. (D.5) tends to zero in the long time limit.

Consider next the average of the remaining terms over the distribution of initial errors. A further drastic simplification will occur in the case of unbiased errors, $\overline{[\![u_j]\!]} = 0$, since all terms in $\delta\mu$ surviving the first averaging will give a vanishing contribution in (D.1)-(D.4). These are: the last term in (D.1): the third last term in (D.2); the third term in (D.3); and the second and third terms in (D.4). Assuming further that initial errors are uncorrelated, $\overline{[\![u_i u_j]\!]} = \overline{[\![u_i^2]\!]}\delta_{ij}^{\rm kr}$, will transform the last third term in (D.3) into the total derivative of $J_{il}^2$, which will give a vanishing contribution through the phase space averaging. Likewise, by averaging the terms in $\delta\mu_i\delta\mu_j$ one obtains $\overline{[\![\delta\mu_i\delta\mu_j]\!]} = \overline{[\![\delta\mu_i^2]\!]}\delta_{ij}^{\rm kr}$. By keeping track of all these steps one arrives, finally, at Eq. (3.62).

# Appendix E

# Proof of equivalence between Eqs. (3.50) and (3.59)

The equivalence between the formulations presented in subsections 3.3.3.2 and 3.3.5 is hereby established in the conditions under which the latter is presented, namely: generic $n$-dimensional systems, the variables of which span a $n$-dimensional phase space and a $n'$-dimensional parameter vector $\mu$ along with the corresponding rectangular $\mathbf{\Phi}(n \times n')$ matrix, with $n, n' \in \mathbb{N}$; term $\delta\mathbf{f}$ considered to be small enough to be neglected ($|\delta\mathbf{f}| \ll |\mathbf{\Phi}\delta\mu|$), thus leading $\mathbf{p}$ in Eq. (3.7) to become $\mathbf{p} = \mathbf{\Phi}\delta\mu$; $g_{ij}$ taken as self-adjoint (corresponding to $\mathbf{A}$ in subsection 3.3.3.2).

In particular, it is intended to be proven that expressions (3.50) and (3.59) are equivalent in such conditions.

For that purpose, we begin by recalling (3.50):

$$||\mathbf{u}^2(\tau)|| = U_0 + U_1\tau + \frac{1}{2}U_2\tau^2 + \frac{1}{6}U_3\tau^3 + \mathcal{O}(\tau^4)\,, \tag{E.1}$$

where $U_l$ ($l = 0, \cdots, 3$) are given by (3.42), (3.44), (3.46) and (3.49) respectively.

Next, we rewrite the terms $U_l$ in indicial form, taking $A_{ij} = g_{ij}$, $p_i = \Phi_{ik}\delta\mu_k$:

## E. PROOF OF EQUIVALENCE BETWEEN EQS. (3.50) AND (3.59)

$$U_0 = \sum_{ij} g_{ij}\, u_i\, u_j \tag{E.2a}$$

$$U_1 = 2 \sum_{ijk} g_{ij}\, u_i \left(J_{jk} u_k + \Phi_{jk} \delta\mu_k\right) \tag{E.2b}$$

$$\begin{aligned} U_2 = 2 \sum_{ijk} g_{ij} \Bigg\{ & \left(J_{ik} u_k + \Phi_{ik} \delta\mu_k\right) \left(J_{jl} u_l + \Phi_{jl} \delta\mu_l\right) \\ & + u_i \left[ \left( \frac{dJ_{jl}}{dt} + J_{jk} J_{kl} \right) u_l + \left( J_{jk} \Phi_{kl} + \frac{d\Phi_{kl}}{dt} \right) \delta\mu_l \right] \Bigg\} \end{aligned} \tag{E.2c}$$

$$\begin{aligned} U_3 = 6 \sum_{ijk} g_{ij} \Bigg\{ & \left(J_{ik} u_k + \Phi_{ik} \delta\mu_k\right) \left[ \left( \frac{dJ_{jm}}{dt} + J_{jk} J_{km} \right) u_m \right. \\ & \left. + \left( J_{jl} \Phi_{lm} + \frac{d\Phi_{jm}}{dt} \right) \delta\mu_m \right] + \frac{1}{3} u_i \left[ \left( \frac{d^2 J_{jm}}{dt} + 2 \frac{dJ_{jk}}{dt} J_{km} \right. \right. \\ & \left. \left. + J_{jk} \frac{dJ_{km}}{dt} + J_{jk} J_{kl} J_{lm} \right) u_m + \left( 2 \frac{dJ_{jl}}{dt} + J_{jk} J_{kl} \right) \Phi_{lm} \delta\mu_m \right. \\ & \left. + \left( J_{jl} \frac{d\Phi_{lm}}{dt} + \frac{d^2 \Phi_{jm}}{dt} \right) \delta\mu_m \right] \Bigg\} \end{aligned} \tag{E.2d}$$

where all intervening terms are evaluated at $\tau = 0$.

We now turn to the square error expansion in (3.59):

$$\begin{aligned} ||\mathbf{u}^2(t)|| = \sum_{ij} g_{ij} \Bigg\{ & u_i\, u_j + 2\, u_i\, A_j\, t + \left(A_i\, A_j + u_i\, B_j\right) t^2 \\ & + \left( A_i\, B_j + \frac{1}{3}\, u_i\, C_j \right) t^3 \Bigg\}. \end{aligned} \tag{E.3}$$

By setting $\widetilde{U}_l$ as the coefficient of $t^l$ (and noting that all intervening terms are evaluated at $\tau = 0$),

$$\widetilde{U}_0 = \sum_{ij} g_{ij}\, u_i\, u_j \tag{E.4a}$$

$$\widetilde{U}_1 = 2 \sum_{ij} g_{ij}\ u_i\, A_j \tag{E.4b}$$

$$\widetilde{U}_2 = \sum_{ij} g_{ij} \left( A_i \, A_j + u_i \, B_j \right) \tag{E.4c}$$

$$\widetilde{U}_3 = \sum_{ij} g_{ij} \left( A_i \, B_j + \frac{1}{3} \, u_i \, C_j \right) , \tag{E.4d}$$

Eq. (3.59) can be rewritten as:

$$||\mathbf{u}^2(\tau)|| = \widetilde{U}_0 + \widetilde{U}_1 \tau + \widetilde{U}_2 \tau^2 + \widetilde{U}_3 \tau^3 + \mathcal{O}(\tau^4) \, . \tag{E.5}$$

The equivalence between Eqs. (E.5) and (E.1) [i.e. Eqs. (3.59) and (3.50)] shall thus be established by proving that $\widetilde{U}_0 = U_0$, $\widetilde{U}_1 = U_1$, $2\,\widetilde{U}_2 = U_2$ and $6\,\widetilde{U}_3 = U_3$, i.e. that $l!\,\widetilde{U}_l = U_l$.

By applying onto $\widetilde{U}_l$ the definition of the coefficients $A_i$, $B_i$ and $C_i$ given in Eqs.(3.56), $\widetilde{U}_l$ become:

$$\widetilde{U}_0 = \sum_{ij} g_{ij} \, u_i \, u_j \tag{E.6a}$$

$$\widetilde{U}_1 = 2 \sum_{ijk} g_{ij} \, u_i \left( J_{jk} u_k + \Phi_{jk} \delta\mu_k \right) \tag{E.6b}$$

$$\begin{aligned} \widetilde{U}_2 = \sum_{ijk} g_{ij} \Bigg\{ & \left( J_{ik} u_k + \Phi_{ik} \delta\mu_k \right) \left( J_{jl} u_l + \Phi_{jl} \delta\mu_l \right) \\ & + u_i \left[ \left( \frac{dJ_{jl}}{dt} + J_{jk} J_{kl} \right) u_l + \left( J_{jk} \Phi_{kl} + \frac{d\Phi_{kl}}{dt} \right) \delta\mu_l \right] \Bigg\} \end{aligned} \tag{E.6c}$$

$$\begin{aligned} \widetilde{U}_3 = \sum_{ijk} g_{ij} \Bigg\{ & \left( J_{ik} u_k + \Phi_{ik} \delta\mu_k \right) \left[ \left( \frac{dJ_{jm}}{dt} + J_{jk} J_{km} \right) u_m \right. \\ & \left. + \left( J_{jl} \Phi_{lm} + \frac{d\Phi_{jm}}{dt} \right) \delta\mu_m \right] + \frac{1}{3} u_i \left[ \left( \frac{d^2 J_{jm}}{dt} + 2 \frac{dJ_{jk}}{dt} J_{km} \right. \right. \\ & \left. \left. + J_{jk} \frac{dJ_{km}}{dt} + J_{jk} J_{kl} J_{lm} \right) u_m + \left( 2 \frac{dJ_{jl}}{dt} + J_{jk} J_{kl} \right) \Phi_{lm} \delta\mu_m \right. \\ & \left. + \left( J_{jl} \frac{d\Phi_{lm}}{dt} + \frac{d^2 \Phi_{jm}}{dt} \right) \delta\mu_m \right] \Bigg\} . \end{aligned} \tag{E.6d}$$

By multiplying $\widetilde{U}_l$ by $l!$ (in particular $\widetilde{U}_2$ by 2 and $\widetilde{U}_3$ by 6) the $U_l$ terms are obtained as in Eq. (E.2). Consequently, Eqs. (3.59) and (3.50) are indeed equivalent in the conditions of section 3.3.5.

# References


Abdallah, A. S. and M. D. Plumbey, 2003: Geometric ICA using nonlinear correlations and MDS., Nara, Japan, Brain Science Institute. 161–166. 27

Abramov, R., 2006: A practical computational framework for the multidimensional moment-constrained maximum entropy principle. *J. Comput. Phys.*, **211**, 198–209. 28

Abramowitz, M. and I. A. Stegun, 1972: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, 1046 pp. 10, 30

Aires, W., B. Rossow, and A. Chedin, 2002: Rotation of EOFs by the independent component analysis: Towards a solution of the mixing problem in the decomposition of geophysical time series. *J. Atmos. Sci.*, **59**, 111–123. 10

Baker, G. A. and P. Graves-Morris, 1996: *Padé approximants*. Cambridge University Press, 233 pp. 76

Biau, G., E. Zorita, H. von Storch, and H. Wackernagel, 1999: Estimation of precipitation by kriging in the EOF space of the sea level pressure field. *J. Climate*, **12**, 1070–1085. 12

Bjerknes, V., T. Hesselberg, and O. Devik, 1911: *Dynamic Meteorology and Hydrography*. 2

Blinnikov, S. and R. Moessner, 1998: Expansions for nearly Gaussian distributions. *Astron. Astrophys. Suppl. Ser.*, **130**, 193–205. 28

Boer, G. J. and T. G. Shepherd, 1983: Large-scale two-dimensional turbulence in the atmosphere. *J. Atmos. Sci.*, **40**, 164–184. 73

Broer, H., C. Simó, and R. Vitolo, 2002: Bifurcations and strange attractors in the Lorenz-84 climate model with seasonal forcing. *Nonlinearity*, **15**, 1205–1267. 105

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908. 2

Buizza, R. and T. N. Palmer, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518. 2

Cassou, C., T. Laurent, J. W. Hurrel, and C. Deser, 2004: North Atlantic winter climate regimes: Spatial asymmetry, stationarity with time, and oceanic forcing. *J. Climate*, **17**, 1055–1068. 7, 8

Cavazos, T., 2000: Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. *J. Climate*, **13**, 1718–1732. 8

Charney, J. G. and J. G. DeVore, 1979: Multiple flow equilibra in the atmosphere and blocking. *J. Atmos. Sci.*, **36**, 1205–1216. 86

Chu, P. and L. Ivanov, 2007: On stochastic stability of regional ocean models in wind forcing. *Non. Proc. Geophys.*, **14**, 665–670. 57

Cohen-Tanoudji, C., B. Diu, and F. Laloë, 1977: *Quantum Mechanics*, Vol. 1. John Wiley and Sons, 898 pp. 56

Comon, P., 1994: Independent component analysis, a new concept? *Signal Process.*, **36**, 287–314. 10, 29, 30, 34

Cover, T. M. and J. A. Thomas, 1991: *Elements of Information Theory*. Wiley, 576 pp. 19, 22, 26

Dalcher, A. and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39A**, 474–491. 2, 57, 156

Daniels, H. E., 1954: Saddlepoint approximations in statistics. *Ann. Math. Statist.*, **25**, 631–650. 32

DelSole, T., 2004: Predictability and information theory. Part I: Measures of predictability. *J. Atmos. Sci.*, **61**, 2425–2440. 10

Dole, R. M., 1983: Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation. *Large-Scale Dynamical Processes in the Atmosphere*, B. J. Hoskins and R. P. Pearce, Eds., Academic Press, 95–109. 7

Edgeworth, F. Y., 1905: The law of error. *Cambridge Philos. Soc.*, **20**, 36–66, 113–141. 10, 28

Gasiorowicz, S., 1996: *Quantum Physics*. 2d ed., John Wiley and Sons, 470 pp. 55

Holton, J. R., 2004: *An Introduction to Dynamic Meteorology*. 4th ed., Academic Press, 535 pp. 118

Hurrell, J. W., 1995: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. *Science*, **269**, 377–387. 8, 40

Hurrell, J. W., M. P. Hoerling, A. S. Phillips, and T. Xu, 2004: Twentieth century North Atlantic climate change. Part I: Assessing determinism. *Clim. Dynam.*, **23**, 371–389. 8, 45

Hyvärinen, A., J. Karhunen, and E. Oja, 2001: *Independent Component Analysis*. John Wiley and Sons, 478 pp. 21, 23

Jaynes, E. T., 1982: On the rationale of maximum-entropy methods. *Proc. IEEE*, **70**, 939–952. 10

Jolliffe, I. T. and D. B. Stephenson, 2003: *Forecast Verification - A practicioner's Guide in Atmospheric Science*. Wiley, 254 pp. 13

Kalnay, E., 2003: *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, 364 pp. 101

Kendall, M. and A. Stuart, 1969: *The Advanced Theory of Statistics*. 3d ed., Griffin, London. 31

Kenney, J. F. and E. S. Keeping, 1951: *Mathematics of Statistics, Part 2*. 2d ed., Van Nostrand, 202 pp. 31

Kidson, J. W. and C. Thompson, 1998: A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate*, **11**, 735–753. 4

Kistler, R., E. Kalnay, W. Collins, S. Saha, G. White, J. Woollen, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, H. van den Dool, R. Jenne, and

M. Fiorino, 2001: The NCEP-NCAR 50-year Reanalysis: Monthly Means CD-ROM and Documentation. *Bull. Am. Met. Soc.*, **82**, 247–267. 35, 101

Kraskov, A., H. Stögbauer, and P. Grassberger, 2004: Estimating mutual information. *Phys. Rev. E*, **69**, 066–138. 9, 26, 28

Krishnamurti, T., J. Sanjay, A. Mitra, and T. Vijaya Kumar, 2004: Determination of forecast errors arising from different components of model physics and dynamics. *Mon. Wea. Rev.*, **132**, 2570–2594. 57

Leith, C. E., 1982: Statistical methods for the verification of long and short range forecasts. *ECMWF Seminars 1981*, 313–333. 156

Livezey, R. E. and W. Y. Chen, 1983: Statistical field significance and its determination by Monte-Carlo techniques. *Mon. Wea. Rev.*, **111**, 46–59. 38

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141. 2, 97

Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513. 82, 156

Lorenz, E. N., 1984: Irregularity: a fundamental property of the atmosphere. *Tellus*, **36A**, 98–110. 105

Lorenz, E. N., 1986: Atmospheric models as dynamical systems. *Perspectives in Nonlinear Dynamics, World Scientific Publishing Co.*, 1–17. 105

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, ECMWF, Reading, Berkshire, United Kingdom, 1–18. 57

Lorenz, E. N., 2005: Designing chaotic models. *J. Atmos. Sci.*, **62**, 1574–1587. 3

Marshall, J. and F. Molteni, 1993: Toward a dynamical understanding of planetary-scale flow regimes. *J. Atmos. Sci.*, **50**, 1792–1818. 3, 117, 118, 153

Marwan, N. and J. Kurths, 2002: Nonlinear analysis of bivariate data with cross recurrence plots. *Phys. Lett. A*, **302**, 299–307. 10

Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate*, **12**, 2256–2284. 4

Nicolis, C., 1992: Probabilistic aspects of error growth in atmospheric dynamics. *Quart. J. Roy. Meteor. Soc.*, **118**, 553–568. 2, 57

Nicolis, C., 2003: Dynamics of model error: Some generic features. *J. Atmos. Sci.*, **60**, 2208–2218. 2, 57, 82, 130

Nicolis, C., 2004: Dynamics of model error: The role of unresolved scales revisited. *J. Atmos. Sci.*, **61**, 1740–1753. 2, 57, 82

Nicolis, C., 2007: Dynamics of model error: The role of the boundary conditions. *J. Atmos. Sci.*, **64**, 204–215. 2

Nicolis, C., R. A. P. Perdigão, and S. Vannitsem, 2009: Dynamics of prediction errors under the combined effect of initial condition and model errors. *J. Atmos. Sci.*, **66**, 766–778. 56, 63, 76, 77, 79, 130, 149, 150, 177, 179

Nicolis, C., S. Vannitsem, and J. F. Royer, 1995: Short range predictability of the atmosphere: mechanisms for superexponential error growth. *Quart. J. Roy. Meteor. Soc.*, **121**, 705–722. 2

Nicolis, G., 1995: *Introduction to nonlinear science*. Cambridge University Press, 254 pp. 86

Nicolis, G. and C. Nicolis, 1995: Chaos in dissipative systems: Understanding atmospheric physics. *Adv. Chem. Phys.*, **91**, 511–570. 92

Nicolis, G. and C. Nicolis, 2007: *Foundations of Complex Systems*. World Scientific. 55, 56, 58, 59, 60, 61

Osborn, T. J., K. R. Briffa, S. F. B. Tett, P. D. Jones, and R. M. Trigo, 1999: Evaluation of the North Atlantic Oscillation as simulated by a climate model. *Clim. Dynam.*, **15**, 685–702. 9, 36

Ott, E., 2002: *Chaos in Dynamical Systems*. 2d ed., Cambridge University Press. 58, 60

Palmer, T. N., 1999: A nonlinear dynamical perspective on climate prediction. *J. Climate*, **12**, 575–591. 7

Paninski, L., 2003: Estimation of entropy and mutual information. *Neural Comput.*, **15**, 1191–1254. 28

Paninski, L., 2004: Estimating entropy on m bins given fewer than m samples. *IEEE Trans. Info. Theory*, **50**, 2200–2203. 28

Papoulis, A., 1991: *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 666 pp. 11, 29

Perdigão, R. A. P., 2004: *Independent Component Analysis of the Low Frequency Geopotential Height Field and its relevance to Precipitation regimes over the Euro-Atlantic region.* Dissertation for the degree in Geophysical Sciences – Meteorology and Oceanography. Department of Physics, University of Lisbon, Portugal. Supervisor: Prof. Doutor Carlos A. L. Pires. 10

Pires, C., 1996: *Prévision Atmosphérique à Long-Terme: Un Problème d'Hybridation Statistico-Dynamique*. Thèse de Doctorat de l'Université Pierre et Marie Curie – Paris VI. Spécialité: Météorologie. 133, 134

Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, **48A**, 96–121. 150

Pires, C. A. and R. A. P. Perdigão, 2007: Non-Gaussianity and asymmetry of the winter monthly precipitation estimation from the NAO. *Mon. Wea. Rev.*, **135**, 430–448. 4, 10, 13, 24, 27, 28

Pozo-Vásquez, D., M. J. E. Parra, F. S. Rodrigo, and Y. Castro-Díez, 2001: A study of NAO variability and its possible non-linear influences on European surface temperature. *Clim. Dynam.*, **17**, 701–715. 8

Pozo-Vázquez, D., M. Esteban-Parra, F. S. Rodrigo, and Y. Castro-Díez, 2001: The association between ENSO and winter atmospheric circulation and temperature in the North Atlantic region. *J. Climate*, **14**, 3408–3420. 4

Pugh, C. C., 1969: On a theorem of P. Hartman. *Amer. J. Math.*, **91**, 363–367. 155

Pugh, C. C., 1987: *Global Stability of Dynamical Systems*. Springer, 150 pp. 155

Reynolds, C. A., P. J. Webster, and E. Kalnay, 1994: Random error growth in NMC's global forecasts. *Mon. Wea. Rev.*, **122**, 1281–1305. 2, 57

Robinson, W. A., S. Li, and S. Peng, 2003: Dynamical nonlinearity in the atmospheric response to Atlantic sea surface temperature anomalies. *Geophys. Res. Lett.*, **30**, 2038. 7

Rogers, J. C., 2002: North Atlantic storm track variability and its association to the North Atlantic Oscillation and climate variability of northern Europe. *J. Climate*, **10**, 1635–1647. 8

Ropelewski, C. and M. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362. 4

Rossler, O. E., 1976: An Equation for Continuous Chaos. *Phys. Lett.*, **57 A**, 397–398. 111

Roulston, M. S. and L. A. Smith, 2002: Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, **130**, 1653–1660. 10

Savijarvi, H., 1995: Error Growth in a Large Numerical Forecast System. *Mon. Wea. Rev.*, **123**, 212–221. 156, 162, 167

Schubert, S. and Y. Schang, 1996: An objective method for inferring sources of model error. *Mon. Wea. Rev.*, **124**, 325–340. 57

Serreze, M. C., F. Carse, and R. G. Barry, 1997: Icelandic low cyclone activity: Climatological features, linkages with the NAO, and relationships with recent changes in the Northern Hemisphere circulation. *J. Climate*, **10**, 453–464. 9, 45

Shannon, C. E., 1948: The mathematical theory of communication. *Bell. Syst. Technol. J.*, **27**, 379–423. 9, 18

Silverman, B. W., 1986: *Density Estimation for Statistics and Data Analysis, Monographs on Statistics and Applied Probability*. Chapman and Hall, 175 pp. 10

Simmons, A. J. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647–677. 82

Simmons, A. J., et al., 2004: Comparison of trends and variability in CRU, ERA-40 and NCEP/NCAR analysis of monthly-mean surface air temperature. ERA-40 project report series 18, 38 pp. 101

Sivia, D. S., 1996: *Data analysis: A Bayesian Tutorial*. Oxford University Press, 200 pp. 10

Smith, L. A., C. Ziehmann, and K. Fraedrich, 1999: Uncertainty dynamics and predictability in chaotic systems. *Quart. J. Roy. Meteor. Soc.*, **125**, 2855–2886. 2

Swanson, K., T. N. Palmer, and R. Vautard, 2000: Observational error structures and the value of advanced assimilation techniques. *J. Atmos. Sci.*, **47**, 1327–1340. 133

Swanson, K., R. Vautard, and C. Pires, 1999: Four-dimensional variational assimilation and predictability in a Quasi-Geostrophic model. *Tellus*, **50A**, 369–390. 133, 134

Tabeling, P., G. Zocchi, F. Belin, J. Maurer, and H. Willaime, 1996: Probability density functions, skewness, and flatness in large Reynolds number turbulence. *Phys. Rev. E*, **53**, 1613. 28

Teixeira, J., C. Reynolds, and K. Judd, 2007: Time step sensitivity of nonlinear atmospheric models: Numerical convergence, truncation error growth, and ensemble design. *J. Atmos. Sci.*, **64**, 175–189. 2, 153, 154, 155, 157, 159, 160, 162, 167

Teixeira, J. and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496. 2

Thompson, P. D., 1957: Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow patterns. *Tellus*, **9**, 275–295. 2

Toth, Z. and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbation. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330. 2

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319. 133

Tribbia, J. J. and D. P. Baumhefner, 1988: The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies. *Mon. Wea. Rev.*, **116**, 2276–2288. 57

Tribbia, J. J. and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: and updated perspective. *Mon. Wea. Rev.*, **132**, 703–713. 57

Trigo, R. M., T. J. Osborn, and J. M. Corte-Real, 2002: The North Atlantic Oscillation influence on Europe: climate impacts and associated physical mechanisms. *Climate Res.*, **20**, 9–17. 4

Trigo, R. M. and J. P. Palutikof, 1999: Simulation of daily temperatures for climate change scenarios over Portugal: A neural network model approach. *Climate Res.*, **6**, 1161–1171. 8

Trigo, R. M., D. Pozo-Vazquez, T. J. Osborn, Y. Castro-Diez, S. Gámis-Fortis, and M. Esteban-Parra, 2004: North Atlantic Oscillation influence on precipi-

tation, river flow and water resources in the Iberian Peninsula. *Int. J. of Climatology*, **24**, 925–944. 4

Vannitsem, S. and C. Nicolis, 1997: Lyapunov vectors and error growth patterns in a T21L3 Quasigeostrophic model. *J. Atmos. Sci.*, **54**, 347–361. 133, 138

Vannitsem, S. and Z. Toth, 2002: Short-term dynamics of model errors. *J. Atmos. Sci.*, **59**, 2594–2604. 2, 72, 76, 130

Wackernagel, H., 2003: *Multivariate Geostatistics: an introduction with applications*. 3d ed., Springer. 12

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences - An Introduction*. Academic Press, 467 pp. 12

Wu, A. and W. W. Hsieh, 2004: The nonlinear association between ENSO and the euro-atlantic winter sea level pressure. *Clim. Dynam.*, **23**, 859–868. 7