

Department of Physics and Astronomy

University of Heidelberg

Master thesis

in Physics

submitted by

Robin Weber

born in Alkmaar, The Netherlands

2014

**A comparison of techniques
for the initialisation of seasonal-to-decadal
climate prediction**

This Master thesis has been carried out by Robin Weber

at the

Institute of Environmental Physics (IUP), Heidelberg

and the

Catalan Institute of Climate Sciences (IC3), Barcelona

under the supervision of

Prof. Dr. Werner Aeschbach-Hertig

and the co-supervision of

Prof. Dr. Francisco Doblas-Reyes

and

Dr. Alberto Carrassi

Ein Vergleich mehrerer Methoden zur Initialisierung von saisonalen bis dekadischen Klimavorhersagen:

Full Field (FFI) und Anomaly Initialisation (AI) sind zwei Methoden, die zur Initialisierung von saisonalen bis dekadischen Klimavorhersagen eingesetzt werden. FFI initialisiert das Modell anhand der Beobachtungen. Das hat zur Folge, dass die Trajektorien zum Attraktor hin abdriften. AI initialisiert das Modell anhand der beobachteten Anomalien, um Drift zu vermeiden. Wir vergleichen beide Methoden unter verschiedenen Szenarien fehlerbehafteter Beobachtungen und Modellen. Außerdem analysieren wir zwei weiterführende Methoden. Least Squares Initialisation (LSI) propagiert Beobachtungsinformation partiell initialisierter Systeme in unbeobachtete Räume des Modells anhand der Kovarianzen der Modellanomalien. Exploring the Parameters Uncertainty (EPU) dient der Korrektur des Drifts während des Modelllaufs.

Unseren Untersuchungen liegt ein einfaches Klimamodell zugrunde. Ergebnisse zeigen, dass bessere Beobachtungen die Vorhersagegenauigkeit von FFI erhöhen. Dagegen hängt die Vorhersagegenauigkeit von AI von Modellverbesserungen ab. Eine erfolgreiche Annäherung des Modellattraktors mit Hilfe von AI ist nur dann gewährleistet, wenn sich die Wahrscheinlichkeitsdichtefunktionen des Modells und des Klimas sich lediglich um die erste Ordnung unterscheiden. Signifikante Unterschiede höherer Ordnung können dazu führen, dass die Verteilung der Anfangsbedingungen weniger mit der Wahrscheinlichkeitsdichtefunktion des Modells übereinstimmt, mit einer verringerten Vorhersagegenauigkeit zur Folge. LSI und EPU führen zu Prognoseverbesserungen, die zum Einsatz in Modellen höherer Komplexität ermutigen.

A comparison of techniques for the initialisation of seasonal-to-decadal climate prediction:

Full Field (FFI) and Anomaly Initialisation (AI) are two schemes used to initialise seasonal-to-decadal climate prediction. FFI initialises the model on the observations, but trajectories drift towards the model's own attractor. AI assimilates the observational anomalies onto the mean of the model climate with the intention of avoiding drift. We compare both approaches under different circumstances of observational and model errors. We also analyse two advanced schemes. Least Squares Initialisation (LSI) propagates observational information of partially initialised systems to the whole domain using the covariance of the model anomalies. Exploring the Parameters Uncertainty (EPU) is a drift correction technique applied during the forecast run.

Experiments are carried out using an idealized coupled dynamics. Results suggest that an improvement of FFI necessitates refinements in the observations, whereas improvements in AI are subject to model advances. A successful approximation of the model attractor using AI is guaranteed only when the differences between model and nature PDFs are limited to the first order. Significant higher order differences leads to an initial conditions distribution that is less representative of the model PDF, resulting in degradation of skill. Both LSI and EPU lead to significantly improved skill scores, encouraging implementation in models of higher complexity.

Contents

1	Introduction	11
2	Backdrop	15
2.1	From numerical weather prediction to climate projection: features of different time scales	15
2.2	What do chaotic simple models have in common with the atmosphere? . .	17
2.3	Extended range prediction	20
2.4	Narrowing uncertainty by means of initialisation	23
2.5	State-of-the-art initialisation techniques	25
2.6	Data assimilation	27
3	Methods: A research test bed for the initialisation of s2d prediction	31
3.1	Research objectives	31
3.2	Problem framework	32
3.2.1	Excursus: Bias is induced by model errors alone	33
3.3	Standard initialisation approaches	34
3.3.1	Full Field and Anomaly Initialisation	34
3.3.2	Properties	35
3.3.3	A word on nudging	36
3.4	Advanced schemes	37
3.4.1	Least Squares Initialisation	37
3.4.2	Exploring the Parameters Uncertainty (EPU)	39
3.5	The idealized coupled model	40
3.6	Experimental Setup	43
3.7	Measured quantities	44
3.7.1	Verification of forecast skill	44
3.7.2	Quantifiers of model bias and drift	45
3.8	Stability analysis	46
3.8.1	Computation of Lyapunov exponents	50
4	Results	53
4.1	Research line 1: Relative performance of Full Field and Anomaly Initialisation	53
4.1.1	Initialisation performance with respect to different model compartments	53
4.1.2	Initialisation performance in the face of observational error	54
4.1.3	Initialisation performance in the face of model bias	55

4.1.4	The role of the initial shock with regards to performance	58
4.1.5	The role of the attractor	60
4.1.6	Approximation of the model PDF	63
4.1.7	Summary of research line 1	70
4.2	Research line 2: Advanced schemes	72
4.2.1	Least Squares Initialisation (LSI)	72
4.2.2	Exploring the Parameter Uncertainty (EPU)	72
4.2.3	Summary of research line 2	75
5	Conclusion	77
A	Lists	81
A.1	List of Figures	81
B	Bibliography	85

Abbreviations

AI	Anomaly Initialisation
AMOC	Atlantic Meridional Overturning Circulation
AMV	Atlantic Multi-decadal Variability
DA	Data Assimilation
DJF	December January February
ENSO	El Niño Southern Oscillation
EPS	Ensemble Prediction System
EPU	Exploring (the) Parameters Uncertainty
FFI	Full Field Initialisation
GCM	General Circulation Model
IPCC	International Panel for Climate Change
LSI	Least Squares Initialisation
MJO	Madden-Julian Oscillation
NAO	North Atlantic Oscillation
NWP	Numerical Weather Prediction
OSSE	Observing System Simulation Experiment
PDF	Probability Density Function
PDV	Pacific Decadal Variability
s2d	Seasonal (to) Decadal
SST	Sea Surface Temperature

1 Introduction

The 20th century witnessed the emergence of the field of climate science that has entered the public consciousness and spawned widespread political debate like no other. Pioneering work, to name only a few, covers the explanation of the Earth's glacial cycles by Milankovich [1941]; the idea of a global ocean conveyor belt by Broecker [1991]; the link between CFCs and the ozone hole by Solomon *et al.* [1986]; and the use of ice cores to reconstruct past climates, e.g. Thompson [2000]. After the discovery of the greenhouse effect by Fourier in 1824, and first studies by Arrhenius concerning the influence of carbon dioxide on surface temperatures in 1896, greenhouse gases (GHGs) have been identified as a driver of climate change [Pierrehumbert, 2010], and the effect to which our planet will be impacted by cumulative anthropogenic emissions is an urgent question. Plass [1956] predicted that the planet would be about 1 °C warmer in 2000 than in 1900 (IPCC [2013] estimate: between 0.65 – 1.06 °C over 1880-2012), and early reports about the effects of CO₂ on climate were addressed to the American National Academy of Sciences by Charney *et al.* [1979]. Combustion processes are a central pillar of today's economy; systematic measurements of rising CO₂ levels in the atmosphere were undertaken by Keeling after 1958 [Keeling *et al.*, 2009]. The economic cost of sealing emissions is weighed against likely future impacts on climate. Along with observations of the Earth's response to anthropogenic GHG forcing, prediction is inherently a central tool of the discussion.

Climate models today synthesize knowledge acquired from studies of the Earth's climate system. The use of climate models in research on carbon dioxide and climate began in the 1970s, and largely improved in the wake of increasing computational power, coupling of additional components, and increasing model resolution and complexity. The implementation of such models in centurial climate projections have simulated a warming response towards GHGs, the strength of which is dependent on the future emissions scenario [Hawkins and Sutton, 2009]. The uncertainties of such predictions are derived from the scenario uncertainty, and model deficiencies [Hawkins and Sutton, 2009].

Climate prediction on shorter time scales has recently begun to play a larger role [Meehl *et al.*, 2009]. In the context of *climate services*, seasonal-to-decadal (s2d) prediction has received a lot of attention. Much like numerical weather prediction (NWP), the idea behind climate services is to build institutions that provide information to the society about how the climate is expected to change within the near future, so that adaptive measures and planning can take place. However, such a shift towards operational climate forecasting bares a new set of demands that need to be met. The first such demand is that s2d models can reproduce seasonal-to-decadal climate variability patterns with sufficient skill. Although climate models used for centurial projections display decadal variability, it has previously not been the goal of models to predict such patterns, which were averaged

out through multiple runs [Haines *et al.*, 2009]. The second central demand posed by climate services is *regional* prediction skill, the lack of which will severely limit any service to a country or other stakeholder. Inferences from centurial projections have trusted only global mean averages; too large have been the model-related errors with respect to, for example, the Atlantic meridional overturning circulation (AMOC), to dare any regional assessments. Seasonal and decadal variability patterns such as the El Niño Southern Oscillation (ENSO) have regional character, so that both demands are connected, and dependent on how well the model performs. This is by no means obvious; NWP relies perhaps more heavily on observations and initialisation, whereas centurial projections are more strongly dependent on the scenario uncertainty [Hawkins and Sutton, 2009].

Model-related uncertainties aside, predicting so called *internal variability* [Hawkins and Sutton, 2009] will require the incorporation of knowledge with regards to the state of today's climate. If, first-of-all, variability patterns exist; secondly, they are predictable; then thirdly, we need to synchronize our simulated climate with the observed pattern. Points one and two have been observationally affirmed; the third must be achieved through a correct *initialisation* of the system. Furthermore, besides a tracking of the variability, using observations to initialise forecasts can correct the up-to-date response to GHG forcing [Lee *et al.*, 2006]. Previous centurial projections using climate models have been initialised from randomly selected preindustrial states [Meehl *et al.*, 2009], neglecting observations on the grounds that initial condition signature is lost on such time scales.

S2d prediction therefore calls for schemes to initialise predictions. Full Field Initialisation (FFI) incorporates observations, wherever available, directly into the model. However, due to the unavoidable presence of model deficiencies, forecasts experience a drift from the observations towards the climate of the model, negatively affecting prediction skill [Magnusson *et al.*, 2012]. Furthermore, initial conditions inconsistent with the model climate can lead to rapid re-adjustments known as initialisation shocks, resulting in short-term loss of skill. Anomaly Initialisation (AI) assimilates only the observed anomalies on an estimate of the model mean climate. It has been devised with the goal of overcoming initialisation shocks, as well as model drift, by initialising the model closer to its own attractor [Smith *et al.*, 2007]. Comprehensive comparisons of FFI and AI on s2d time scales using state-of-the-art coupled climate models have recently appeared [Magnusson *et al.*, 2012; Smith *et al.*, 2013; Hazeleger *et al.*, 2013]. Results have indicated improved skill after initialisation at the seasonal time scale, with a slightly better skill for FFI [Magnusson *et al.*, 2012; Smith *et al.*, 2013]. On the decadal scale, studies have shown skill either in favor of AI [Smith *et al.*, 2013] or FFI [Hazeleger *et al.*, 2013].

State estimation theory in geosciences is commonly referred to as data assimilation [Daley, 1991], and has a long-standing tradition in NWP. The goal is to provide the best possible estimate of the system's state based on statistical or dynamical information, which is then used as the initial condition for predictions [Kalnay, 2003]. Observations are assimilated into the models according to their relative accuracies, and unobserved variables and model grid points are taken into account by means of their correlations, allowing for propagation of observational information and reduction of inconsistencies. Data assimilation (DA) has dramatically contributed to enhance forecast skill in NWP, and is now regarded with attention by the s2d community. Coupled data assimilation schemes are

being developed for climate models with the hope of reducing inconsistencies between separate model compartments, but have a long way to go [Dee, 2013]. Such schemes are in contrast to simple FFI and AI, and promise better initial conditions.

This study aims to contribute to the discussion on adequate initialisation approaches for s2d prediction, and is concerned with two main objectives. The first objective compares the performance of AI and FFI for different observational and model error scenarios. The second objective introduces and assesses the skill of two advanced formulations. Least Squares Initialisation (LSI) utilizes correlations between variables of different model compartments in order to propagate observational information and reduce inconsistencies in the spirit of coupled DA. Exploring the Parameters Uncertainty (EPU) is a technique that corrects the model drift during a forecast run, based on a short time approximation of the model error due to parametric uncertainty.

Our research is carried out using the low order climate model of Peña and Kalnay [2004]. It is based on the Lorenz-3-variable model [Lorenz, 1963], with a slow component strongly coupled to a fast component mimicking the tropical ocean-atmosphere coupling, and a weakly coupled fast component in analogy of the extratropical atmosphere. Initialisation schemes are assessed in the framework of an Observing System Simulation Experiment (OSSE, [Bengtsson *et al.*, 1981]) test bed, in which observations are sampled from a defined *true* trajectory, against which forecast skill is assessed as well.

Parts of this study have formed a publication [Carrassi *et al.*, 2014], currently under revision. A unified formalism based on the notation and concepts of DA theory from which FFI and AI can be derived, has been proposed. We make use of this formalism in our methodology.

Chapter 2 covers the background related to our research. Chapter 3 introduces the methodology, the aforementioned schemes, and the low order climate model. Results are given in chapter 4. Finally, conclusions and future work are summarized in chapter 5. We also encourage the reader to make use of the list of acronyms that follows the table of contents.

2 Backdrop

2.1 From numerical weather prediction to climate projection: features of different time scales

Numerical weather prediction (NWP) is an initial value problem. A model representing the necessary dynamics is integrated starting from an initial state informed by the most recent observations. The quality of a forecast $\mathbf{x}^f(t)$ will depend on how well the model $\mathbf{F}(\mathbf{x}^f)$ represents the real dynamics, as well as the quality of the initial conditions $\mathbf{x}^f(t_0)$. The system state is given by a state vector containing all information of all variables at all geographically located model grid points, the superscript f stands for the forecast state, and t represents the forecast *lead* time. The error $\mathbf{e}^f(t) = \mathbf{x}^f(t) - \mathbf{x}^{nat}(t)$ at any forecast time t (in the scale resolved by the model) is unknown, where \mathbf{x}^{nat} is a projection of the true continuous dynamics onto the discretized model grid. A set of observations \mathbf{y}^{obs} samples nature, subject to instrumentational errors and having a much smaller dimension than the model state due to a limited observational network. Merging the observations with the model state results in the best attainable representation of the nature state, which is the goal of the data assimilation process described in Sect. 2.6. The resulting *analysis* state is used for the initial conditions, as well as for forecast skill verification, establishing a proxy for the unknown nature state.

Despite the shortcomings intrinsic in the field with regards to the impossibility of perfect observations or their perfect geographical distribution, and the impossibility of a perfect model, the skill of today's weather forecasts up to several days in the future bears witness to how far NWP has advanced. This is widely due to the increased power of supercomputers allowing for much finer numerical resolutions and fewer approximations;

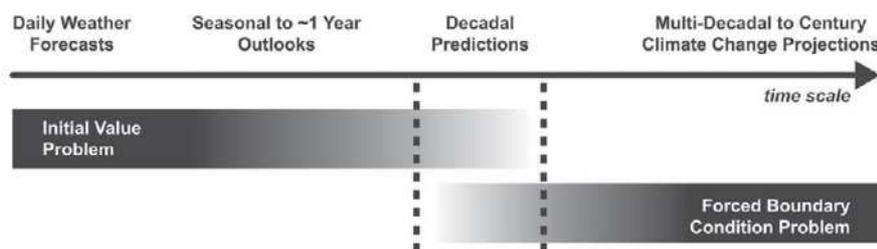


Figure 2.1: Schematic illustrating progression from initial value problems with daily weather forecasts at one end, and century climate change projections as a forced boundary condition problem at the other. Seasonal and decadal prediction falls in between. After Meehl *et al.* [2009].

improved representation of small-scale physical processes within the models; more accurate methods of data assimilation incorporating observations into the model and resulting in improved initial conditions; and a wider observational network [Kalnay, 2003].

Lorenz [1963] discovered what is now referred to as deterministic chaos: based on a simple system representing a convection cell, he found that solutions were unstable with respect to small modifications, so that slightly different initial states can evolve into considerably different states. Such a “sensitivity to initial conditions” represents an unsurmountable prediction barrier dependent on the finite accuracy of observations [Lorenz, 1982]. It has given way to an ensemble approach initialising several trajectories from perturbed, equally likely initial conditions, in order to assess the probability of a forecast by looking at the forecasted ensemble spread [Palmer, 1993]. The Ensemble Prediction System (EPS) thus explores the probability of a forecast by evolving perturbations according to the dynamics given by the model. This is particularly useful because the evolution of probability density functions (PDFs) using a high order nonlinear model is not feasible in practice. However, the ensemble size is far smaller than the system’s dimension and can only account for the system’s largest instabilities.

Long term climate projection is concerned with assessing the probability of climatic changes given a scenario of additional boundary conditions. Such endeavors have received widespread media attention in response to reports of the International Panel for Climate Change (IPCC). Whereas weather forecast models assume all components besides the atmosphere (ocean, land and cryosphere) as well as CO_2 and aerosol concentrations to be fixed for the short duration of interest, climate models make use of the full variability of all components of the system that play a role on time scales of about a hundred years, albeit at a coarser resolution. Observations of the Earth’s system are not directly incorporated in centurial predictions, because all memory of the initial conditions is expected to be lost on such time scales. Multiple runs (ensembles) of such models are made in order to average over the internal variability of the system. Uncertainty in climate representation is sampled using multi-models or multi-parameter results from the same model [Haines *et al.*, 2009]. Notice that the aforementioned prediction barrier is overcome through a paradigm shift from forecasting specific system states in NWP to time-averaged probability distributions of such states constrained by the system’s “climate”.

Figure 2.1 [Meehl *et al.*, 2009] indicates that climate projection is a forced boundary condition problem mainly determined by CO_2 and other aerosol or greenhouse gas concentrations. Seasonal and decadal prediction falls between daily weather forecasts using high resolution models, and lower resolution centurial projections, sharing features inherent to both categories. Seasonal forecast models implemented for time scales spanning up to twelve months have an active ocean component into which real ocean observations are assimilated in order to initialise the prediction [Anderson, 2008]. Due to the large signal of the El Niño Southern Oscillation (ENSO) throughout the Pacific and beyond, models are tuned to get the tropical Pacific to work well [Haines *et al.*, 2009].

In contrast to seasonal forecasting, decadal forecast horizons between one and ten years demand models that no longer operate under fixed external forcings from aerosols, CO_2 , or the solar cycle [Haines *et al.*, 2009]. Furthermore, it has been shown [Haines *et al.*,

2009; Smith *et al.*, 2007] that initialisation impacts forecast skill, and recent advances of observing technologies for the slower components of the climate system give rise to new potential. The Argo profiling float array in the oceans monitor the top two km of ocean heat content and density since 2007, and satellites launched by the ESA (SMOS) in 2009 and Nasa (Aquarius) in 2011 observe soil moisture and ocean salinity conditions. Emerging coupled data assimilation approaches [Dee, 2013] seek to optimally incorporate such observations of slow system components while maintaining congruence in the coupling of separate components, rather than assimilating observations into each compartment autonomously (discussed later in Sect. 2.6). Such schemes are still in their infancy, however, and current seasonal-to-decadal (s2d) forecasts resolve to initialisation of individual compartments resulting in initial states that might be incongruent with the model climate. Classical Full Field and Anomaly Initialisation under investigation in this thesis assimilate observations directly, in the absence of any data assimilation. Note that s2d prediction involves high-resolution models for better simulation of both regional climate and climate extremes in contrast to standard coupled models used for climate projection [Meehl *et al.*, 2009]. Future *seamless* climate predictions have been envisaged, predicting a number of time scales using different versions of the same model [Meehl *et al.*, 2009].

2.2 What do chaotic simple models have in common with the atmosphere?

In his groundbreaking paper on *Deterministic Nonperiodic Flow*, Lorenz [1963] found nonperiodic solutions of a simple forced dissipative system representing a convection cell. Although nonperiodic behaviour is a common feature in natural systems, distinguished especially in turbulent flow, the existence of deterministic nonperiodic solutions had not yet been established. For *linear* systems, constant or periodic forcing leads to a constant or periodic response, so that nonperiodic solutions were sometimes regarded as the result of nonperiodic or random forcing. Lorenz [1963] showed that nonperiodicity is not implicitly connected to randomness, but can be fully deterministic, with the addition that such systems are characterized as having a *sensitive dependence on initial conditions*. Besides having laid the foundation for a new branch of science in *chaos theory*, the existence of nonperiodic deterministic solutions based on a highly truncated model of convection dynamics has had profound practical implications in Earth system modelling. *Chaos* is defined as aperiodic long-term behaviour in a deterministic system that exhibits sensitive dependence on initial conditions [Strogatz, 2000]. Solutions do not settle down to fixed points, periodic trajectories, or quasi-periodic trajectories; the irregular behaviour stems from the system's nonlinearity and is not a consequence of random or noisy inputs; and nearby trajectories separate exponentially fast. An *attractor* is defined to be a closed, minimal, and invariant set that attracts an open set of initial conditions [Strogatz, 2000]. Any trajectory that starts in the attractor stays in the attractor for all time, and any trajectory that starts in its basin of attraction converges towards the attractor as $t \rightarrow \infty$. A *strange attractor* is an attractor of a system exhibiting sensitive dependence on initial con-



Figure 2.2: Evolution of an ensemble of initial points on the Lorenz (1963) attractor, for three sets of initial conditions in different phase space regions. Predictability is a function of the initial state. After Palmer [1993].

ditions [Strogatz, 2000]. A famous example of a strange attractor is the Lorenz-3-variable model's [Lorenz, 1963] “butterfly”, proven to exist by Tucker [1999].

Today, the chaotic nature of the atmosphere is widely accepted [Trevisan and Palatella, 2011]. This conclusion has been drawn from observational studies, as well as from work on low order truncations of atmospheric dynamics [Palmer, 1993]. First of all, weather forecasts starting from very similar initial conditions can evolve into very different atmospheric states [Buizza, 2000]. Then, simple models such as the Lorenz equations, or 10-component barotropic equations [de Swart, 1990], show qualitative similarities with the behaviour of the large scale atmosphere in the existence of a regime structure, as well as distinct time scales. Such time scales are connected to the *quasi-stationary* regimes, and correspond to an oscillation time around a regime centroid, and a residence time within a regime [Palmer, 1993]. In reality, weather regimes have been observed to be a feature of low-frequency atmospheric variability, in which planetary flow patterns appear repeatedly at fixed geographical locations and persist beyond the life times of individual weather disturbances [Ghil and Robertson, 2002].

Another distinguishing property of the Lorenz equations mirrored in the atmosphere is the existence of distinct phase space regions of high (low) instability associated with low (high) *predictability*. Figure 2.2 illustrates the phase space evolution of three different ensembles of initial points on the Lorenz [1963] attractor [Palmer, 1993]. Trajectories initialised from the ensemble in the top panel remain close together throughout the entire prediction. Trajectories initialised from the ensemble in the bottom left panel remain close initially, but eventually diverge. Trajectories initialised from the ensemble in the bottom right panel diverge almost immediately. In the last example, two entirely different

and equally probable system states can evolve from the ensemble of initial conditions, making forecasting fruitless. Thus, predictability is a function of the initial state, and the evolution of an ensemble can serve as a measure of forecast probability. Ensemble Prediction Systems are now common practice in NWP [Buizza *et al.*, 1999].

Perturbation growth is given by the *Lyapunov exponents*. In numerical studies of the Lorenz attractor, one finds that $\|\delta(t)\| \propto \|\delta_0\|e^{\gamma t}$, where $\gamma \approx 0.9$ is the (leading) Lyapunov exponent [Strogatz, 2000]. This value represents an average over the entire attractor, and describes the growth rate of a perturbation in the *linear* regime, i.e. for $\|\delta_0\| \rightarrow 0$. The linear error doubling time is then on average given by $t_2 \propto \frac{1}{\gamma} \ln 2$, a system-intrinsic value fixing a *predictability horizon* dependent on the initial error size. The exponential divergence *saturates* when the separation of trajectories is comparable to the diameter of the attractor, as they cannot evolve further apart [Strogatz, 2000]. Lyapunov exponents as well as their computation are discussed in Sect. 3.8. Suffice it to say that the calculation of such exponents are neither feasible nor particularly desirable in NWP, the former for the reason that the model dimensions are too large to get hold of the system's instabilities, the latter due to the fact that the fastest growth rate occurs on the scale of the Brownian motion, saturating rapidly.

Exponentially growing errors saturate eventually, i.e. their growth rate subsides as the magnitude of the errors increases [Lorenz, 1982]. A study by Lorenz [1969] indicated that even if the larger scales could be observed perfectly, the inevitable uncertainty in the smaller scales would after a day or so induce errors in the larger scales, comparable to the larger-scale initial errors which presently result from inadequate observations. The induced errors would then grow as if they had been present initially. This was later observationally confirmed [Somerville, 1979]. Thus, small scale features with fast doubling times that are unlikely to be resolved incur a bound on the accuracy of one-day forecasts, ultimately resulting in a mean *predictability barrier* [Lorenz, 1982]. Early estimates of atmospheric doubling times with realistic models amounted to about five days [Charney *et al.*, 1966], based on deviation rates of perturbed integrations from the control. The estimates decreased as models became more refined. Lorenz [1982] found an error doubling time of 2.5 days by comparing the differences between the 1-day forecast with the 2-day forecast from the preceding day and so on. Using the Lorenz [1982] method, Bengtsson and Hodges [2006] assess a potential increase in forecast skill compared with today by about 3 days in the extra-tropics, and as much as a week in the tropics.

In this section we have illustrated how work with simple models has found many analogs in the real atmosphere, covering sensitivity to initial conditions, regimes, characteristic time scales, predictability, and error saturation. Palmer [1998] investigates the response of the Lorenz equations to a small-amplitude imposed forcing as an analogue to anthropogenic climate change, finding a signal strongly dependent on the system's natural variability. In our study, we follow the philosophy of Hoskins [1983], that understanding is achieved through the interaction of NWP and conceptual models belonging to a continuous hierarchy, extended by Palmer [1998] to climate models too.

2.3 Extended range prediction

In Sect. 2.2, we have discussed the deterministic predictability barrier of approximately two weeks [Lorenz, 1982] as a consequence of the chaotic feature of the atmosphere. Prediction horizons beyond this limit are known as *extended range* [Palmer, 1993]. The objective of extended range predictions is to infer quantities describing the PDF of the atmospheric state, that can remain predictable on much longer time scales. Predictability can arise from *external* factors that alter the likelihood of residence in atmospheric attractors [Palmer, 1993]. In principle, a quantity can be defined as predictable if its probability of occurrence is larger than a random guess obtained from a *very long* statistical average, where *very long* might indicate a sufficiently larger time period (e.g. an order of magnitude) compared to the forecasted *mean* period. We exemplify this on the basis of the quasi-periodic solar forcing that modulates the weather PDF outside of the tropics on the time scale of the Earth's orbit. It is easy to infer that the month of August of the next year will most likely be on average warmer than the month of January for any location on the extratropical Northern hemisphere. A prediction of this kind in which we have predictive skill compared to a random draw from a multi-year statistical average appears trivial, but we can also establish that the same can be inferred for the prediction of the relative mean temperature difference between January and August of any year up to many thousands of years in the future. In relation, the half-year accuracy relative to the long prediction horizon already seems a lot more impressive. This kind of accuracy is solely due to the (very predictable) quasi-periodic nature of the Earth's orbit, as well as the impact due to the magnitude of the solar forcing on the Earth's atmosphere. The *Milankovich cycles* [Milankovich, 1941] represent documented signals in the Earth's *archives* (e.g. polar ice sheets) of much slighter changes in solar forcing, baring witness to the Earth's climatic response to a quasi-periodic forcing. Interestingly, this is the expected response of a *linear* system to such a forcing [Lorenz, 1963], indicating that a linear approximation of the Earth's *first order, equilibrium* response to a forcing is somewhat justified. In summary, the PDF of the Earth's atmospheric state, i.e. our *weather*, is clearly a function of boundary conditions such as the forcing.

Abstracting from the example above, if climatic parameters that exhibit a forcing on the Earth's atmosphere are predictable to some extent, then there is also hope to make inferences about the atmospheric PDF in response to the forcing. Such predictability due to evolving boundary conditions is known as *predictability of the second kind*, in contrast to *predictability of the first kind* associated with initial conditions [Kalnay, 2003]. The most prominent of such parameters are sea-surface temperatures (SSTs), but variables describing soil moisture content, land cover, sea-ice extent, and snow cover [Zwiers and Kharin, 1998; Doblas-Reyes *et al.*, 2013] all play a role too, also due to insulation and albedo effects. The slow nature of the *cryo-* and *hydrospheres* is much in favour of the predictability of such climate parameters. Note also that increased forcing due to anthropogenic greenhouse gases has now become an important source of predictability too.

In more general terms, climates in which a large fraction of the interannual variance of seasonal means originates from sources other than high frequency weather are often described as *potentially predictable* [Madden, 1976], where the word *potential* indicates

that the predictability also depends on the predictability of the anomalous forcing [Rowell, 1998]. This can be assessed in terms of a signal-to-noise ratio $F = \sigma_{tot}^2 / \sigma_{\epsilon}^2$ of the total variance divided by the unpredictable weather noise [Madden, 1976; Madden and Shea, 1978; Madden, 1989]. It is estimated using purely observational data at each point, can be plotted globally, and includes secondary sources of predictability, not just that due to SSTs [Rowell, 1998]. However, the underlying model that separates the signal from the unpredictable noise assumes a constant signal through each season, which is often untrue [Rowell, 1998]. An alternative method relies heavily on the underlying model's climate skill, and assumes potential predictability to be derived entirely from oceanic forcing. In this approach, potential predictability is measured using an ensemble of climate simulations, forced by the same observed interannually varying SSTs, but started from different initial atmospheric conditions. The sensitivity to initial conditions is used to quantify the random component of interannual variability, whereas the relative similarity is used to quantify the potentially predictable component of the total variance [Rowell, 1998].

The potential for skillful decadal predictions depends largely on whether models simulate sufficient decadal climate variability both in terms of magnitude as well as structure [Meehl *et al.*, 2009]. Zwiers and Kharin [1998] compare the ratios of the simulated total variance with the observed total variance, and the simulated weather noise with the observed weather noise for 30 different models within the Atmospheric Model Intercomparison Project (AMIP). The ratios indicate a large spread amongst models in terms of how well they simulate the observed total variance, as well as the weather noise for some variables. Figure 2.3 shows the ratio F between the interannual variance of the seasonal mean and the corresponding weather noise for December-January-February (DJF) 850 hPa temperature T_{850} as simulated by four different models calculated according to the alternative method described above. All models simulated less T_{850} potential predictability than computed from the observations. Although there is consistency between models in terms of the proportion of the globe at which significant predictability is found (e.g. strong evidence of potential predictability is seen in the tropics), there is considerable variation in the spatial pattern of significant F statistics [Zwiers and Kharin, 1998]. Encouragingly, Hegerl *et al.* [2007] have shown that the temperature variability of coupled climate models over global and continental space scales are realistic, even on time scales up to multiple decades.

Atmospheric long-term variability patterns exist, that have different time scales and spatial impact, and are commonly known as the interannual El Niño Southern Oscillation (ENSO), the decadal North Atlantic oscillation (NAO), the Pacific decadal variability (PDV), the Atlantic multi-decadal variability (AMV), and the Madden-Julian oscillation (MJO), to name only a few [Smith *et al.*, 2012]. The largest source of seasonal forecast skill is the ENSO [Smith *et al.*, 2012; Doblas-Reyes *et al.*, 2013], the first successful prediction pioneered by Cane *et al.* [1986] with a simplified coupled ocean-atmosphere model. This is in line with the predictability signal in Fig. 2.3, although such studies far exceed actual forecast quality achieved due to assumptions such as perfectly predicted boundary forcings [Doblas-Reyes *et al.*, 2013]. Warm sea-surface temperature anomalies in the tropical Pacific lead to increased ocean-to-atmosphere heat flux as well as a coupled feedback, impacting the structure of tropospheric rainfall latent heat release

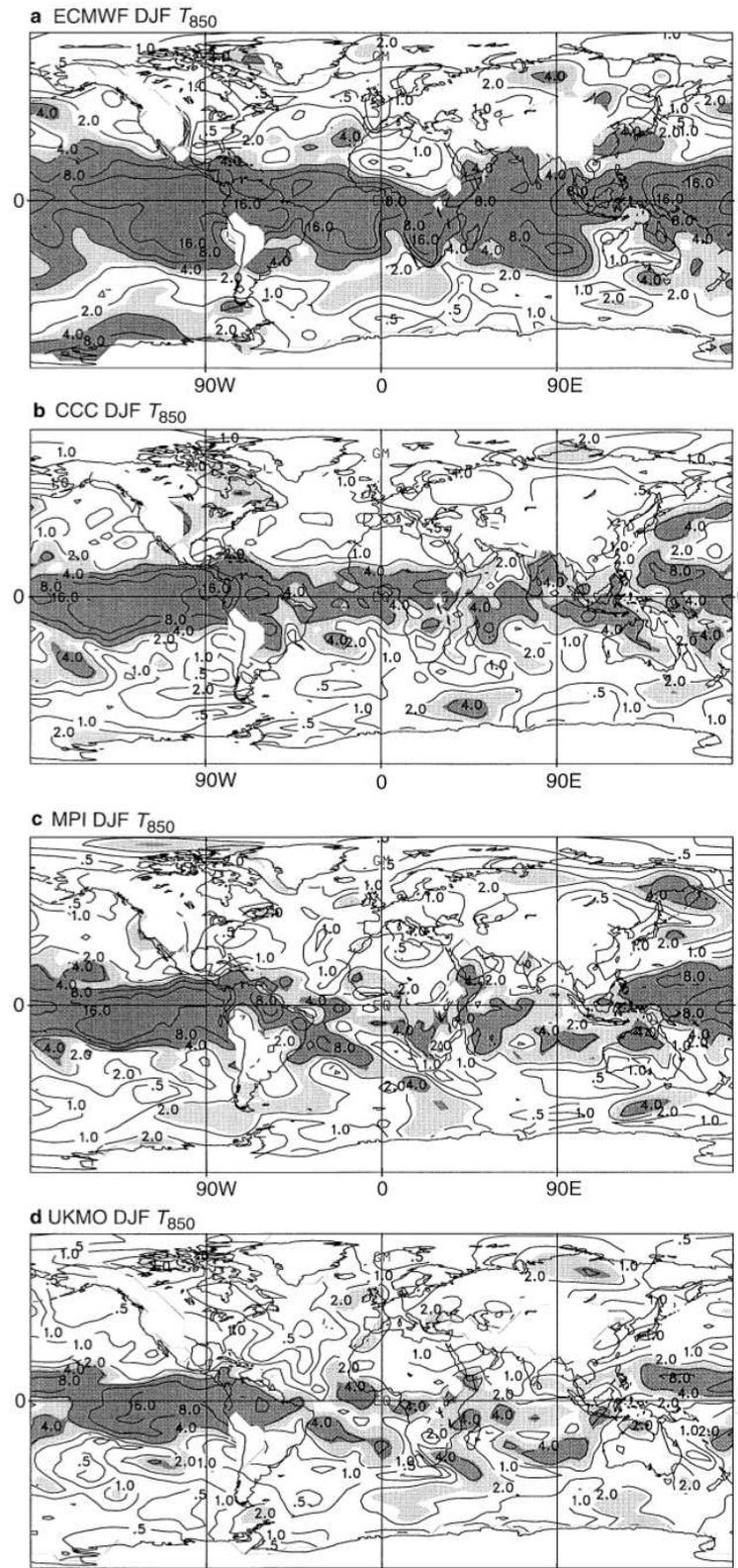


Figure 2.3: The potential predictability ratio F between the interannual variance of the seasonal mean and the corresponding weather noise induced variance for analysed DJF mean T_{850} as simulated by **a** ECMWF, **b** CCC model, **c** MPI model, **d** UKMO model. Contours are 0.5, 1, 2, 4, 8, and 16. Light (dark) shading corresponds to F ratios that are significantly greater than one at the 5% (1%) significance level. High terrain areas have been masked. After Zwiers and Kharin [1998].

with widespread teleconnections in remote regions of the globe [Doblas-Reyes *et al.*, 2013]. ENSO is now accurately predicted many months ahead, with dynamical models slightly more accurate than statistical techniques [Smith *et al.*, 2012]. For example, at 8 months multimodel correlation coefficients for Niño-3.4, a region in the tropical Pacific, are approximately 0.75, and then they drop to 0.6 and 0.5 at 10 months and 12 months respectively. Seasonal forecasts of Atlantic tropical storm activity are skilful and issued operationally [Smith *et al.*, 2012], too. In comparison, decadal prediction is still in its infancy [Haines *et al.*, 2009] and will be followed up in our discussion on initialisation below.

2.4 Narrowing uncertainty by means of initialisation

Climate predictions are subject to three sources of uncertainty to varying degrees, depending on the forecast time horizon and spatial scale [Hawkins and Sutton, 2009]. The first is the uncertainty connected to the chaotic features of the atmosphere, discussed in Sect. 2.2. Predictions aside, an atmospheric state can be chosen at random from the “observed” atmospheric PDF. This is also known as the *internal variability* of the system [Hawkins and Sutton, 2009]. The second source of uncertainty is related to *model deficiencies*, and the third is the *scenario uncertainty* associated with future radiative forcing due to GHG emissions [Hawkins and Sutton, 2009]. Figure 2.4 shows the total variance for the global decadal mean surface air temperature predictions split into the three sources of uncertainty as a function of lead time. On climate projection time scales, the scenario uncertainty contributes towards the largest total variance of the mean global temperature, whereas the internal variability is almost entirely insignificant. At the other end of the spectrum, the insets show that during the first two decades the global as well as local total variance is comprised only of uncertainties related to model inadequacy and internal variability. Additionally, we can see how the variance due to internal variability drops after making a decadal instead of an annual average, and that local uncertainties are likely to be higher than those related to global averages. The relative importance of internal variability and model uncertainty will differ for other variables. Nevertheless, it is likely that the uncertainty in regional climate predictions for the next few decades is dominated by model uncertainty and internal variability, that are potentially reducible through scientific progress [Hawkins and Sutton, 2009].

Improving climate models is an ongoing process, which is much enhanced in the wake of an improved observational network. In order to narrow the uncertainty with regards to the internal variability, its evolution can be predicted by *initialising* dynamical models with the current state of the climate system. However, initialisation is non-trivial, and difficulties must be overcome in order to achieve such improvements in reality [Smith *et al.*, 2012]. Predictions beyond a few weeks rely on observations of the slow coupled components, including e.g. the sub-surface ocean, which are sparse. Then, constraining a model with observations generally disrupts its dynamical balance, leading to rapid re-adjustments, known as initialisation shocks, which can lead to loss of forecast skill

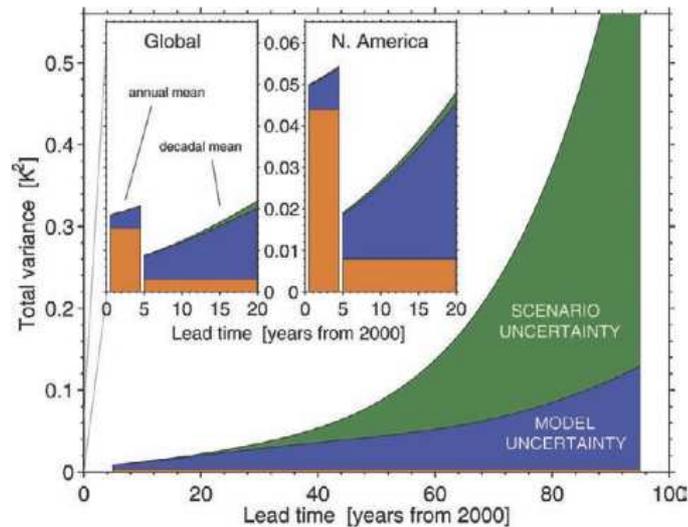


Figure 2.4: Total variance for the global decadal mean surface air temperature predictions, split into the three sources of uncertainty. Orange fields represent the internal variability component. Insets: As in the main panel, but only for lead times less than 20 years for (left) the global mean and (right) a North American mean. Lead times shorter than 5 years are plotted using annual mean data to highlight how the internal variability component is vastly reduced when considering decadal mean data. After [Hawkins and Sutton \[2009\]](#).

[[Smith *et al.*, 2012](#); [Doblas-Reyes *et al.*, 2013](#)]. Furthermore, model deficiencies result in a difference, or *bias*, between model simulations and observations averaged over a given period. Figure 2.5 gives an example of model bias for the HADCM3 general circulation model, showing a global map of the difference in model simulated and observed annual near surface temperatures averaged over the period 1960-2009. During the forecast, the model will drift back towards its preferred climate state, introducing errors that could be large compared to the predictable signal [[Smith *et al.*, 2012](#)]. Such drift can be neglected in NWP, because forecast horizons are short, and the model is regularly constrained by observations and thus kept close to the real climate [[Magnusson *et al.*, 2012](#)]. Also, in NWP the initial condition error growth masks the bias, which is related to model errors. Finally, imperfect model simulations of internal variability as well as imperfect responses to external forcing [[Smith *et al.*, 2012](#)] that might increase the bias over the forecast run, will limit the skill improvement achievable through initialisation.

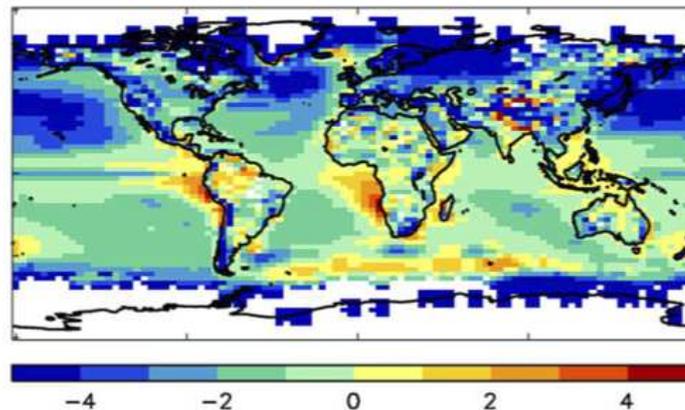


Figure 2.5: Example of model bias, shown by the difference in HADCM3 model simulated and observed annual near surface temperatures averaged over the period 1960-2009. After [Smith *et al.* \[2013\]](#).

2.5 State-of-the-art initialisation techniques

Full Field Initialisation (FFI) assimilates the observations directly into the model. The bias is corrected a-posteriori by applying a lead time dependent bias correction in post-processing. In seasonal forecasting, it is also made dependent on the seasonal cycle [[Magnusson *et al.*, 2012](#)]. A robust estimation of the bias requires a large data set of *hindcasts* (i.e. retrospective forecasts). However, biases can be flow-dependent, related to different initial conditions and the nonlinear nature of the system [[Magnusson *et al.*, 2012](#)]. If the bias is large enough, the nonlinear terms can become non-negligible, making such a linear calibration process insufficient [[Magnusson *et al.*, 2012](#)]. Furthermore, imperfect model responses to GHG forcing can lead to mean changes in the bias [[Smith *et al.*, 2012](#)], not accounted for in the hindcast period.

Anomaly Initialisation (AI) assimilates the *observed anomalies* into the model, i.e. the differences between the observations and the observed means (anomalies) are added onto the simulated mean of the model climate [[Magnusson *et al.*, 2012](#)]. In effect, this is the same as subtracting the model bias from the observations. The rationale is to avoid an initialisation shock occurring from an initial state far away from the model attractor, as well as to avoid model drift [[Magnusson *et al.*, 2012](#)]. Avoidance of shock is by no means guaranteed, since the structure of the observed anomaly may not be consistent with the model mean state, e.g. application of an observed sea-ice anomaly in regions where the model never has sea-ice [[Magnusson *et al.*, 2012](#)]. Avoiding model drift allows the a-posteriori bias correction to be independent on forecast lead time, and therefore more robust. However, calibration of the forecasts face the same problems as FFI, which are connected to the nonlinearities arising when a system's mean state is different [[Magnusson *et al.*, 2012](#)].

Practical concerns arise for both approaches. Computation of bias estimators for decadal prediction experiments is expensive. The seasonal independence of the bias correction

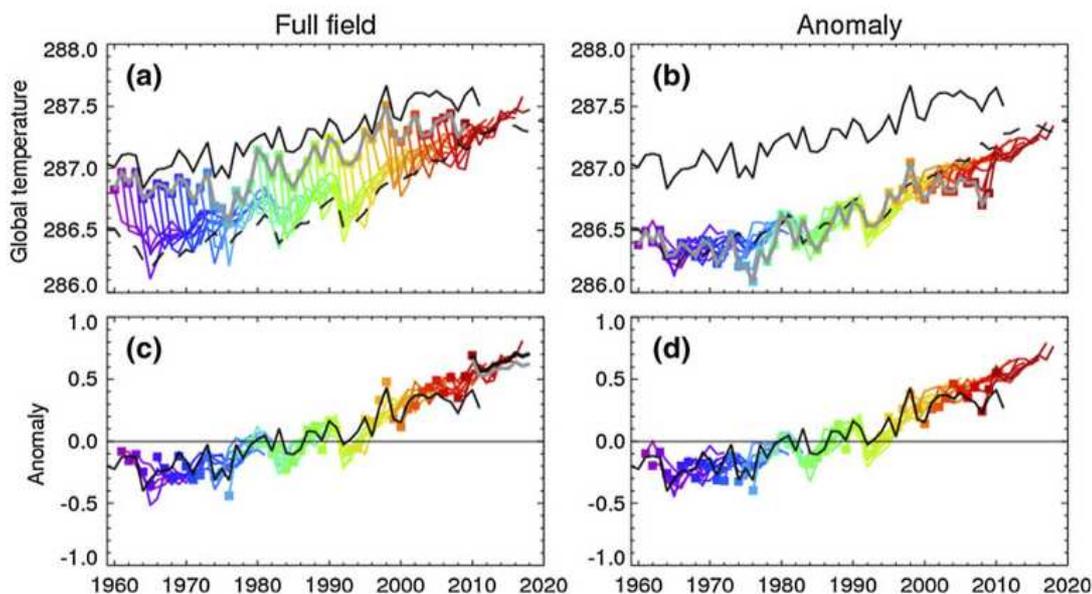


Figure 2.6: Illustration of bias correction for full field (left) and anomaly (right) initialisation. Thin black curves show the observed time series of annual mean global temperature. Coloured curves show the ensemble mean hindcasts, with different colours showing different start dates. (a) and (b) show absolute values (K). (c) and (d) show anomalies after adjustment for model biases. The solid grey and black dashed curves in (a) and (b) show values from the analyses and mean uninitialised model control runs respectively. After [Smith *et al.* \[2013\]](#).

technique when using Anomaly Initialisation thus allows less computation for equal level of “robustness”. On the other hand, the model climatological mean over a long period has to be computed for the implementation of AI. Moreover, the sampling period used for the observed climatology must be consistent with that used for the model climatology. Some regions like the southern oceans had not been observed prior to the advent of Argo, or only sporadically, so that there is a lack of information to derive a long-term climatology [[Magnusson *et al.*, 2012](#)].

Figure 2.6 [[Smith *et al.*, 2013](#)] illustrates the evolutions of forecasts (i.e. model integrations) after FFI (top left) and AI (top right). Given that such integrations have been run over a past record from 1960-today, they are also termed *hindcasts*, with the total period after 1960 known as the *hindcast period*. The thin black curve represents the observed annual mean global temperature, which has increased by about 0.5 K since 1960. The black dashed curve shows the mean of uninitialised long control runs, indicating a global average model bias of about -0.5 K. Note that the model control runs show a similar increase in global annual mean temperature due to greenhouse gas forcing. The grey curves show values from the *analyses*, which have already been described in Sect. 2.1 as the initial conditions after a merging of observations with the model state (here: according to either FFI or AI). The coloured curves show the ensemble mean hindcasts, with different colours corresponding to different start dates. After Full Field Initialisation we observe

that the forecasts drift within approximately the first year towards a cooler average model climate, before following the increasing model trend. After Anomaly Initialisation, the forecasts start close to the model climate and do not exhibit a discernable drift. Due to the cold bias of the model, it is more indicative to look at whether the model predicts the observed anomalies, rather than absolute values. The bottom panels display the observed as well as predicted anomalies after application of the bias correction procedure. Note that the observed record shows a recent stagnation of the warming trend in the last decade or so. As seen here, this stagnation has not been predicted by many models, which has caused much debate in the scientific community [Guemas et al., 2013]. This shows how the bias can change due to incorrect responses to greenhouse gases, that neither AI nor FFI can avoid [Smith et al., 2012].

Recent decadal prediction studies have demonstrated skill beyond the seasonal time scale using FFI, e.g. by Yeager et al. [2012]. Several have adopted Anomaly Initialisation, e.g. Smith et al. [2007]. It is important to assess their relative merits, and few such studies exist [Magnusson et al., 2012; Smith et al., 2013]. It is unclear which approach is best for seasonal-to-decadal prediction, and both are currently being evaluated [Smith et al., 2012].

2.6 Data assimilation

Data assimilation (DA) is a field that deals with the incorporation of data from observations into real physical models. Its main ambition is to obtain the most accurate estimate of the true state of a system given the information at hand. For most geophysical applications, the dimension of the physical model is greater than the dimension of the observational phase space. This means that any given observational input vector can be related to more than one estimate of the model state, making the problem underdetermined. However, knowledge of the past state of the system can further constrain the range of possibilities of the current model state. In NWP for example, in which the model is updated with new observations at regular short time intervals, the model state *implicitly* contains information about the past observations too. Accordingly, both incoming observational data and a calculated model trajectory are merged to infer an estimate of the truth that utilizes the availability of both sources of information and is expected to be more accurate than either sources alone. Accuracy is maximized by acquiring knowledge of the error statistics of both sources of information. If one source is noted to be more accurate than another, then weights to each source can be given appropriately. In an idealized scenario in which errors are normally distributed, a unique set of weights minimizing the error variance of the estimate, or so called *analysis*, can be found [Kalnay, 2003].

In a *Bayesian* approach, the model state uncertainty after a short forecast is also known as the *prior*, and the observational uncertainty is known as the *likelihood*. The resulting PDF, a product of the two PDFs around the mean model or observed state, is known as the *posterior*. This approach is associated with Bayes' theorem of conditional probabilities, and interpretes the incoming data as a *sequence*, ultimately resulting in the minimization of a cost function. The prior is sometimes also referred to as the *background* or the *first*

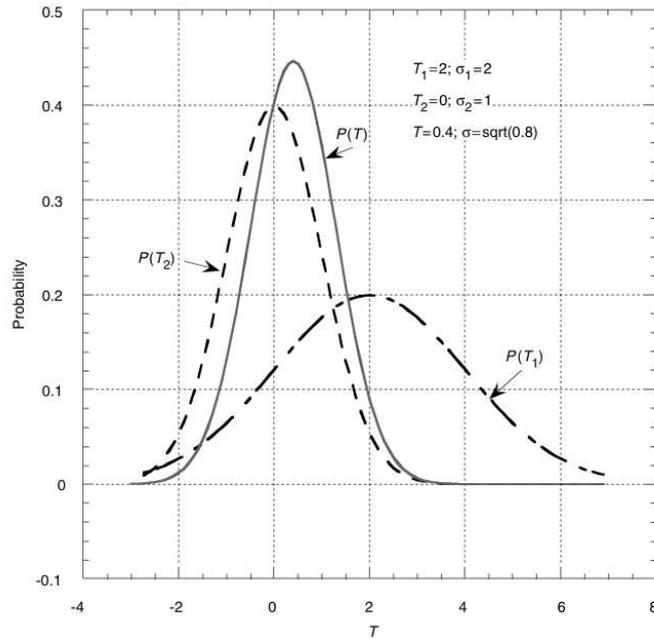


Figure 2.7: Illustration of the properties of the probability distribution of the analysis T , given observations T_1 and T_2 , using either the least squares or the Bayesian approach. After Kalnay [2003].

guess. In the following we only give an overview of the basic ideas behind DA, as well as an explanation of the equations necessary for our analysis. For a complete description of the many DA techniques, see Kalnay [2003].

We introduce main concepts using an example from Kalnay [2003]. Assume two independent measurements of the temperature at a given location and time. These can be written in the form

$$T_1 = T_t + \epsilon_1 \quad T_2 = T_t + \epsilon_2 \quad (2.1)$$

where T_t represents the *true* temperature, and $\epsilon_{1,2}$ are the respective errors associated with both measurements. We assume unbiased instruments, i.e. $\overline{\epsilon_1} = \overline{\epsilon_2} = 0$, and knowledge about the variances of the Gaussian observational errors σ_1^2 and σ_2^2 . Furthermore, we assume uncorrelated errors, i.e. $\overline{\epsilon_1 \epsilon_2} = 0$. Given this statistical information about both observations, we can estimate T_t from a linear combination of the two observations:

$$T = a_1 T_1 + a_2 T_2 \quad (2.2)$$

Imposing the condition of an unbiased *analysis* T , we write $\overline{T} = T_t$, so that $\overline{a_1 T_1} + \overline{a_2 T_2} = T_t$ and thus, given unbiased errors, $a_1 + a_2 = 1$. T will be the *best estimate* of T_t if the

coefficients are chosen to minimize the mean squared error of T :

$$\sigma_a^2 = \overline{(T - T_t)^2} = \overline{(a_1 T_1 + a_2 T_2 - T_t)^2} = \overline{(a_1(T_1 - T_t) + a_2(T_2 - T_t))^2} \quad (2.3)$$

Substituting $a_2 = 1 - a_1$, and minimizing with respect to a_1 gives

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \quad (2.4)$$

The weights of the observations are proportional to their accuracy. Substituting the coefficients of Eq. 2.4 into Eq.2.3, we obtain a relationship between the analysis variance and the observational variances:

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \quad (2.5)$$

Thus, if the coefficients are *optimal*, the errors *unbiased*, *uncorrelated* and their statistics *Gaussian*, then the precision of the analysis is the sum of the precisions of the measurements [Kalnay, 2003]. Figure 2.7 illustrates the probability distribution of the analysis $P(T)$, given the observations T_1 and T_2 and their associated uncertainties. The resulting analysis is more accurate than either observed temperatures.

Assume that one of the two temperatures $T_1 = T_b$ is the forecast or background value, and the other is an observation $T_2 = T_o$. Using Eq. 2.2 and Eq. 2.4, we can rewrite the analysis $T = T_a$ as

$$T_a = T_b + W(T_o - T_b) \quad (2.6)$$

where $(T_o - T_b)$ is known as the *observational innovation*, *update*, or *increment*, i.e. the new information brought by the observation. W is the optimal weight given by

$$W = \sigma_b^2(\sigma_b^2 + \sigma_o^2)^{-1} \quad (2.7)$$

Such a rearrangement of equations gives rise to the following interpretation of the problem. The analysis is obtained by adding the innovation to the first guess, weighted by the optimal weight. This weight is the background variance divided by the total error variance. The larger the background error, the larger the correction to the background. Finally, as before, the analysis precision is given by the sum of the background and observation precisions [Kalnay, 2003].

Equation 2.6 is equal to the Kalman Filter analysis update [Kalman, 1960]. Equation 2.6 bares the same form for multidimensional problems, in which T_b and T_a are three dimensional fields \mathbf{x}_b and \mathbf{x}_a of the order of 10^8 [Trevisan and Palatella, 2011] and T_o is a set of

observations \mathbf{y}_o of the order of 10^5 or 10^6 [Kalnay, 2003]. The error variances are then replaced by *error covariance matrices*, and the optimal weight by the *Kalman gain matrix* \mathbf{W} [Kalnay, 2003]:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{W}[\mathbf{y}_o - H(\mathbf{x}_b)] \quad (2.8)$$

where

$$\mathbf{W} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (2.9)$$

H is the *forward observational operator* mapping from model space into the observational space and thus converting the background field into “first guesses of the observations” [Kalnay, 2003]; \mathbf{H} is its linearized form; \mathbf{B} and \mathbf{R} are the respective background and observational error covariances of differing dimensions, transformed into the observational or model phase spaces with the help of \mathbf{H} .

We end this chapter with a few final comments. First of all, Eq. 2.8 and 2.9 are by no means limited to the field of NWP. What *is* unique to NWP, however, is that the background state is obtained from a short, recent forecast. Such a forecast has a much smaller error compared to a background state obtained from the entire climatology. This is practically feasible because forecast models are accurate enough, and the DA cycle is repeated every few hours or so. Thus, despite a limited observational network and an underdetermined problem of finding a model state in accordance with the observations, the model state is further implicitly constrained by past observations. The DA cycles “keep” the model close to the truth.

The second important point is connected to the error covariance matrices. The observational errors are reasonably assumed to be uncorrelated, so that the matrix \mathbf{R} is diagonal [Kalnay, 2003]. On the other hand, the background error covariance matrix \mathbf{B} reproduces the forecast error variance along its diagonal, but also the cross-correlation of forecast errors in its off-diagonal elements. This detail is crucial, because it allows the inference of correlations between variables and grid points [Kalnay, 2003]. In this way, even variables / grid points that have not been observed are updated due to their correlations with other observed variables / grid points. Observational information is *propagated* from observed to unobserved “regions” of the model by means of the forecast error structure, achieving *dynamical consistence*. *Coupled data assimilation schemes* are still in their infancy, but their goal is to propagate observational information across separate, but coupled compartments, e.g. the ocean and the atmosphere. In s2d prediction, compartments are currently assimilated with observations and initialised separately. This can lead to dynamical inconsistencies in the analyses, making forecasts more vulnerable to initialisation shocks. Finally, forecast errors, i.e. the entries of the matrix \mathbf{B} , are unknown. The NMC method [Parrish and Derber, 1992] estimates them by comparing 12h and 24h forecasts, and treating the 12h forecast as the “truth”. The assumption is that the model is a perfect representation of the underlying real atmospheric dynamics.

3 Methods: A research test bed for the initialisation of s2d prediction

3.1 Research objectives

We will begin this section by restating the objectives of our research already discussed in the Introduction. Our research is composed of two main research lines. The first line is related to the relative performance of Anomaly and Full Field Initialisation under different scenarios. The second line investigates the performance of advanced schemes.

The first branch of the first line is the comparison of Anomaly and Full Field Initialisation for different observational error, as well as different observational distribution scenarios. We investigate how initialisation limited to individual “oceanic” or “atmospheric” model compartments affects the performance of either algorithm.

The second branch of the first line explores the role of parametric model error on the relative performance of both algorithms. Anomaly Initialisation has been devised to tackle the effects of drift on forecast skill [Magnusson *et al.*, 2012]. Drift arises from errors in the model. Thus, studying the effects of model error on the different algorithms’ forecast skill is the foundation of any justifiable comparison.

The first branch of our second research line is the study of an advanced initialisation scheme that intends to propagate observational information of partially initialised systems to the whole model domain according to the forecast or *background* error structure of the model. Least Squares Initialisation (LSI) is based on standard practices in data assimilation (see Sect. 2.6) common in numerical weather prediction, from which it derives its name. We propose the use within a climate prediction context by replacing the forecast error covariance matrix with the covariances of the model anomalies.

The second branch of our second research line studies the performance of an online drift correction technique applied during the forecast run after initialisation. Exploring the Parameters Uncertainty (EPU) is designed to estimate, and subtract, the forecast bias related to parametric model error. This is done over short time intervals in which a linear approximation of the forecast error can be justified.

In order to meet the objectives, Carrassi *et al.* [2014] have developed a unified formalism using notation and concepts of data assimilation theory from which all of the above schemes can be derived, described in the following sections 3.2, 3.3 and 3.4. Our research is carried out using a simplified dynamics, introduced in Sect. 3.5, for better control and a statistically robust analysis. The experimental setup follows in Sect. 3.6. Verification of skill, and other measured quantities are summarized in Sect. 3.7. Finally, our setup requires an understanding of the stability features of the dynamical system. Section 3.8 covers the stability analysis of the systems we use. The methodology and experimental

setup introduced in this chapter is consistent with Carrassi *et al.* [2014].

3.2 Problem framework

In the following section we will formalise the problem and state the assumptions upon which our setup is based. We write the prognostic climate model under the form of an autonomous dynamical system¹:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \lambda) \quad (3.1)$$

where \mathbf{x} and λ are the state and parameter vectors of dimension I and P respectively. We define a true solution that will represent a “target”, which the models shall seek to predict. We call it “the nature”. We assume the nature to be given as

$$\frac{d\mathbf{x}^{nat}}{dt} = \mathbf{F}(\mathbf{x}^{nat}, \lambda^{nat}) + \mathbf{G}(\mathbf{x}^{nat}, \lambda^{nat}) \quad (3.2)$$

where $|\mathbf{G}| \ll |\mathbf{F}|$. The term \mathbf{G} represents the processes in nature that are not (well) represented in the prognostic models. Note that it does not account for errors arising from unresolved subscales, a weighty source of error in real climate models. Hence, model and nature span the same phase space of dimension I . In our experiments, we shall limit our focus to model error originating from parameter misrepresentation only, i.e. $\delta\lambda = \lambda - \lambda^{nat}$ and $\mathbf{G} = 0$.

We further assume that observations forming a vector $\mathbf{y}_i^o = \mathbf{y}^o(t_i)$ of dimension M are available at equally spaced times $t_i = i\tau, i = 0, 1, \dots$, where τ is a fixed time interval between successive observations. In typical applications, the observations vector has a much smaller dimension than the model state vector, i.e. $M \ll I$. The observed variables are assumed to be unbiased, affected by a Gaussian white noise of zero mean and standard deviation σ^o , i.e. $\epsilon^o \in N(0, \sigma^o)$.

$$\mathbf{y}^o = H(\mathbf{x}^{nat}) + \epsilon^o \quad (3.3)$$

where H is the observation operator mapping from the nature to the observational subspace of the model domain. In this formulation, the observational error accounts for both the instrumental as well as the representativity error² connected to H [Kalnay, 2003].

¹A dynamical system is prescribed by a fixed rule that determines the time dependence of a point in geometrical space. It is autonomous if this rule is time-invariant [Ott, 2002].

²Localised observations undergo value fluctuations subject to subscale dynamics not represented in the model. Therefore, observations can misrepresent larger scale model dynamics they are intended to sample. This leads to a *representativity error* [Kalnay, 2003].

3.2.1 Excursus: Bias is induced by model errors alone

The following derivation can be found in Carrassi *et al.* [2008]. We can write an expression for the forecast bias at time t as follows:

$$\mathbf{e}^{bias}(t) = \overline{\mathbf{e}^f}(t) = \overline{\mathbf{x}^f(t) - \mathbf{x}^{nat}(t)} \quad (3.4)$$

where \mathbf{e}^{bias} signifies the bias, \mathbf{e}^f the forecast error, and the overbar indicates the mean over an ensemble of initial conditions. The drift is given by a time derivative, i.e. the slope, of the bias.

$$\mathbf{d}(t) := \frac{d}{dt} \mathbf{e}^{bias}(t) \quad (3.5)$$

We now illustrate that these quantities are caused by errors in the model alone. The forecast error at time t is

$$\frac{d\mathbf{e}^f(t)}{dt} = \frac{d\mathbf{x}^f(t)}{dt} - \frac{d\mathbf{x}^{nat}(t)}{dt} = \mathbf{F}(\mathbf{x}^f, \lambda) - \mathbf{F}(\mathbf{x}^{nat}, \lambda^{nat}) \quad (3.6)$$

Let us first assume that the model is perfect, i.e. $\lambda = \lambda^{nat}$. Hence,

$$\frac{d\mathbf{e}^f}{dt} = \mathbf{F}(\mathbf{x}^f, \lambda^{nat}) - \mathbf{F}(\mathbf{x}^{nat}, \lambda^{nat}) \approx \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}^f} \cdot \mathbf{e}(t) \quad (3.7)$$

where we have linearized for small errors. We can solve this differential equation by separation of variables:

$$\ln \mathbf{e}^f = \int_{t_0}^t \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}^f} dt \quad (3.8)$$

This can also be written as

$$\mathbf{e}^f(t) = \mathbf{M} \cdot \mathbf{e}^f(t_0) \quad (3.9)$$

where $\mathbf{M} = \exp\left(\int_{t_0}^t \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}^f} dt\right)$ is the linear model propagator. If the initial conditions are unbiased, i.e. $\mathbf{e}^f(t_0) = 0$, then

$$\mathbf{e}^{bias}(t) = \overline{\mathbf{e}^f(t)} \approx \overline{\mathbf{M} \cdot \mathbf{e}^f(t_0)} = \mathbf{M} \cdot \overline{\mathbf{e}^f(t_0)} = 0 \quad (3.10)$$

We have shown that in the perfect model setup, for unbiased initial conditions and under

a linear hypothesis, no bias or drift occurs.

Let us now assume that our model equations are subject to parametric error. Hence, $\lambda \neq \lambda^{nat}$, $\delta\lambda = \lambda - \lambda^{nat}$. Then,

$$\frac{d\mathbf{e}^f(t)}{dt} = \mathbf{F}(\mathbf{x}^f, \lambda) - \mathbf{F}(\mathbf{x}^{nat}, \lambda^{nat}) \approx \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\mathbf{x}^f} \cdot \mathbf{e}(t) + \left. \frac{\partial \mathbf{F}}{\partial \lambda} \right|_{\lambda} \cdot \delta\lambda \quad (3.11)$$

Following similar steps as earlier, we obtain

$$\mathbf{e}^f(t) \approx \mathbf{M}_{t,t_0} \cdot \mathbf{e}(t_0) + \int_{t_0}^t \mathbf{M}_{t,\tau} \left. \frac{\partial \mathbf{F}}{\partial \lambda} \right|_{\mathbf{x},\lambda} \delta\lambda d\tau \quad (3.12)$$

where the first term is the general solution and the second term comprises a specific solution of the differential equation. Assuming again unbiased initial conditions,

$$\mathbf{e}^{bias}(t) = \overline{e^f(t)} \approx \overline{\int_{t_0}^t \mathbf{M}_{t,\tau} \left. \frac{\partial \mathbf{F}}{\partial \lambda} \right|_{\mathbf{x},\lambda} \delta\lambda d\tau} \quad (3.13)$$

Equation 3.13 gives the bias evolution in the presence of unbiased initial conditions, and model error coming only from misrepresentation of parameters. The linear analysis represents a good approximation for small errors at short times. We can thus conclude that in the case of parametric errors and unbiased initial conditions, the drift can only be caused by model deficiencies.

3.3 Standard initialisation approaches

We follow the formulation by Carrassi *et al.* [2014] of the standard initialisation algorithms for s2d prediction using the notation and concepts of data assimilation. Accordingly, the resulting initial state \mathbf{x}^a obtained after implementation of the initialisation procedure is called the *analysis*. In the absence of an initialisation procedure, the predictions are initialised on a *background* state \mathbf{x}^b , which is obtained from a long control run of the model after a transient *spin-up* period. Thus, the background state is by definition a state of the model, $\mathbf{x}^b := \mathbf{x}^m$. We shall assume that the observation operator is linear, given by a $M \times I$ matrix, \mathbf{H} .

3.3.1 Full Field and Anomaly Initialisation

The Full Field Initialisation approach reads as follows:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{H}^T [\mathbf{y}^o - \mathbf{H}\mathbf{x}^b] \quad (3.14)$$

On whichever grid points the observations are available, the background state is replaced by the observations. Where they remain unavailable, the background state is left unchanged. The term in the square brackets reveals the observational update to the background, also referred to as the *innovation* (see Sect. 2.6). If the observations are sampled at the grid points of the model and are measured directly, the observational operator is diagonal, its only entries encompassing the terms one or zero. If the entire system is observed, i.e. observations are available at all grid points, the observational operator is equal to the identity, $\mathbf{H} = \mathbf{I}$.

Similarly, for Anomaly Initialisation we obtain the equation

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{H}^T [\mathbf{y}^{ps_o} - \mathbf{H}\mathbf{x}^b] \quad (3.15)$$

where the observations are merely replaced by pseudo-observations

$$\mathbf{y}^{ps_o} = \mathbf{y}^o - (\bar{\mathbf{y}}^o - \mathbf{H}\bar{\mathbf{x}}^b) \quad (3.16)$$

The overbars indicate a time average, and the term given in brackets in Eq. 3.16 is the (negative) bias. This is easily shown in the case of full observational coverage of the system: $\bar{\mathbf{y}}^o = \bar{\mathbf{x}}^{nat}$ and $\mathbf{H}\bar{\mathbf{x}}^b = \bar{\mathbf{x}}^b$, hence $(\bar{\mathbf{y}}^o - \mathbf{H}\bar{\mathbf{x}}^b) = \bar{\mathbf{x}}^{nat} - \bar{\mathbf{x}}^b = -\mathbf{e}^{bias}$. The fact that it is negative owes to the definition of the bias in Eq. 3.4.

3.3.2 Properties

The initial, or *analysis* error can be written as $\mathbf{e}^a = \mathbf{x}^a - \mathbf{x}^{nat}(t_0)$. Assuming for simplicity observation of the full system (i.e. $\mathbf{H} = \mathbf{I}$) and using Eq. 3.14 and 3.15, the analysis errors for the respective algorithms are:

$$\mathbf{e}_{FFI}^a = \mathbf{e}^o, \quad \mathbf{e}_{AI}^a = \mathbf{e}^o + \mathbf{e}^{bias} \quad (3.17)$$

where \mathbf{e}^a , \mathbf{e}^o and \mathbf{e}^{bias} stand for the analysis and observational errors, and the bias respectively. Under the assumption of unbiased observations, the mean analysis errors sum up to:

$$\bar{\mathbf{e}}_{FFI}^a = 0, \quad \bar{\mathbf{e}}_{AI}^a = \mathbf{e}^{bias} \quad (3.18)$$

whereas for the initial root-mean-square error we get:

$$RMS E_{FFI}^a = \sigma^o, \quad RMS E_{AI}^a \approx \sqrt{\sigma^{o^2} + \mathbf{e}^{bias^2}} = \sigma^o \sqrt{1 + \frac{\mathbf{e}^{bias^2}}{\sigma^{o^2}}} \quad (3.19)$$

In real applications the bias is usually larger than the observational error ($\mathbf{e}^{bias} \gg \sigma^o$). In Eq. 3.19 we can see that the analysis rmse of AI in this case is much less sensitive to the observational error. Thus, the mean initial error of FFI is reduced after refinements in the observations, whereas the mean initial error of AI is effectively reduced when the prognostic model is improved.

On short forecast time scales such as in numerical weather prediction, efficiently reducing the initial error is desirable. For longer forecast horizons such as in s2d prediction for which growth of random error has already saturated, error due to model deficiencies plays a larger role. Therefore, the above stated initial error properties need not be decisive. Anomaly Initialisation intends to circumvent long-term drift, in the case of success resulting in a slower error growth rate despite a large initial error.

3.3.3 A word on nudging

A model can respond with a rapid initial adjustment to initial conditions that are inconsistent with its climate. It is prone to occur with Full Field Initialisation, when natural observations do not comply with the model, and is often referred to as a dynamical shock [Magnusson *et al.*, 2012]. Such a shock can sometimes be remedied using Anomaly Initialisation, but not if the structure of the observed anomaly is still, e.g. geographically, inconsistent [Magnusson *et al.*, 2012]. A different approach has been adopted from synchronisation theory [Bocchetti *et al.*, 2002], known as *nudging*. The idea is to *nudge* the model towards the desired state by relaxing it to a limited set of observations over a period of time. This is a common initialisation procedure for the ocean, in which it is relaxed towards observed SSTs [Magnusson *et al.*, 2012]. We do not assess the performance of this scheme in our study, but have added it for completion.

Nudging consists of the addition of a term to the prognostic equations that acts like an extra coupling term. The strength of the coupling $1/\tau$ is expressed as a relaxation time scale, depending on the properties of the variable chosen to be affected by the nudging. The time scale τ should be large enough to avoid a dynamical shock, but small enough to balance error growth [Hoke and Anthes, 1976]. The equation for Full Field Nudging (FFN) can be written as

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) + \mathbf{H}^T \mathbf{Q}[\mathbf{y}_i^o - \mathbf{H}\mathbf{x}^b] \quad t \leq t^{init} \quad (3.20)$$

with \mathbf{Q} being the diagonal $M \times M$ nudging matrix containing entries in units of time^{-1} . \mathbf{y}_i^o signifies the set of nudged observations (or their interpolations), with $i = 1, 2, \dots$ corresponding to their succession. The initial condition \mathbf{x}^a is obtained through integration up to the initialisation time, i.e. the start date, t^{init} . For Anomaly Nudging (AN), the observations are merely replaced by the pseudo observations in Eq. 3.16. FFN has been implemented operationally by Magnusson *et al.* [2012] and AN by Smith *et al.* [2007, 2013] to nudge oceanic variables.

3.4 Advanced schemes

In this section we follow the formulation by Carrassi *et al.* [2014] of two advanced algorithms that comprise the second main line of our research. Least Squares Initialisation (LSI) has the aim of improving the initial conditions of partially initialised systems by utilizing the information from observations of limited subspaces of the model domain, and propagating it to the entire model space by means of an approximation \mathbf{B}^m of the background error covariance matrix \mathbf{B} . The approximation \mathbf{B}^m contains the information on the covariances of the model anomalies in its off-diagonal elements. Anomalies of different variables and different geographical locations are often linked, and the knowledge of one such anomaly can inform a guess about another anomaly. (see Sect. 2.6 on data assimilation practices). The background error covariances are, in reality, time-dependent (e.g. different summer / winter anomaly structures). Given that \mathbf{B} has a dimension of $I \times I$ with I of the order of 10^8 , the computational requirements are enormous. We use a weaker time-invariant approximation.

Exploring the Parameters Uncertainty (EPU) is a drift correction technique applied during the forecast run after initialisation. It approximates the forecast bias originating from parametric model error over short time intervals that are congruent with the hypothesis of linearity. The approximation is based on a guess strategy that samples parametric errors occurring within the range of parametric uncertainty. The forecast bias is subtracted successively over the forecast run.

3.4.1 Least Squares Initialisation

Full Field (Anomaly) Initialisation replaces the background state with observations (observational anomalies) wherever available, and leaves it unchanged elsewhere. Thus, observations (observational anomalies) are not weighed according to their accuracy, but rather treated as if they were perfect. Furthermore, discrepancies can arise e.g. between neighbouring grid points if some remain unchanged. Data assimilation algorithms have been conceived in order to obtain a best estimate of the system given the relative accuracies of the observations and the background, as well as minimise inconsistencies between the observational and background information. These *first and second goals* can be considered as separate, but they are both tackled using the same background error covariance matrix \mathbf{B} . This works because \mathbf{B} contains the error information about the accuracy of the background along its diagonal; in its off-diagonal elements it contains the information about the correlation between different model variables, grid points, and levels (see Sect. 2.6).

LSI is based on a minimisation of the analysis error covariance, given the background \mathbf{B} and observational error \mathbf{R} covariances. The analysis \mathbf{x}^a is obtained from a linear combination of the background state \mathbf{x}^b and the observational update $[\mathbf{y}^o - \mathbf{H}\mathbf{x}^b]$:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1}[\mathbf{y}^o - \mathbf{H}\mathbf{x}^b] \quad (3.21)$$

The observational update is weighted according to the relative errors of both sources of information, where the relative errors are contained in the weighting term $\mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]$ (see Sect. 2.6). Observational errors from different instrumentation are known, and typically uncorrelated, so \mathbf{R} is assumed and set to be diagonal. However, the determination of \mathbf{B} is not as simple, and will be explained in the following.

Recall from Sect. 2.6 that in the context of NWP in which Eq. 3.21 holds, the background state \mathbf{x}^b is obtained from a short-time forecast. The model is sequentially updated with observations according to Eq. 3.21, and kept close to the “true” state. The important detail is that the forecast becomes the background in the following DA cycle. Thus, the background error covariance is obtained from the forecast error covariance. The background error after a short-time forecast is then defined as $\mathbf{B} = (\mathbf{e}^f)(\mathbf{e}^f)^T = (\mathbf{x}^f - \mathbf{x}^{nat})(\mathbf{x}^f - \mathbf{x}^{nat})^T$, where \mathbf{e}^f is the forecast error. Applying Eq. 3.21 in the context of s2d prediction, the background state \mathbf{x}^b is now obtained after a long control run of the model. Therefore, NWP methods for the estimation of \mathbf{B} no longer hold. Instead, Least Squares Initialisation is based on an approximation of \mathbf{B} :

$$\mathbf{B}^m = \overline{\alpha(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T} \quad (3.22)$$

where the overbars indicate long climatological time averages, and α can be described as an adjustment coefficient. Eq. 3.22 approximates \mathbf{B} using the covariances of the model anomalies. The assumption is that the model is unbiased, the weakness of which is somewhat compensated with α . With regards to the *first goal* and for $\alpha = 1$, the diagonal elements of \mathbf{B}^m reproduce the variance of the climate, underlining that \mathbf{x}^b is a random state of the climate associated with a large error equal to the climate variance. The observations are always more accurate in comparison. Thus, the function of \mathbf{B}^m is not primarily a representation of the accuracy of the background, which would have been in accordance with the *first goal*. The aim of Eq. 3.22 is to find a representation of the correlations between grid points and variables in its off-diagonal elements, in order to propagate information to unobserved domains of the model, and reduce inconsistencies. This is in line with the *second goal*. Notice also that \mathbf{B}^m is time-invariant. The computational requirement of a time dependent background covariance matrix for s2d prediction would be too large. \mathbf{B}^m is obtained from a single, long control run of the model, and can be used without further computational effort for initialisation. The presence of a climate change would further limit its performance.

[Smith and Murphy \[2007\]](#) have applied a similar strategy in the initialisation of the ocean. In our study we implement LSI under scenarios in which observed / unobserved model domains are represented by entire model compartments. In this situation LSI propagates observational information across model compartments, which is an implicit goal of coupled data assimilation schemes. Very advanced schemes will employ a time-dependent error covariance matrix.

3.4.2 Exploring the Parameters Uncertainty (EPU)

Under the assumption of unbiased observations, and using the framework set in Sect. 3.2, we show in Sect. 3.2.1 that biases in the forecasts are caused by model deficiencies only. In our online drift correction approach, we focus on model deficiencies related to errors in the model parameters alone.

In Sect. 3.2.1 we found an expression for the evolution of the bias:

$$\mathbf{e}^{bias}(t) = \overline{\mathbf{e}^f}(t) \approx \overline{\int_{t_0}^t \mathbf{M}_{t,\tau} \frac{\delta \mathbf{F}}{\delta \boldsymbol{\lambda}}|_{\mathbf{x},\boldsymbol{\lambda}} \delta \boldsymbol{\lambda} d\tau} \quad (3.23)$$

The approximation in Eq. 3.23 is due to the assumption of linearity. In realistic applications it cannot be solved, because the dimensions of the models involved are huge. We can expand Eq. 3.23 in a Taylor series up to the first order in time. Then,

$$\mathbf{e}^{bias}(t) \approx \overline{\frac{\delta \mathbf{F}}{\delta \boldsymbol{\lambda}}|_{\mathbf{x},\boldsymbol{\lambda}} \delta \boldsymbol{\lambda} [t - t_0]} \quad (3.24)$$

Eq. 3.24 represents an approximation of the bias over a short time interval. The accuracy of the approximation will depend on the time duration of the linear regime that is proportional to the largest (in absolute value) Lyapunov exponent of the dynamics [Nicolis, 2003].

The principle behind EPU is to subtract short time estimations of the bias, using Eq. 3.24, successively over the entire forecast run:

$$\mathbf{x}^{un}(t_i) = \mathbf{x}_i^{un} = \mathbf{x}_i^f - \mathbf{e}_i^{bias} = \mathbf{x}_i - \mathbf{C}_i \Delta T_{Bias} = \mathbf{x}_i - \frac{\delta \mathbf{F}}{\delta \boldsymbol{\lambda}}|_{\mathbf{x}_{i-1},\boldsymbol{\lambda}} \delta \boldsymbol{\lambda}_i \Delta T_{Bias} \quad (3.25)$$

where \mathbf{x}^{un} represents the corrected *unbiased* forecast state; ΔT_{Bias} signifies the bias correction time interval; $t_i = i \Delta T_{Bias}$ for $i = 1, 2, \dots$; and the compact form $\mathbf{x}(t_i) = \mathbf{x}_i$ is used to simplify the notation. The parameter vector $\boldsymbol{\lambda}$ is fixed over the entire model run, but $\delta \boldsymbol{\lambda}_i$ is the sampled parametric error at time t_i .

Equation 3.25 introduces the operator $\mathbf{C}_i = \frac{\delta \mathbf{F}}{\delta \boldsymbol{\lambda}}|_{\mathbf{x}_{i-1},\boldsymbol{\lambda}} \delta \boldsymbol{\lambda}_i$. Its first term $\frac{\delta \mathbf{F}}{\delta \boldsymbol{\lambda}}$ describes the functional dependence of the model on the uncertain parameters, which changes with the state of the model. Thus, it is a rectangular matrix projecting parametric error into the phase space of the model, with rows and columns corresponding to the dimensions of the state and parametric error respectively. It can be computed at any time along the model integration, constrained only by the model's dimensionality in case of large systems.

The second term of the operator \mathbf{C}_i relates to the parametric error vector $\delta \boldsymbol{\lambda}$. It is by nature unknown, and is sampled from the range of uncertainty $\Delta \boldsymbol{\Lambda} = (\boldsymbol{\lambda}_{min}, \boldsymbol{\lambda}_{max})$, that we assume

to have been identified, according to

$$\delta\lambda_i \in \begin{cases} U(0, \lambda_{max} - \bar{\lambda}), & \text{if } \lambda > \bar{\lambda} \\ U(\lambda_{min} - \bar{\lambda}, 0), & \text{if } \lambda < \bar{\lambda} \end{cases}$$

where the index i is associated with successive time intervals $t_i = i\Delta T_{Bias}$ for $i = 1, 2, \dots$; $U(a, b)$ is the uniform distribution in the interval (a, b) ; $\bar{\lambda}$ is the mean value of the range $\Delta\lambda$. $\bar{\lambda}$ plays the role of the “most probable” parameter, where in the case of $\lambda > \bar{\lambda}$ the parametric error is sampled as positive, and for $\lambda < \bar{\lambda}$ it is sampled as negative. Its role is to discriminate between over- and under-estimation of the unknown λ^{nat} . However, $\lambda^{nat} \neq \bar{\lambda}$ in general, so that the above guess strategy might erroneously select positive or negative errors.

3.5 The idealized coupled model

We base our experiments on an idealized coupled model introduced by [Peña and Kalnay \[2004\]](#). Three versions of the Lorenz 3-variable model [[Lorenz, 1963](#)] are coupled in order to mimic the behaviour of subsystems of fast and slow time scales:

$$\begin{aligned} \frac{dx_e}{dt} &= \sigma(y_e - x_e) - c_e(Sx_t + k_1) \\ \frac{dy_e}{dt} &= rx_e - y_e - x_e z_e + c_e(Sy_t + k_1) \\ \frac{dz_e}{dt} &= x_e y_e - bz_e \\ \\ \frac{dx_t}{dt} &= \sigma(y_t - x_t) - c(SX + k_2) - c_e(Sx_e + k_1) \\ \frac{dy_t}{dt} &= rx_t - y_t - x_t z_t + c(SY + k_2) + c_e(Sy_e + k_1) \\ \frac{dz_t}{dt} &= x_t y_t - bz_t + c_z Z \\ \\ \frac{dX}{dt} &= \tau\sigma(Y - X) - c(x_t + k_2) \\ \frac{dY}{dt} &= \tau(rx - Y - SXZ) + c(y_t + k_2) \\ \frac{dZ}{dt} &= \tau(SXY - bZ) - c_z z_t \end{aligned} \tag{3.26}$$

The capital letters represent the slow system referred to as the “ocean”, and the lower cases represent variables corresponding to the fast compartments referred to as the “extratropical atmosphere” (denoted with a subscript e) and the “tropical atmosphere” (with

subscript t). $\sigma = 10$, $b = 8/3$, and $r = 28$ are the standard values of the Lorenz system; c_e is the coupling strength of the “extratropical atmosphere” with the “tropical atmosphere”, c is the coupling strength of the “tropical atmosphere” with the “ocean” in the x - and y -variables, and c_z the coupling strength of the “tropical atmosphere” with the “ocean” in the z -variables; $k_1 = 10$ and $k_2 = -11$ are “uncentering” parameters introducing a phase lag between subsystems, and S and τ represent spatial and temporal scaling factors respectively. Note that when $c = c_z = 0$, the original Lorenz model is recovered.

The slow feature of the “ocean” is effectuated by setting $\tau = 0.1$. In accordance with [Peña and Kalnay, 2004], the spatial scaling factor is set to $S = 1$. However, the relative amplitudes of the compartments are characterized by the specific values of the coupling parameters in spite of the unit scaling factor. The choice of $c = c_z = 1$ results in a strongly coupled case reminiscent of the tropical El Niño-Southern Oscillation (ENSO), in which the “slave” “tropical atmosphere” has a small amplitude and undergoes regime changes clearly modulated by the slow “ocean” compartment. The “extratropical atmosphere” is weakly coupled ($c_e = 0.08$) with the “tropical atmosphere” component of the “ENSO”.

The model integration is carried out using a second order Runge-Kutta scheme with time step $\delta t = 0.01$. We establish a time definition on the basis of the duration of an ocean cycle, so that one “year” corresponds to 240 time steps. The simulated “ENSO” is the defining feature of our model, on the basis of which we expect to assess long-term predictability. Our time definition is connected to this feature through the cycles of the “ocean”.

Figures 3.1(a), 3.1(b) and 3.1(c) display the attractors of the “ocean”, “tropical atmosphere”, and “extratropical atmosphere” respectively. These have been obtained after omittance of a long transient spin-up and visualizing forty “year”-long trajectories in phase space. The “ocean” oscillates between a “normal” regime lasting typically three to twelve “years”, and an “El Niño” regime lasting only one “year”. The “tropical atmosphere” is visibly faster with a smaller amplitude. The “extratropical atmosphere” is close to the characteristic Lorenz model due to its weak “extratropical” coupling. Figure 3.1(d) displays the time series of the x -variables of all three compartments, illustrating the differences in scale as well as frequency between the slow, large-amplitude “ocean” and the faster “atmospheric” compartments.

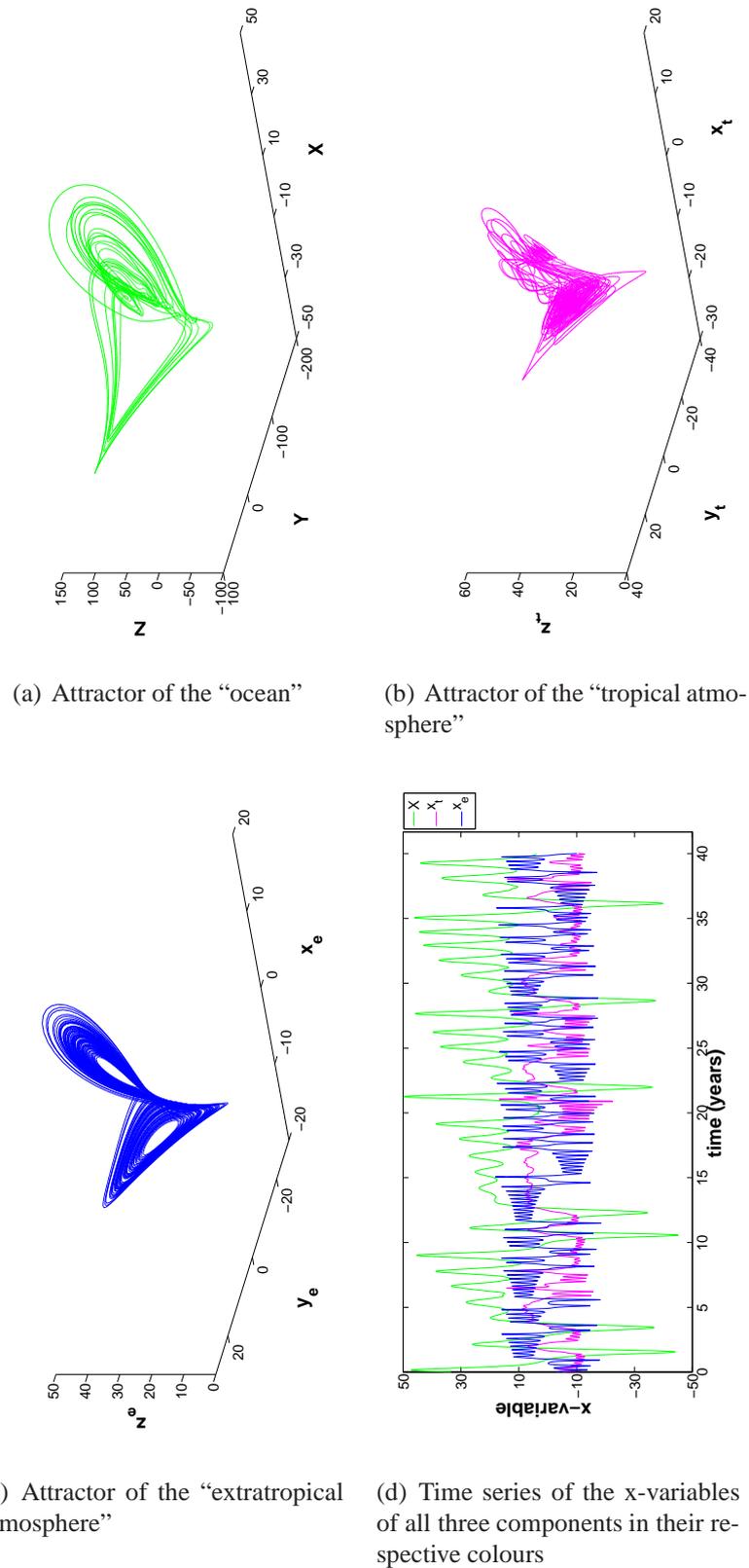


Figure 3.1: (a)-(c): Attractors of the 3-component coupled system. The x -, y - and z - axes correspond to the same variables. The “ocean” (a) is strongly coupled to the “tropical atmosphere” (b), which in turn is weakly coupled to the “extratropical atmosphere” (c). The time series of the x -variable of all three compartments are plotted in (d).

3.6 Experimental Setup

Our experimental setup is based on the standard observation system simulation experiment (OSSE [Bengtsson *et al.*, 1981]) configuration in which the simulated nature evolution is sampled at discrete times to generate the series of simulated observations. The idealized coupled model specified in Sect. 3.5 is integrated over an initial period of 60 000 time steps in order to reach a state on its own attractor. This period is also called the spin-up. Starting from the new initial condition given after the spin-up, the model is integrated over a trajectory of forty “years”. This trajectory is assigned the role of “the nature”, which will be targeted by the predictions. Thus, the nature is prescribed by a matrix \mathbf{N} of the dimension 9×9600 , its rows and columns determined by the number of variables and time steps respectively. For any state \mathbf{x}_i^{nat} of the nature, $\mathbf{x}_i^{nat} \in_i \mathbf{N}$, with $i \in \{1, \dots, 9600\}$, $[\mathbf{x}_i^{nat}] = [9 \times 1]$, and \in_i signifying that the vector belongs to the i th column of the matrix. Values for the “observations” of the variables are generated by sampling the nature trajectory and then adding a Gaussian white noise ϵ^o , with zero mean and standard deviation vector $\sigma^o \in N(0, \sigma^o)$. σ^o is set to a percentage of the system’s natural variability, with dimension $[\sigma^o] = [9 \times 1]$. The observations are distributed homogenously every “month”, i.e. every twenty time steps, for the first thirty years. They are prescribed by the matrix \mathbf{O} of dimension 9×360 , and given by $\mathbf{y}_l^{obs} \in_l \mathbf{O}$, with $l \in \{1, \dots, 360\}$ and $[\mathbf{y}_l^{obs}] = [9 \times 1]$.

We can now make a prediction of the nature by running the dynamical system starting from an observation. In order to simulate the fact that real climate models are imperfect, we introduce errors in the idealized coupled dynamics given by Eq. 3.26, and refer to these as erroneous models. Erroneous models are generated through the misspecification of the “tropical” coupling parameters c and c_z , or the forcing parameter r . Note that r plays the role of the Rayleigh number [Strogatz, 2000], which is an important parameter describing the balance between bouyancy force on a pocket of fluid as a result of thermal expansion with the loss of energy to thermal diffusion and viscosity [Hilborn, 1994]. However, it is directly connected to the temperature difference between the bottom and top of the convection cell described by the Lorenz equations [Hilborn, 1994]. A larger r corresponds to a larger difference in temperature, i.e. a larger forcing. Both coupling parameters are sampled from the range of (0.2 – 1.5) with step 0.1, and the forcing parameter from (29 – 68) with step 1.

Similarly to the nature trajectory, a control run is obtained by integrating the erroneous model for forty “years” after a spin-up period. The control run represents a reference point with regards to the assessment of forecast improvement after initialisation. It also gives the background state of the system \mathbf{x}^b , which is needed if the system is only partially initialised.

Following the typical hindcast format of climate prediction studies (see Fig. 3.2 or Sect. 2.5), ten-“year” predictions are made every “month” over thirty “years”, summing up to a total hindcast period of forty “years” coinciding with the length of the nature trajectory. The predictions are initialised using the “monthly” observational information according to either FFI or AI, comprising a total of 360 start dates. The forecasted trajectories are prescribed by the matrix \mathbf{F}_k^p , where the subscript k specifies the erroneous model, and the superscript $p = 1, 2, \dots, 360$ signifies the associated start date. A forecast state is given by

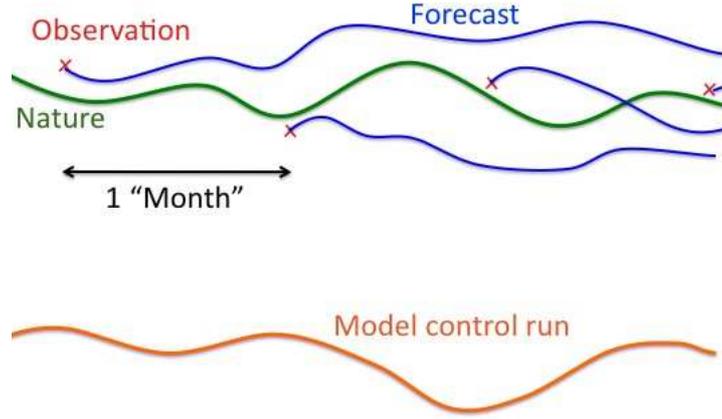


Figure 3.2: Schematic of the experimental setup in phase space. The observations (red crosses) are sampled each month from the nature (green line) and used to initialise the forecasts (blue lines). The model control run is shown (orange line) systematically below the nature, indicating a model bias.

$\mathbf{x}_j^f \in_j \mathbf{F}_k^p$, with $j \in \{1, \dots, 2400\}$ and $[\mathbf{x}_j^f] = [9 \times 1]$. In the case of a fully initialised system, the observation operator is given by $\mathbf{H} = \mathbf{I}$ and $[H] = [9 \times 9]$. If a single compartment is initialised, its dimension reduces to $[\mathbf{H}] = [3 \times 9]$. The model biases $\mathbf{H}\bar{\mathbf{x}}^m - \bar{\mathbf{y}}^o$ needed for the AI scheme (Eq. 3.15) are estimated using the sample of 360 observations / start dates, as well as the corresponding sample of the control run.

The model errors lead to different forecast biases, allowing for a kind of scenario control of models. An important constraint for the sampled parametric errors is the similarity of the stability features with those of the nature. The sampled models must be suitable representations of the nature on some specified level. This level is given by the degree of instability of the model compartments, which is reproduced by the positive Lyapunov exponents of a dynamical system. The instability features of a dynamical system give a notion of its predictability. A fair assessment of the relative forecast skill of AI and FFI with respect to the model bias requires similar predictability of the models involved.

The nature has two positive Lyapunov exponents, $\gamma_{1,2}^{nat} = 0.9063, 0.3150$. A total of 109 erroneously coupled models are sampled according to the parameter range given above, and under the constraint of having similar positive exponents. We refer to this set of models as $\{c, c_z\}$ -models. The forty erroneously forced models are referred to as r -models. See Sect. 3.8 for the analysis of the stability features of the selected erroneous models.

3.7 Measured quantities

3.7.1 Verification of forecast skill

The field of climate prediction is populated with different types of skill scores [von Storch and Zwiers, 1999]. The phase space of our deterministic coupled model is euclidean,

allowing us to use a skill measure based on the distance of the initialized model trajectory from the nature trajectory. This distance is given at a given lead time as an average root-mean-square error (RMSE) over all variables, where the RMSE of the variables are normalized with respect to their natural variances in order to account for differences in amplitude:

$$RMS E_{tot}(t) = \sum_{i=1}^9 \sqrt{\frac{(e_i^f(t))^2}{(\sigma_i)^2}} / 9 = \sum_{i=1}^9 \sqrt{\frac{(x_i^f(t) - x_i^{nat}(t))^2}{(\sigma_i^{nat})^2}} / 9 \quad (3.27)$$

where $i = 1, \dots, 9$ indicates the variables, and the overbar indicates the mean over the sample of predictions / initial conditions. Recall that in Anomaly Initialisation the bias is added onto the observations, i.e. the initial distance to the nature trajectory is larger. In order to facilitate a fair comparison between AI and FFI, we *debias* the total RMSE given in Eq. 3.27 by subtracting the bias (Eq. 3.4 in vector notation):

$$RMS E(t) = RMS E_{tot}^{deb}(t) = \sum_{i=1}^9 \sqrt{\frac{(e_i^f(t) - e_i^{bias}(t))^2}{(\sigma_i)^2}} / 9 \quad (3.28)$$

The root-mean-square skill score relates the forecast skill to a reference skill given by the non-initialised model control run, i.e. the background:

$$RMS SS(t) = RMS SS_{tot}^{deb}(t) = 100 \left(1 - \frac{RMS E^f(t)}{RMS E^b(t)} \right) [\%] \quad (3.29)$$

An $RMS SS(t) > 0\%$ means that the sample of 360 initialised predictions has better skill than the equivalent sample of non-initialised predictions.

3.7.2 Quantifiers of model bias and drift

We have defined the bias as a mean error, and the drift as the time derivative thereof, in Eq. 3.4 and 3.5. Our analysis is performed on the basis of the biases of single variables, as well as a global quantity describing the total model bias that we define as the *RMS Bias*. The *RMS Bias* is computed by taking the average of the root mean square differences between the model control run and nature means over all variables, normalized by their own variances. It describes the mean size of the bias over all nine normalized variables of the model.

$$RMS Bias = \sum_{i=1}^9 \sqrt{\frac{(\bar{x}_i^b - \bar{x}_i^{nat})^2}{(\sigma_i^{nat})^2}} / 9 \quad (3.30)$$

Such a “global” measure of the bias of a model allows for a broad comparison among models with respect to a single indicator. It can be interpreted as the total shift in phase space of the centre of the model’s attractor with respect to nature, impartial with regards to the contribution of each variable independently. The *RMS Bias* coincides with the global normalized bias for large lead times:

$$\lim_{t \rightarrow \infty} \sum_1^9 \sqrt{\frac{e_i^{bias}(t)^2}{(\sigma_i^{nat})^2}} / 9 = RMS\ Bias \quad (3.31)$$

For our purposes a quantification of the drift becomes relevant. Numerically, we calculate the drift of a variable d_i by taking the absolute value of the bias within a time period T_d , while respecting its sign and dividing by the lead time associated with this value, t_{max} .

$$d_i^{num} = \pm \frac{\max|e_i^{bias}(t = 0 : T_d)|}{t_{max}} \quad (3.32)$$

Drift can occur slowly over s2d time scales, or very rapidly in the form of an initialisation shock occurring due to inconsistent initial conditions with respect to the model attractor [Magnusson *et al.*, 2012]. The time scale with respect to which drift is investigated is determined by T_d .

3.8 Stability analysis

The stability properties of the erroneous configurations are centered on the computation of their spectrum of Lyapunov exponents, γ_i for $i = 1, 2, \dots, 9$. The Lyapunov exponents of a dynamical system are quantities that characterize the *linear* rate of separation of infinitesimally close trajectories (see also Sect. 2.2). We now introduce a definition of the Lyapunov exponent given in Ott [2002], followed by a more conceptual understanding given in Kalnay [2003].

The linear evolution of a perturbation $\mathbf{z}_0 = \delta\mathbf{z}_0$ is given by

$$\mathbf{z}(t) = \mathbf{M}_{t,t_0} \cdot \mathbf{z}_0 \quad (3.33)$$

where $\mathbf{M}_{t,t_0} = \exp(\int_{t_0}^t \frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}}|_{\mathbf{x}} dt)$ is the tangent linear model or *propagator*. $\mathbf{z}(t)/|\mathbf{z}(t)|$ gives the direction of the infinitesimal displacement of the trajectory from $\mathbf{x}(t)$, and $|\mathbf{z}(t)|/|\mathbf{z}_0|$ is the factor by which it grows or shrinks.

For initial condition \mathbf{x}_0 and initial orientation of the displacement given by $\mathbf{u}_0 = \mathbf{z}_0/|\mathbf{z}_0|$,

the Lyapunov exponent is defined as

$$\gamma(\mathbf{x}_0, \mathbf{u}_0) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln(|\mathbf{z}(t)|/|\mathbf{z}_0|) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln |\mathbf{M}_{t,t_0} \cdot \mathbf{u}_0| \quad (3.34)$$

Assume an initial arbitrarily small sphere of perturbations of volume V within the phase space of an n -dimensional dynamical system. These are subject to an evolution according to the model equations. If the system is unstable with respect to any axis i , then the sphere will undergo growth along the axis i proportional to $e^{\gamma_i t}$. An n -dimensional system is thus characterized by n Lyapunov exponents. The total volume of the sphere $|\mathbf{V}|$ at time t will be given by $V e^{(\gamma_1 + \dots + \gamma_n)t}$.

By convention, the spectrum of exponents are aligned according to $\gamma_1 \geq \gamma_2 \geq \dots \geq \gamma_n$. Therefore, if a system is unstable, at least the first Lyapunov exponent is greater than zero, $\gamma_1 > 0$. Moreover, a Hamiltonian system in which the total energy is conserved is characterized by $\sum_{i=1}^n \gamma_i = 0$. For dissipative systems, the total sum of the exponents is negative, $\sum_{i=1}^n \gamma_i < 0$. In this case the total volume \mathbf{V} approaches zero volume, i.e. $\dim[\mathbf{V}] \leq n$ if the system is n -dimensional [Kalnay, 2003]. Note that the initial sphere \mathbf{V} in the phase space of an unstable dissipative system will evolve into an ellipsoid after a short time. The sphere grows or decays linearly along each of the n axes in the initial linear phase. For long times, the growth along the unstable axes continues, but the axes are bounded by the volume approaching zero. As a consequence, in the longer nonlinear time regime the axes of the ellipsoid fold in phase space, distorting into a banana shape. This occurs over and over again, so that after an infinite time, the original sphere has evolved into a structure of zero volume given by the strange attractor of the system. The dimension of this fractal structure has been estimated by Kaplan and Yorke [1979] to be

$$d = k + (\gamma_1 + \dots + \gamma_k)/|\gamma_{k+1}| \quad (3.35)$$

where the sum of the first k Lyapunov exponents is positive, and the sum of the first $k + 1$ exponents is negative [Kalnay, 2003].

In our experiments, the positive Lyapunov exponents described by the nature are given by $\gamma_{1,2}^{nat} = 0.9063, 0.3150$. The code for calculating the Lyapunov exponents was provided to the author by his supervisors. Autonomous dynamical systems have at least one exponent equal to zero [Kalnay, 2003], coinciding with our third exponent $\gamma_3 = 0$. This can be understood by considering the specific case of two separate initial conditions on the same model trajectory, that on average have the same distance from each other. The dissipation of the system is given by the sum of the nine Lyapunov exponents $\sum_{i=1}^9 \gamma_i^{nat} = -28.59$. This can be verified by calculating the divergence of the system [Kalnay, 2003], $\text{div } \mathbf{F} = -2(\sigma + 1 + b) - \tau(\sigma + 1 + b) = -28.5$; the small difference between both values is due to numerical error.

We have selected a sample of coupling configurations, i.e. $\{c, c_z\}$ -models and r -models, on the basis that these reproduce similar stability features given by the nature. This is

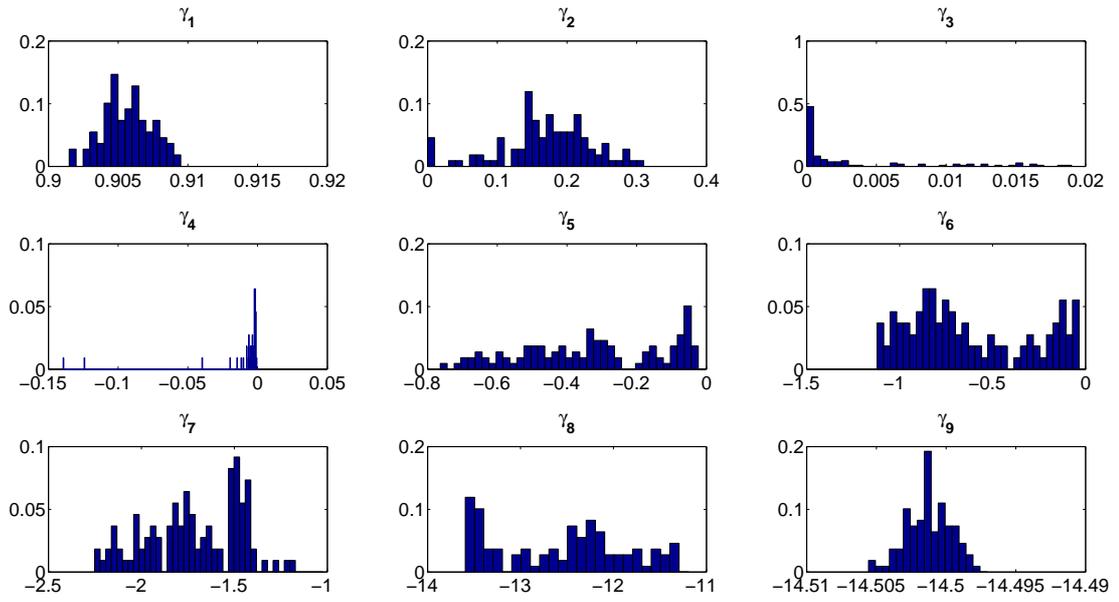
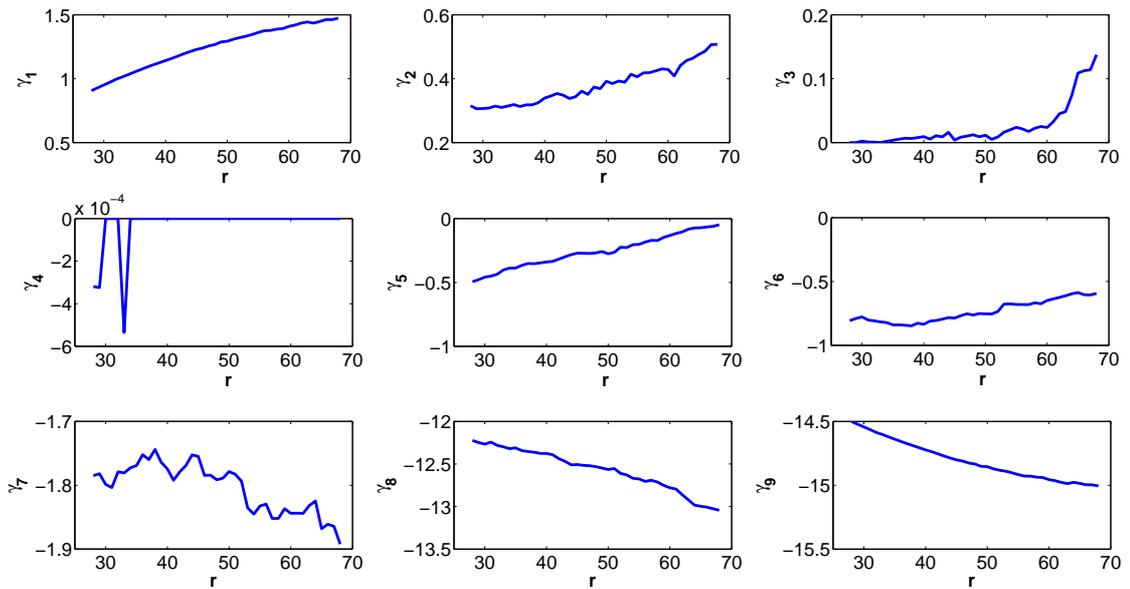
(a) $\{c, c_z\}$ -models: Distribution of Lyapunov exponents(b) r -models: Lyapunov exponents as a function of r

Figure 3.3: Characterisation of the configurations using Lyapunov exponents. (a) For $\{c, c_z\}$ -models: Distributions of the nine Lyapunov exponents. (b) For r -models: Lyapunov exponents as a function of the r .

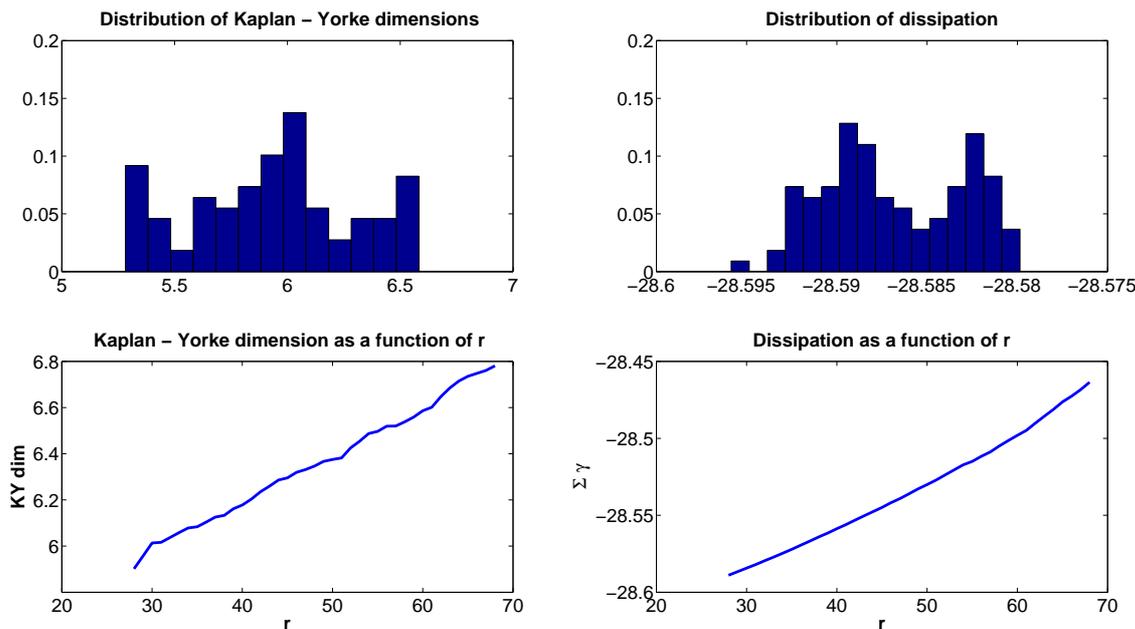


Figure 3.4: Distributions of the Kaplan-Yorke dimensions and the dissipation (top left and right) for $\{c, c_z\}$ -models. Kaplan-Yorke dimension and dissipation (bottom left and right) as a function of r .

based on the assumption that real climate models are imperfect, but reproduce the general behaviour of nature. The nine Lyapunov exponents for $\{c, c_z\}$ -models are given as distributions in Fig. 3.3(a). We can see that the first exponent varies only in the second number after the decimal, in contrast to the second exponent showing changes of above 50% for some configurations. This suggests that the first exponent represents the instability in the “extratropical atmosphere”, which is subject to a weak and correctly reproduced coupling parameter $c_e = 0.08$. The distribution of the second exponent is centered about a mean of $\overline{\gamma_2} \approx 0.2$, smaller than in the nature. We also observe that about half of the configurations have a third exponent equal to zero, and the rest have a positive, but small third exponent. Configurations that have a non-zero third exponent have a zero fourth exponent (not shown).

In order to highlight the trend as a function of the erroneous forcing for r -models, the values are displayed with respect to r , Fig. 3.3(b). We observe an increasing trend for the first two positive exponents with respect to the forcing, indicating a faster rate of divergence of initially close trajectories. We further observe that almost all models have a positive third exponent, and of significant value larger than 0.1 for very large r . The increased instability of the erroneously forced configurations signifies a decreased predictability.

Figure 3.4 displays the distributions of the Kaplan-Yorke dimensions and the total sum of the Lyapunov exponents (top left and right panels, respectively) for $\{c, c_z\}$ -models, as well as their dependence on the erroneous forcing for r -models (bottom panels). For $\{c, c_z\}$ -models, the attractor dimensions vary within the range of 5.3 – 6.6, and although the

dissipation rate appears bimodal, it varies only in the second number after the decimal. For r -models, both quantities show trends with respect to the forcing. The dissipation rate of the models decreases, the attractor dimensions increase.

3.8.1 Computation of Lyapunov exponents

In the following we describe the technique for numerically calculating Lyapunov exponents of chaotic flows given by [Benettin *et al.* \[1980\]](#). We first consider the first exponent γ_1 . Choose an arbitrary perturbation \mathbf{z}_0 and iterate it for a long time using the tangent linear model. Then, $\gamma_1 = \lim_{n \rightarrow \infty} \frac{1}{n} \ln(|\mathbf{z}_n|/|\mathbf{z}_0|)$ where n resembles the n th iteration. If $\gamma_1 > 0$, $|\mathbf{z}_n|$ typically becomes so large resulting in computer overflow. This is overcome by normalizing $|\mathbf{z}|$ periodically at times $\tau_i = i\tau$ for $i = 1, 2, \dots$ where τ is a fixed small time interval. The magnitudes α_i used for renormalization are stored in order to obtain the largest Lyapunov exponent:

$$\gamma_1 = \lim_{k \rightarrow \infty} \frac{1}{k\tau} \sum_{i=1}^k \ln \alpha_i \quad (3.36)$$

In practice, k is chosen sufficiently large after which convergence within an acceptable tolerance has been reached. How do we know that we have computed the first exponent, and not any other? First of all, the initial perturbation must have a component in the direction of the instability associated with the first exponent. Second of all, the rate of change associated with the first exponent is largest, so that for long integration times, the direction of the instability converges towards the Leading Lyapunov Vector.

The calculation of the remaining exponents is done as follows. Recall that the sum of the Lyapunov exponents describe a rate of change of a volume in phase space. Thus, by keeping track of the evolution of a volume in phase space through successive renormalization, one can infer the sum of the exponents. The second exponent is computed by evolving a parallelogram initially spanned by two independent arbitrary perturbations $|\mathbf{z}_0^1|$ and $|\mathbf{z}_0^2|$. The iterated vectors $|\mathbf{z}_n^1|$ and $|\mathbf{z}_n^2|$ span a parallelogram of area A_n . The area will be distorted in a way that $A_n \propto \exp[n(\gamma_1 + \gamma_2)]A_0$, so that we have

$$\gamma_1 + \gamma_2 = \lim_{n \rightarrow \infty} \frac{1}{n} \ln(A_n/A_0) \quad (3.37)$$

As with the calculation of the first exponent, the areas must be successively renormalized in order to avoid computer overflow. However, not only the magnitude, but the orientation of the iterated vectors is a problem for large times, as they become more coincident in the direction of dominant growth. Hence, the previous successive normalization procedure is generalized so that at each time τ_i the evolving pairs of vectors are replaced by two orthonormal vectors spanning the same two-dimensional subspace after Gram-

Schmidt orthogonalization. We then obtain

$$\gamma_1 + \gamma_2 = \lim_{k \rightarrow \infty} \frac{1}{k\tau} \sum_{i=1}^k \ln \alpha_i^{(2)} \quad (3.38)$$

where $\alpha_i^{(2)}$ is the paralelogram area before normalization at time τ_i . γ_2 can now be deduced from Eq. 3.38 and prior calculation of γ_1 using Eq. 3.36. The remaining exponents are computed in the same way by choosing the dimension of the initial volume accordingly. The calculation of any Lyapunov exponent can be generally expressed as

$$\gamma_l = \lim_{k \rightarrow \infty} \frac{1}{k\tau} \sum_{i=1}^k \ln(\alpha_i^{(l)} / \alpha_i^{(l-1)}) \quad (3.39)$$

See also Ott [2002].

4 Results

Our results are divided into two parts, according to our two main research lines. The relative performance of AI and FFI is studied with respect to the initialisation of different model compartments (Sect. 4.1.1), different observational error scenarios (Sect. 4.1.2), different model bias scenarios (Sect. 4.1.3), different drift scenarios (Sect. 4.1.4), and with respect to the differences between model and nature attractors (Sect. 4.1.5). The last three sections (Sect. 4.1.3-4.1.5) are related to different model parametric error configurations. Part two investigates the performances of LSI (Sect. 4.2.1) and EPU (Sect. 4.2.2). Either the RMSE or the root-mean-square skill score (RMSSS) are used to verify forecast skill. Both are debiased for fair comparison. Their definitions can be found in Eq. 3.28 and 3.29 respectively. We also quantify the “global” model bias using the *RMS Bias* (Eq. 3.30), and the drift using Eq. 3.32.

We refer to erroneously coupled / forced models as $\{c, c_z\}$ - / r -models (Sect. 3.6). Note that the “extratropical atmosphere” is hardly affected by the miscoupling (i.e. almost identical to the nature) for $\{c, c_z\}$ -models, but it differs from the nature for r -models. In order to simplify our analysis where necessary, we use example configurations that represent the general behaviour of the majority of the configurations. These are $\{c = 0.8, c_z = 0.9\}$ and $\{c = 0.3, c_z = 1.2\}$ for $\{c, c_z\}$ -models, corresponding to an average and a very large value of the *RMS Bias* respectively. The positive Lyapunov exponents for either configuration are $\gamma_{1,2} = 0.9036, 0.1895$ and $\gamma_{1,2,3} = 0.9032, 0.2162, 0.0153$, respectively. Compare these values to those of the nature, $\gamma_{1,2}^{nat} = 0.9063, 0.3150$. The second configuration has additional third positive exponent. Example r -models are given by $r = 34, 68$, corresponding to small and large *RMS Biases*, and are discussed in the text where necessary.

Unless specified otherwise, and apart from Sect. 4.1.2 in which the observational error scenario is varied, the observational error standard deviation is set to $\sigma^o = 2.5\%$. The results of Sect. 4.1.1, 4.1.2, 4.1.3 (in part), 4.2.1, and 4.2.2 can be found in [Carrassi *et al.*, 2014].

4.1 Research line 1: Relative performance of Full Field and Anomaly Initialisation

4.1.1 Initialisation performance with respect to different model compartments

We begin by assessing the performance of AI and FFI for the initialisation of single compartments. In Fig. 4.1, the RMSSS of AI and FFI is displayed as a function of lead time for both example configurations after a “monthly” averaging. Initialisation is carried out

either in the full system (black), the “extratropical atmosphere” alone (blue), the “tropical atmosphere” alone (magenta), or in the “ocean” alone (green). The first point is that by far the largest prediction skill is obtained when the system is fully initialised. For single compartments, it does not exceed 20% significantly, whereas for the full system the same level of skill is reached after the 20th “month” (FFI) or the about the 10th “month” (AI). Initialising only single compartments, the largest as well as the longest prediction skill is obtained for the “ocean”, with similar performance for both AI and FFI. For time horizons larger than 40 “months”, the skill scores for the fully initialised system and the system initialised with the ocean alone converge, illustrating how the system’s memory is efficiently stored in its slowest compartment. The total maximum forecast skill for long time horizons does not, however, exceed 5%.

When the “extratropical atmosphere” is initialised, we observe a distinct difference in the performance of AI and FFI. The longer skill after Full Field Initialisation suggests that the initial error is efficiently reduced. The “extratropics” represent the compartment with the largest instability and fastest error growth, and reducing the initial error helps delay the skillful forecast horizon. In the “tropical atmosphere” we do not observe a skillful performance after either AI or FFI, which suggests that the stronger coupling with an uninitialised ocean counteracts a reduction of the initial error.

Overall, the “ocean” component is indispensable in order to retain some forecast skill for time horizons longer than 40 “months”. On short time scales, only initialisation of the full system results in skill larger than 20%.

4.1.2 Initialisation performance in the face of observational error

One might guess that, as a general rule, a reduction of the observational error will lead to a better forecast. This is certainly true for the ideal case of a perfect model, in which forecast errors will in fact only come from imperfect initial conditions, i.e. a limited observational network and observational errors. A perfect model with perfect initial conditions will lead to a perfect forecast for all time. In practice, the forecast time horizon for which an improvement in the initial conditions of the prediction will lead to an improvement in forecast will depend on the growth rate of random errors, which is connected to the positive Lyapunov exponents of the model replicating nature (see Sect. 2.3).

In the case of Anomaly Initialisation applied in a “real” situation of an imperfect model, the bias is added onto the observations in order to approximate the model attractor. If the bias is significantly larger than the observational error, our original assumption that a reduction of the observational error will result in an improvement of the forecast might not hold. In Sect. 3.3.2 we derived that for $e^{bias} \gg \sigma^o$ as in the case of real applications, the mean analysis error of AI is less sensitive to the observational error in comparison with FFI.

Figure 4.2 shows the RMSSS of both algorithms as a function of the observational accuracy for six forecast horizons, full observation of the system, and the example model of $\{c = 0.8, c_z = 0.9\}$. We observe that the forecast skill after Full Field Initialisation deteriorates for large observational error scenarios, whereas Anomaly Initialisation is insensitive to the observational error. Take note of how FFI’s advantage over AI decreases for larger

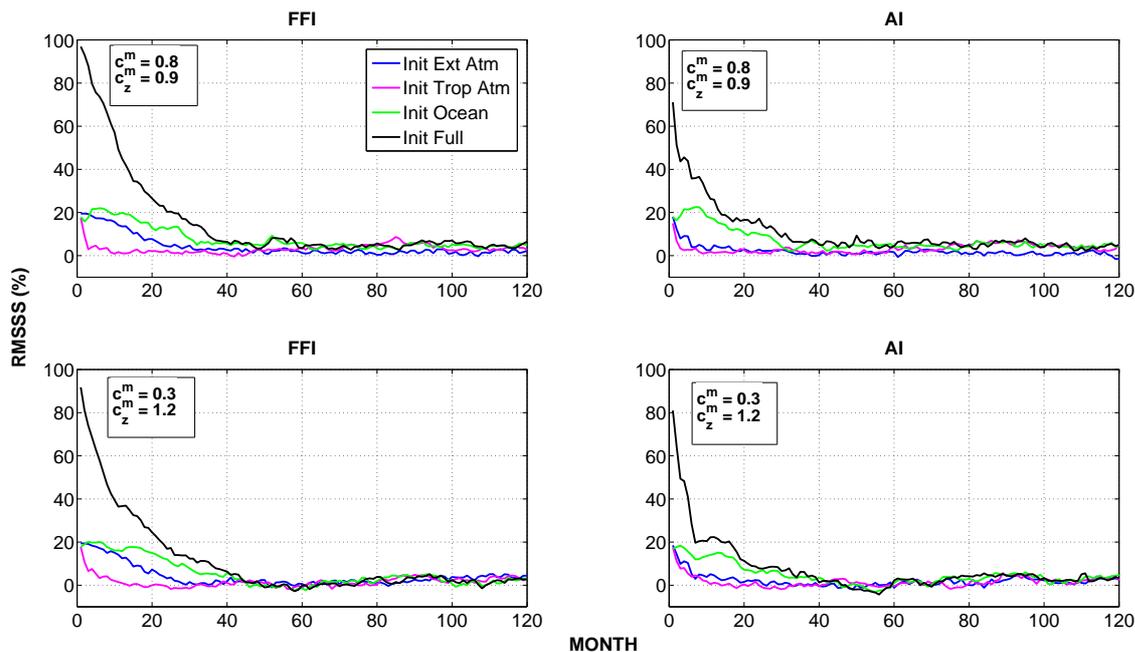


Figure 4.1: RMSSS as a function of the forecast lead time for FFI (left panels) and AI (right panels). Top / Bottom panels refer to the configurations $\{c = 0.8, c_z = 0.9\}$ / $\{c = 0.3, c_z = 1.2\}$ respectively. Different colors represent initialisation of the full system (black), “ocean” (green), “tropical atmosphere” (magenta) and “extratropical atmosphere” (blue).

forecast lead times, eventually yielding to AI for the longest forecast horizon. The initial rapid decrease of this advantage is due to fast loss of skill from the “atmospheric” compartments that are better initialised with FFI. The yielding to AI for the longest forecast horizon is due to a better performance of AI in the “ocean” in our example model, and is in line with the findings of [Toth and Peña \[2007\]](#), in which their simple mapping algorithm identical to AI results in slower error growth. For large observational errors, AI improves over FFI at an earlier forecast horizon, seen for the forecast horizon of 4-5 “years” in Fig. 4.2. The large observational errors result in a faster error growth rate for FFI and an earlier yielding to AI.

4.1.3 Initialisation performance in the face of model bias

In the following section we assess the performance of both initialisation algorithms as a function of the model bias. Introducing parametric error in a set of nonlinear coupled differential equations results in unforeseeable dynamical changes in the system. We have described in Sect. 3.8 the analysis of the stability properties and changes in attractor structure of the differently coupled / forced models discussed here.

Figure 4.3(a) shows the RMSSS for both algorithms as a function of the *RMS Bias* for

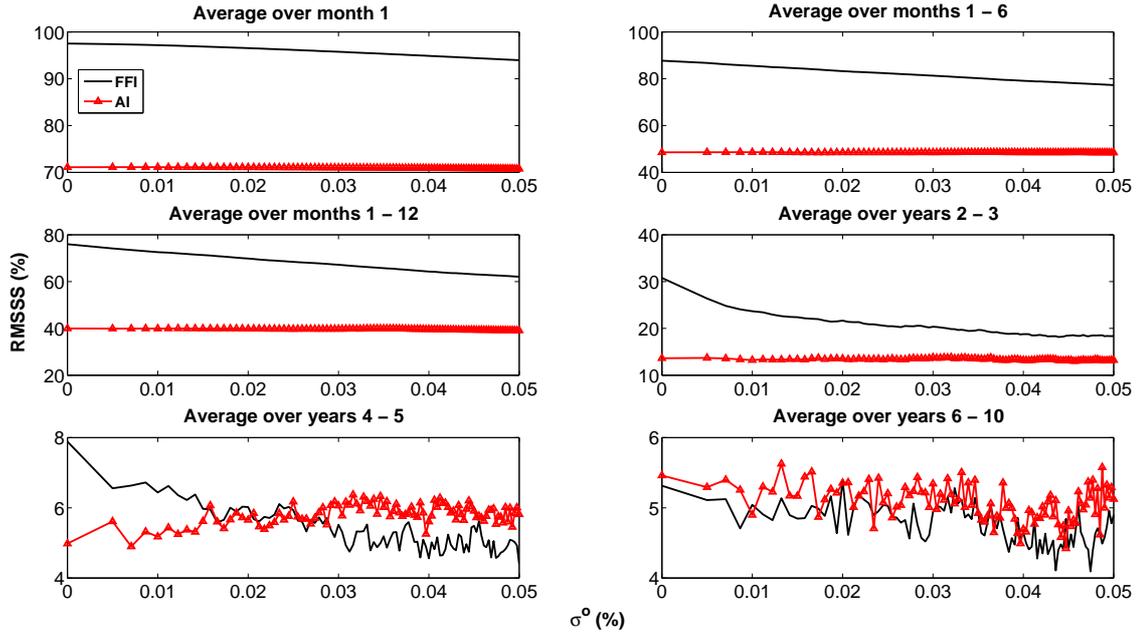


Figure 4.2: RMSSS as a function of the standard deviation of the observational error, σ^o , expressed as a fraction of the system’s natural variability. The model configuration corresponds to $\{c = 0.8, c_z = 0.9\}$. The six panels refer to six different averaging periods and are indicated in the corresponding labels. FFI (black line), AI (red line).

erroneously coupled models ($\{c, c_z\}$ -models) and for the same six forecast horizons as before. Overall, but especially in the initial stage of the forecast, the performance of FFI shows little relation to the nominal size of the model bias. The performance of AI, on the other hand, shows a clear dependency on the model bias within the first forecast “year”. Although for each model FFI performs better than AI for the first forecast “year”, both algorithms appear to converge after the second lead “year”. This points towards comparable skill in the “ocean”, which accounts for most of the forecast skill after the first forecast “year”. We cannot, however, confirm a systematic yielding of FFI to AI for long time horizons as we did previously in our example model in figure 4.2, even in scenarios with larger observational errors that would favor AI.

Figure 4.3(b) shows the RMSSS as a function of the *RMS Bias* for parametric model error in r , which we refer to as r -models. Here we find the opposite behaviour to what was observed for erroneous coupling parameters. AI outperforms FFI for short lead times and for a progressively larger *RMS Bias* associated with larger r . This improvement is marginally observed for the second and third forecast “years”, before converging to zero in the fourth and fifth “years”.

In Sect. 4.1.2, we observed that FFI improves after an improvement of the observational network. Here, with regards to the performance of AI, we observe two cases. The first in which AI improves only after improving the model (Fig. 4.3(a)). The second in which the performance of AI appears to be independent of the model deficiency (Fig. 4.3(b)).

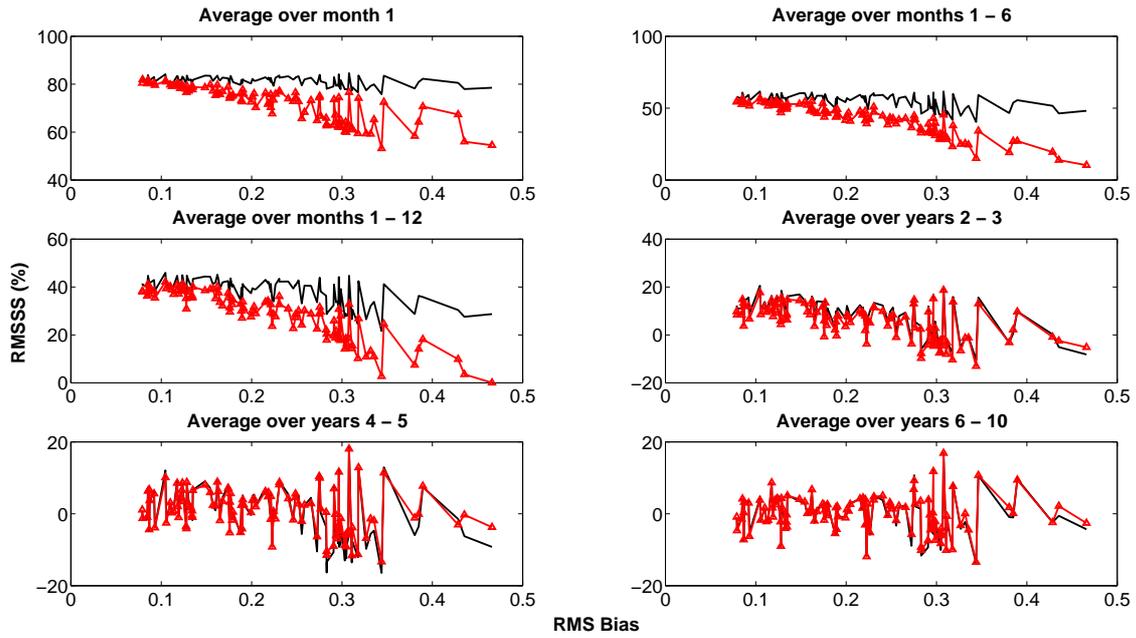
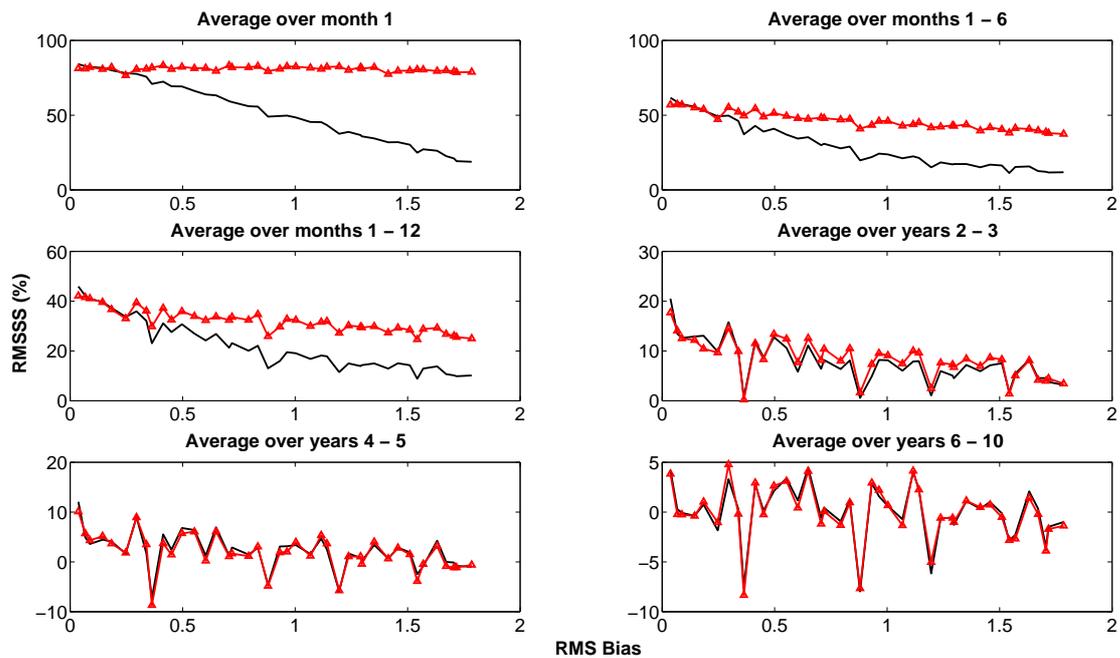
(a) RMSSS as a function of the RMS Bias for $\{c, c_z\}$ -models.(b) RMSSS as a function of the RMS Bias for r -models

Figure 4.3: RMSSS as a function of the RMS Bias for (a) $\{c, c_z\}$ -models and (b) r -models. The six panels refer to six different averaging periods and are indicated in the corresponding labels. FFI (black line), AI (red line).

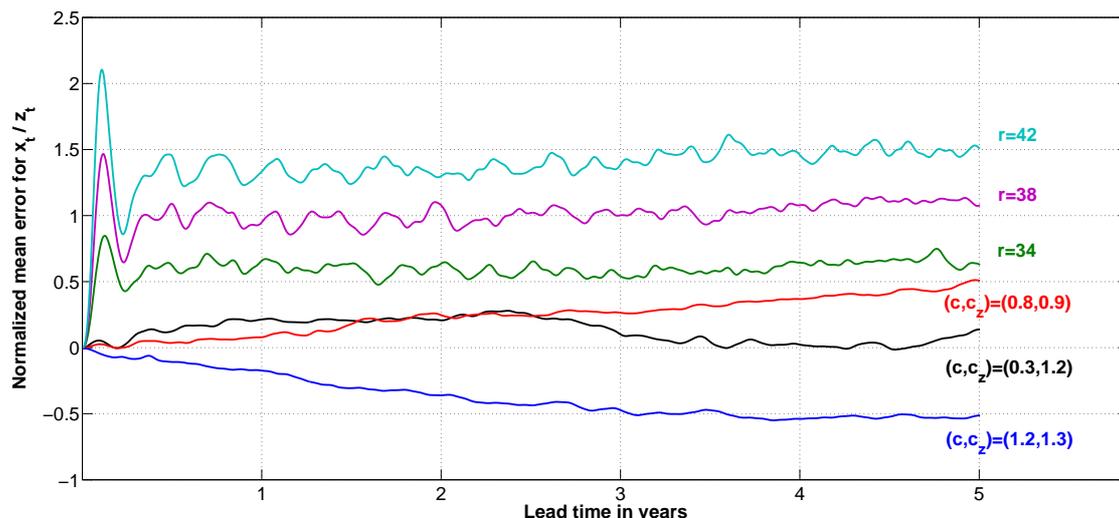


Figure 4.4: Normalized mean forecast error (i.e. normalized bias) of the x_t / z_t variables, for three $\{c, c_z\}$ -models / three r -models respectively, as a function of the lead time in “years”. The different colors correspond to the different models indicated in the plot.

The different behaviour of AI in Fig. 4.3(a) and 4.3(b) is connected to the different types of model error, and motivates the following analysis.

4.1.4 The role of the initial shock with regards to performance

Figure 4.4 illustrates, using specific $\{c, c_z\}$ - and r -models, how the normalized mean forecast error, i.e. normalized bias, evolves over the course of the predictions for the variable z_t in the r -models, and the variable x_t in the $\{c, c_z\}$ -models, which have been chosen here for their larger magnitude in comparison with the remaining variables. First of all, we observe a different drift behaviour, given by the slope of the bias shown, for different parametric errors. For $\{c, c_z\}$ -models, the drift is considerably slower, occurring over long time scales. For r -models, an increasingly large initial peak occurs for larger r , representing a rapid initial adjustment of the model to the initial conditions. This rapid initial adjustment is reminiscent of the initialisation shock that is often observed to follow initialisation [Magnusson *et al.*, 2012]. We also observe that the bias increases for larger r too.

We attempt to quantify the drift described in Eq. 3.32 in order to investigate a relationship between forecast skill and drift. The drift analysis interval T_d is set to one “month”. One might object that the interval of one “month” is too small to sample the drift evolution. On the other hand, take note of how within the first “month” the relative performance of AI / FFI is already determined for the remaining time horizons too, justifying this focus on the initial stage of the forecast. Furthermore, investigating the drift within the first forecast “month” allows us to concentrate on the time scale relevant for the initialisation shock observed above (Fig. 4.4) for the r -models.

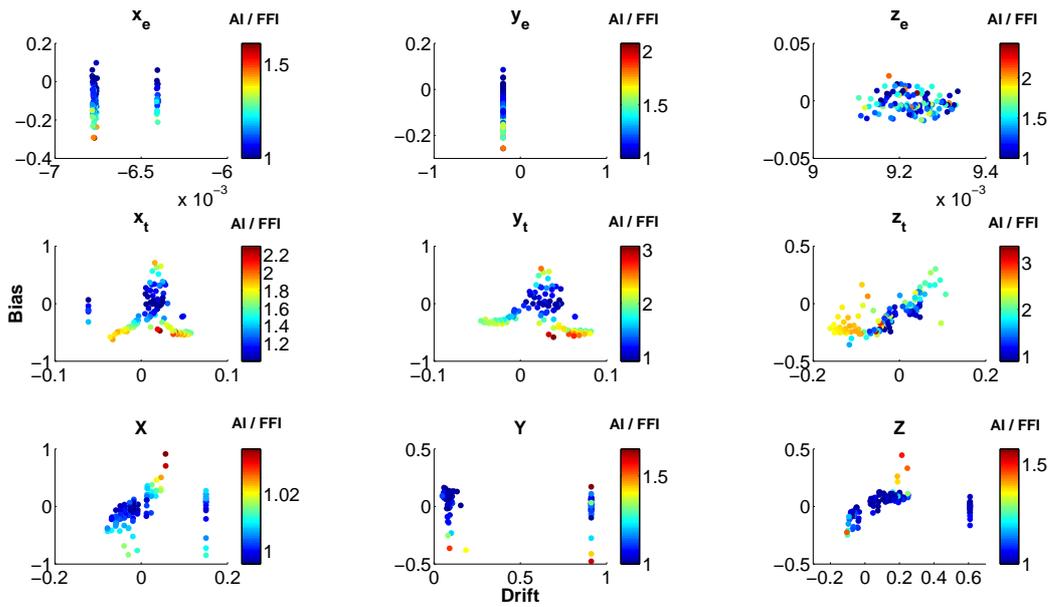
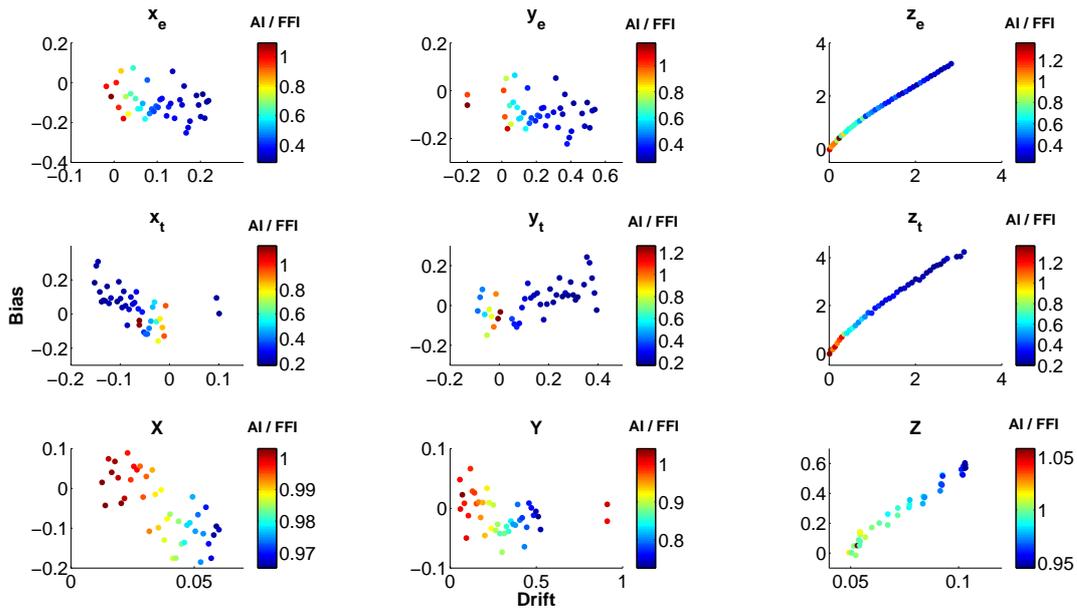
(a) RMSE ratio of AI / FFI (colorbar) for the 109 $\{c, c_z\}$ -models.(b) RMSE ratio of AI / FFI (colorbar) for the 41 r -models.

Figure 4.5: RMSE ratio of AI / FFI (colorbar) for all nine variables as a function of the bias on the y-axis and the initial drift on the x-axis. (a) Each coloured point in a single panel corresponds to one of the 109 $\{c, c_z\}$ -models. (b) Each coloured point in a single panel corresponds to one of the 41 r -models.

Figure 4.5(a) shows the RMSE ratio of AI to FFI for $\{c, c_z\}$ -models with regards to both the normalized bias, on the y-axis, and the initial drift, on the x-axis. Instead of using the *RMS Bias*, we plot the bias for each variable individually. Each coloured point within a panel corresponds to a single $\{c, c_z\}$ -model, making a total of 109 points in each of the nine panels. The colour blue relates performance more in favour of AI, and the colour red is more in favour of FFI. Take note of the different scales for each colour bar. Note as well that all ratios are above the value of one, establishing FFI's better performance over AI for all $\{c, c_z\}$ -models.

Figure 4.5(a) shows a dependence of the relative skill of AI and FFI on the model bias, as well as an independence on the initial drift. This is visible in the vertical colour gradient for each variable. For large positive or negative biases, the ratio is more in favour of FFI, corresponding to yellow/red colours at the “bottom” or “top” of each single scatter plot, and blue colours in the centre of the scatter plots around zero bias. No particular relationship can be observed with regards to the initial drift, with exception of the z_t -variable, in which a larger drift corresponds to ratios in favour of FFI.

Figure 4.5(b) is identical to the previous figure, but shows the RMSE ratio of AI to FFI for the 41 r -models. Note how the colour scales have shifted to values below one, corresponding to a better performance of AI as observed previously in figure 4.3(b). The most important feature we observe is a strikingly clear relationship between the initial drift and the relative performance of AI and FFI, evident in the horizontal colour gradient. With increasing positive or negative drift, the colour ratios change from red to blue, with only a few exceptions to this general trend.

We also observe that for r -models, the x - and y -variables have smaller biases, but a larger initial drift compared to those of the $\{c, c_z\}$ -models. On the other hand, the z -variables show very large biases, making it apparent that the *RMS Bias* for r -models in figure 4.3(b) is due mostly to the large biases in the z -variables. One might infer that the large bias in the z -variables induces an initialisation shock that is manifest in a large initial drift also in the remaining variables. If we calculate the initial drift in the same way but over a three “month” period, we find that for the “atmospheric” x - and y -variables the large initial drift even changes sign (not shown), indicating an erratic behaviour that further supports the hypothesis of a dynamical shock mechanism. The drift of the z -variables is exemplified in Fig. 4.4, showing an initial monotonic “surge” towards a peak larger than the mean, followed by a compensatory oscillation.

We have so far observed that Anomaly Initialisation fails for the erroneously coupled models, but succeeds for the erroneously forced models (Fig. 4.3). The analysis of the initial drift has indicated that an initialisation shock occurs for r -models, which AI successfully counter-balances. In the following section we look for further evidence by visualising the respective model and nature attractors.

4.1.5 The role of the attractor

In order to investigate our hypothesis that AI outperforms FFI in the face of an initialisation shock, it helps to visualise the model and nature attractors and observe the predicted trajectories after initialisation. We have seen that the initial shock occurs for errors in the

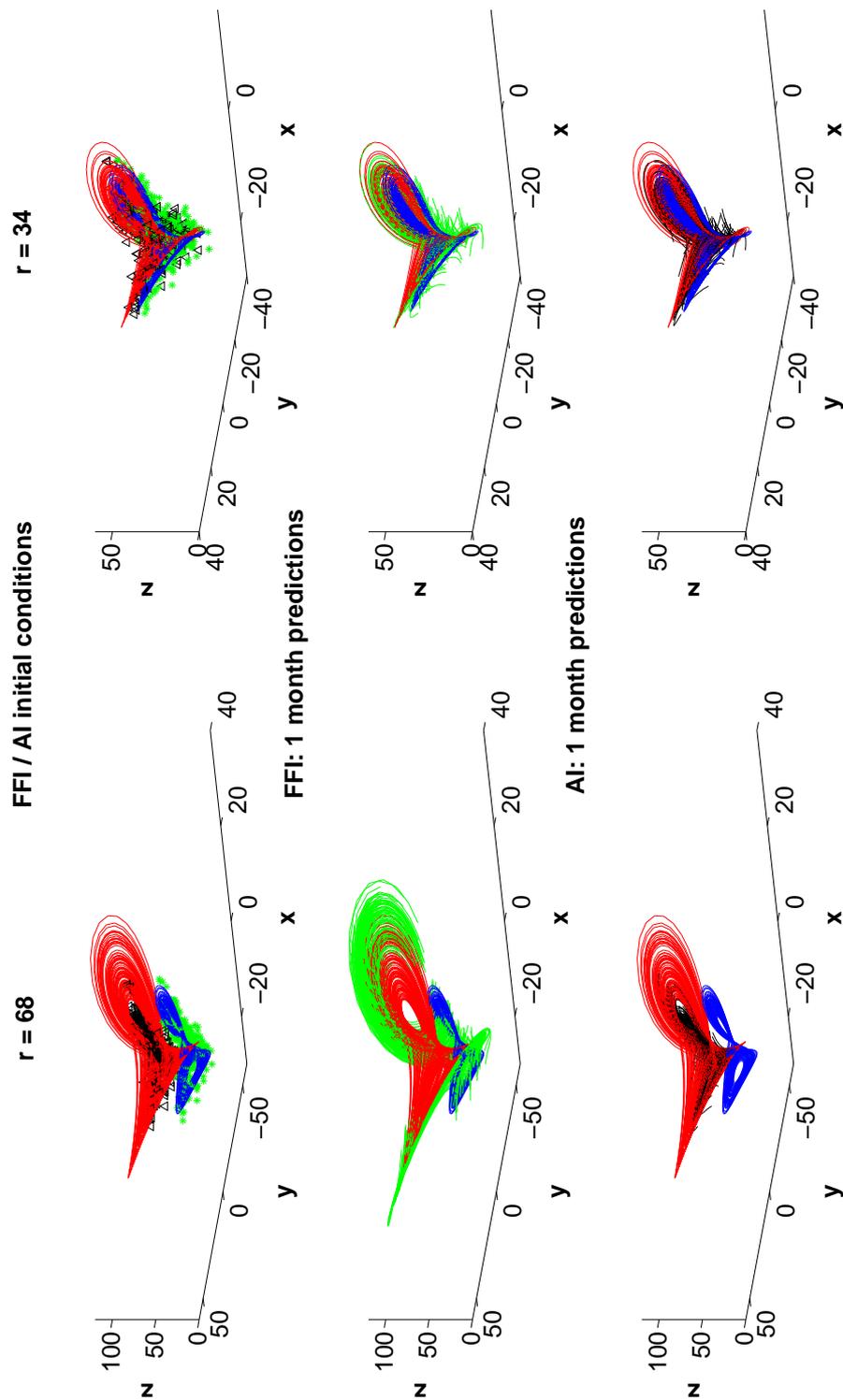


Figure 4.6: Three-dimensional view of the “extratropical” attractors of the model (red) and nature (blue) for $r = 68$ (left panels) and $r = 34$ (right panels). Top panels show the initial conditions of AI (black triangles) and FFI (green stars). Middle panels show the 1-“month” predictions initialised with FFI. Bottom panels show the 1-“month” predictions initialised with AI.

forcing, which is why we first focus on r -models. The top left panel of figure 4.6 displays the 40-“year” *climatological* trajectory (i.e. model control run) of the “extratropical atmosphere” of an example r -model in red ($r = 68$), along with the simulated nature in blue. The case $r = 68$ shows the largest bias among all the r -models considered. The green stars depict the monthly sampled observations that comprise the 360 initial conditions for FFI; the black triangles depict the initial conditions for AI.

First of all, we observe similarity in shape between the model and nature attractors. The model attractor differs in a larger variance accounting for its larger size with respect to the nature, as well as a “translation” along the z -axis in phase space that comprises its bias. We have depicted the “extratropics” for easy visualisation, but the behaviour is qualitatively the similar for the “tropical atmosphere” and “ocean”. The figure illustrates how Anomaly Initialisation constitutes a “translation” of the set of observations closer to the model attractor.

The middle and bottom left panels of figure 4.6 show the one “month” long predictions for FFI and AI respectively starting from the initial conditions displayed in the top left panel. Clearly, initialisation further away from the model attractor in the case of FFI results in the systematic favoring of trajectories along or even outside of the “wings” of the attractor. The outer wings of the attractor can be considered as the model’s more “extreme climate”, for the reason that it is both largest in amplitude, and rare. Within the 40-“year” climatological model trajectory, this region of the attractor has only been visited once or not at all. This serves as an extreme example of dynamical forecast errors resulting from drift / model error. In particular, the “overshooting” of the mean error of the z -variable observed in figure 4.4 is a one-dimensional manifestation of the model’s dynamical compensatory response to an initialisation far outside the attractor. After Anomaly Initialisation (bottom left), the predictions are clearly much more in line with the model climate, and as a consequence more in line with the natural climate after a bias correction too.

The case in which model and nature attractors are as “far away” from each other as they appear for $r = 68$ certainly seems unrealistic. Nevertheless, the right panels of figure 4.6 depict analog behaviour for a model with $r = 34$, which corresponds to a distinguished, but smaller model bias, and for which AI still outperforms FFI. The evolution of the mean error of the z -variable of this model has been shown previously in figure 4.4 too.

Making clear inferences from visualising attractors of the $\{c, c_z\}$ -models and their predictions is more complicated. The individual attractor structures differ significantly from one another, and from nature. We are also restricted to the “tropical atmosphere” or “ocean”, because the “extratropics” are unaffected by the errors in the coupling.

The top panel of Fig. 4.7 displays a single case for the example model $\{c, c_z\} = \{0.8, 0.9\}$ and nature attractors of the “tropical atmosphere” in red and blue respectively, along with the initial conditions for FFI as green stars and for AI as black triangles. The perspective has been altered for better visualisation (compare with Fig. 3.1). We observe a subtle, but important difference between the model and nature attractors. One can identify two separate “regimes” for the nature, where one is more pronounced than the other. The model reproduces a similar structure, giving more weight however to the regime that is less pronounced in the nature. This important, but complicated change in attractor structure of the model can not be compensated by a translation of the set of observations as

AI intends. As a consequence, the result after initialisation with AI is less consistent with the model attractor. The bottom panel of Fig. 4.7 displays the one “month” predictions, showing how the weaker regime is “over-predicted” after Anomaly Initialisation.

Results show that Anomaly Initialisation has been designed to remedy drift by translating the set of observations closer to the attractor. The precondition for this approach to be successful, however, is that the difference in the model and nature PDFs is mainly of the first order. This might be the explanation for why AI performs poorer than FFI for the $\{c, c_z\}$ -spectrum of models, in which their PDFs differ in higher orders than only the first order, i.e. the bias.

4.1.6 Approximation of the model PDF

The objective of Anomaly Initialisation is to better approximate the model attractor. This is intended by adding the bias to the observations distribution to obtain a better representation of the model distribution. Figure 4.8(a) illustrates the probability density functions (PDFs) for each variable for the nature (blue), and for our example model (red), which is chosen here again to be $\{c, c_z\} = \{0.8, 0.9\}$. Additionally, the distribution of the differences between model and nature PDFs is given (green), as well as the nature distribution “corrected” by means of an addition of the bias of the variable (black), from which the initial conditions for Anomaly Initialisation are sampled. Comparing the nature and model distributions, we observe that in the “extratropical atmosphere” they are almost identical, which is expected due to the equal, weak “extratropical” coupling in the model as in the nature. In the “tropical atmosphere”, the nature PDF is given by a bimodal distribution in the x_t - and y_t -variables, for which one of the modes is more distinguished than the other. The model PDF can be described in terms of a bimodal distribution over the same region in phase space, but over-representing the mode that is weaker in the nature. Such a misrepresentation of the natural modes due to an erroneous coupling in the model results in a clear bias. The initial conditions for Anomaly Initialisation are sampled from the distribution given by the addition of this bias onto the nature (black). However, the distribution gained from adding the bias onto the nature does not approximate the model PDF, as it is intended to do. Thus, the initial conditions for AI are sampled from a distribution that much less represents the model PDF than the initial conditions of FFI sampled from the nature distribution. This is due to the fact that, although a difference in the first order moments is present, and this difference constitutes the bias, the addition of this bias onto the nature distribution does not result in an adequate correction, because the higher order differences are not accounted for. In the “ocean”, the distributions from which FFI and AI are sampled (blue and black respectively) are almost entirely equal, explaining the similar performance of AI and FFI in the “ocean” as we have observed for this example model. Figure 4.8(b) depicts the same distributions described in figure 4.8(a), but for an example model with erroneous forcing of $r = 68$. We can see that the distributions from which FFI and AI are sampled (blue and black respectively) are very similar for all x - and y -variables. The key difference can be observed in the z -variables, for which a bias of the model distribution with respect to the nature distribution is present. The distribution from

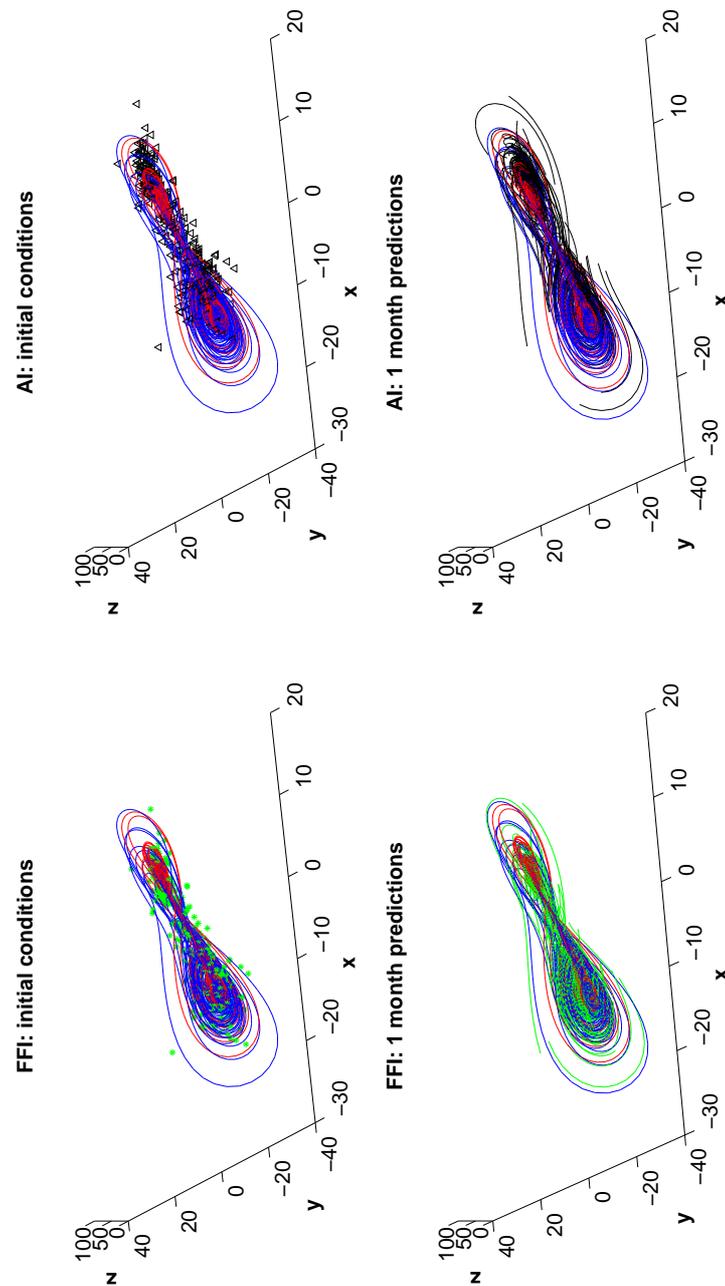


Figure 4.7: Three-dimensional view of the “tropical” attractors of the model (red) and nature (blue) for $\{c, c_z\} = \{0.8, 0.9\}$. The initial conditions of FFI (green stars) and AI (black triangles) are shown in the top left and right panels respectively. The one “month” predictions are shown in the bottom panels after FFI (left) and AI (right). Note that the perspective has been altered with respect to Fig. 4.6 for better recognition.

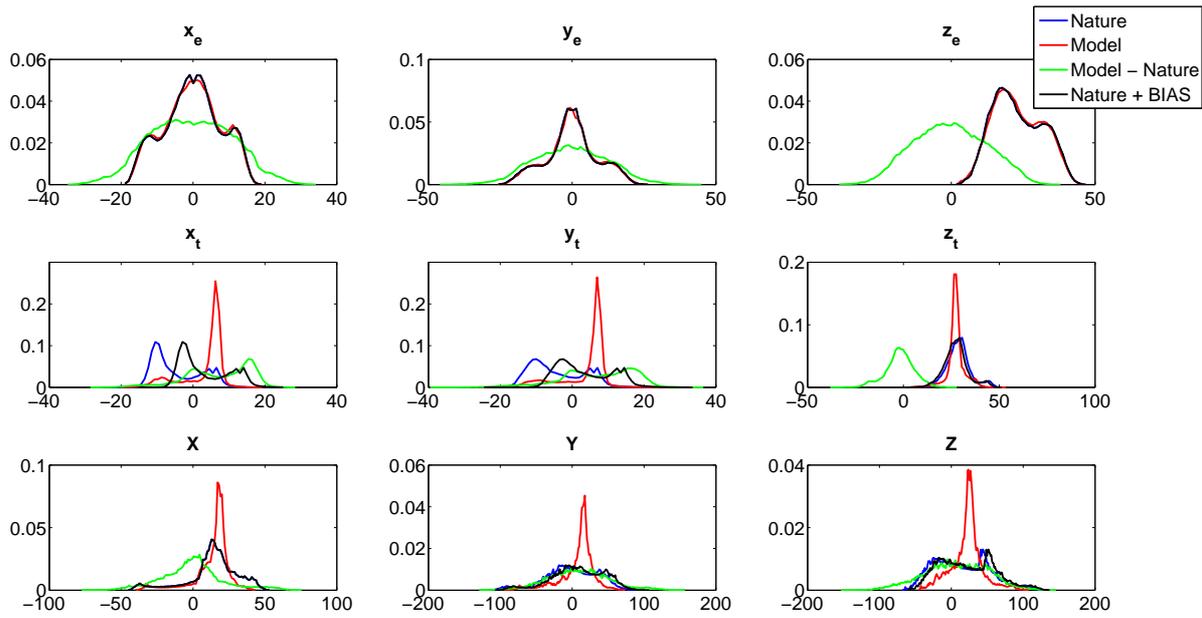
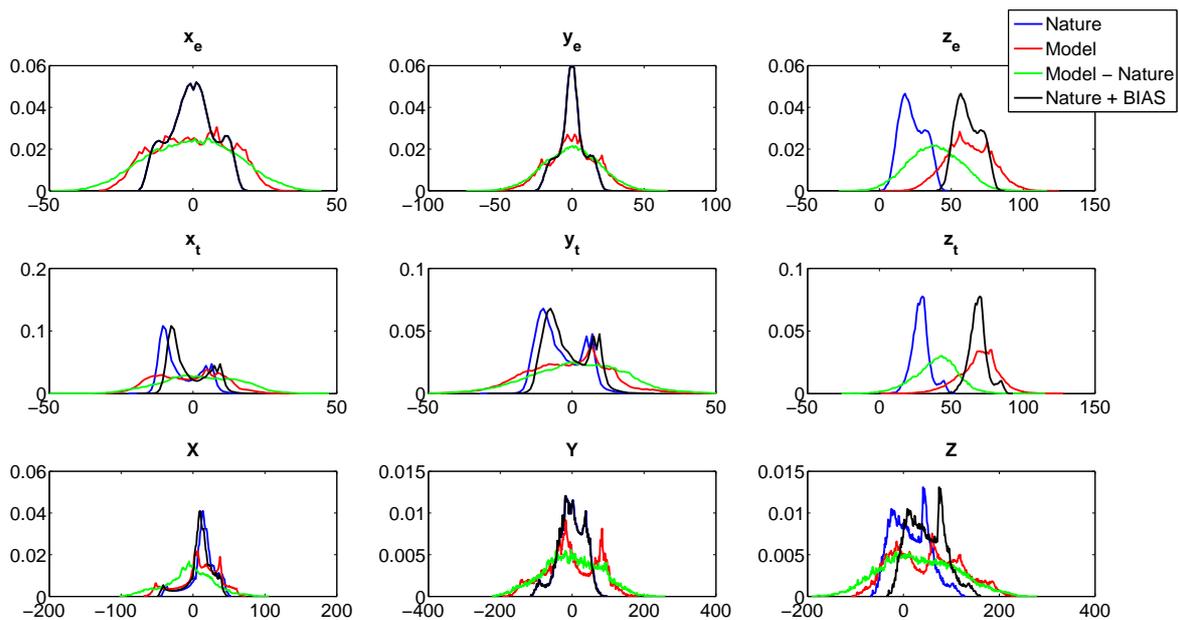
(a) $\{c_x, c_z\} = \{0.8, 0.9\}$ (b) $r = 68$

Figure 4.8: PDFs of the nature (blue), model (red), the difference between model and nature (green), and the distribution associated with the initial conditions for AI (black). The nine panels correspond to the nine variables as indicated. (a) $\{c_x, c_z\} = \{0.8, 0.9\}$, (b) $r = 68$.

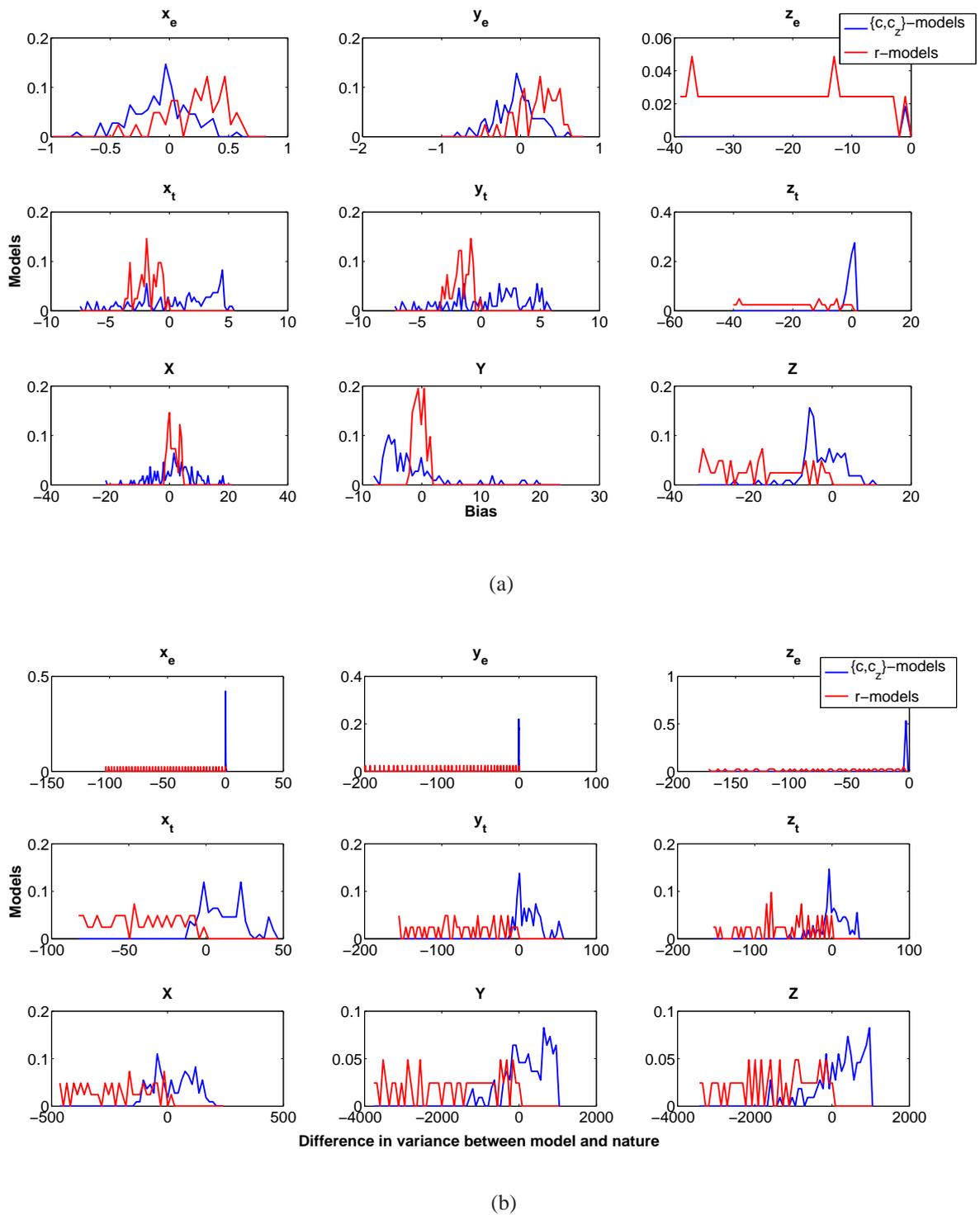
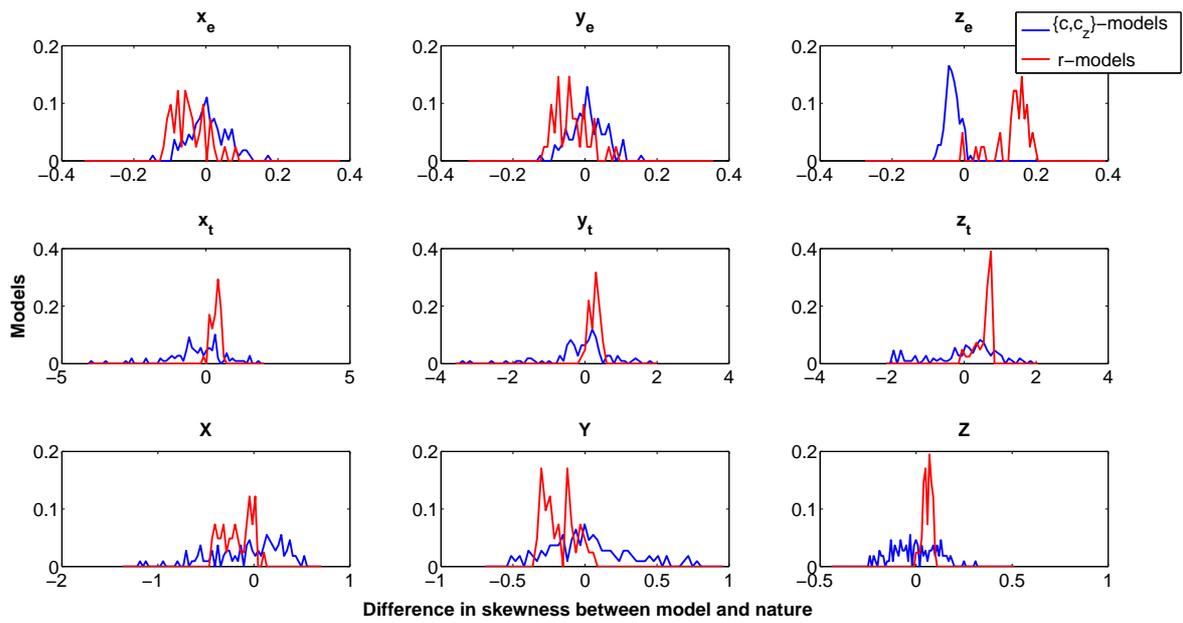
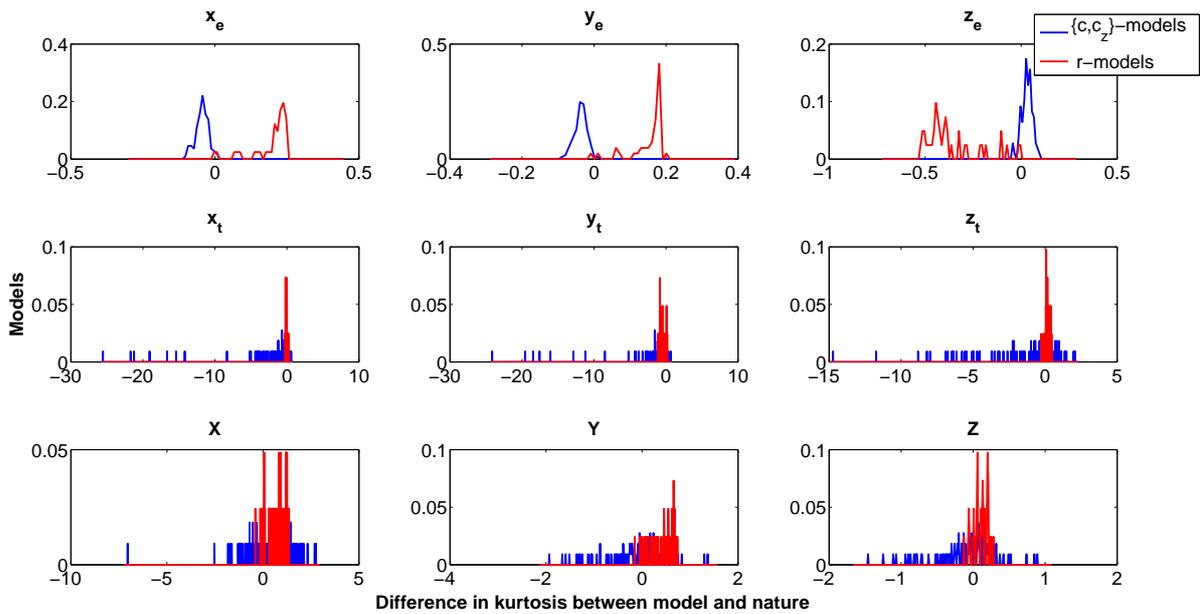


Figure 4.9: Distributions of (a) the biases and (b) the differences in second order moments for $\{c, c_z\}$ -models (blue) and r -models (red). Panels correspond to the nine variables as indicated.



(a)



(b)

Figure 4.10: Difference in (a) third order moments and (b) fourth order moments between model and nature plotted as a distribution for $\{c, c_z\}$ -models (blue) and r -models (red). Panels correspond to the nine variables as indicated.

which the initial conditions for AI are sampled (black) approximates the model PDF (red) much better than the nature distribution (blue) from which the initial conditions of FFI are sampled. This is due to the fact that the differences between the model and nature PDFs are well-approximated in terms of a difference in both first order moments. As opposed to our previous $\{c, c_z\} = \{0.8, 0.9\}$ -model, for this model the addition of the bias onto the nature distribution can correct the difference in the first order moments, resulting in a better approximation of the model PDF. This feature is much less distinguished in the “ocean”, explaining why the better performance of AI over FFI for r -models is less pronounced here.

In order to extend the analysis of the PDFs for all models, figure 4.9(a) displays the distributions of the bias for $\{c, c_z\}$ -models (blue) and r -models (red) for all nine variables. Figures 4.9(b) to 4.10(b) follow in the same fashion, displaying the distributions of the differences between the higher order moments of model and nature PDFs up to the fourth order. Figure 4.9(a) shows how the bias is comprised almost entirely in the z -variables for r -models, and more broadly distributed over the variables for $\{c, c_z\}$ -models. Figure 4.9(b) shows how the variance of all variables of the r -models is very broadly distributed, owing to the fact that the variance increases for an increase in forcing. Figures 4.10(a) and 4.10(b) show how the differences in skewness and kurtosis are much more broadly distributed for $\{c, c_z\}$ -models rather than r -models. These results clearly confirm the findings of our example models in the previous figures 4.8(a) and 4.8(b). The differences between model and nature PDFs for $\{c, c_z\}$ -models occur not only in the first moment of the distributions, but also in the third and fourth moments. Contrarily, for r -models the differences between model and nature PDFs are generally well-approximated in the first order, explaining why Anomaly Initialisation is a promising alternative to Full Field Initialisation. The fact that larger differences occur in the second moment for r -models do not turn out to be a disadvantage, because the larger variances of the models increase the probability that the initial conditions after Anomaly Initialisation will project onto the model attractor.

It appears as if Anomaly Initialisation will work better when the projection of its initial conditions onto the model PDF is successful. Similarity of probability distributions can be measured in terms of the Bhattacharyya coefficient [Bhattacharyya, 1943]. The Bhattacharyya coefficient is an approximate measurement of the amount of overlap between two statistical samples.

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)} \quad (4.1)$$

where p and q are discrete probability distributions over the same domain X . Therefore, $0 \leq BC \leq 1$, where 1 corresponds to maximum similarity. Note that if the distributions do not overlap, the BC coefficient does not distinguish between distributions “far” or “very far” from each other; both cases result in $BC = 0$. We want to measure the similarity between the model PDFs ($p(x)$) and the “corrected” distributions from which the initial conditions of AI are sampled ($q(x)$).

Figure 4.11(a) displays the Bhattacharyya coefficient as a function of the bias for $\{c, c_z\}$ -

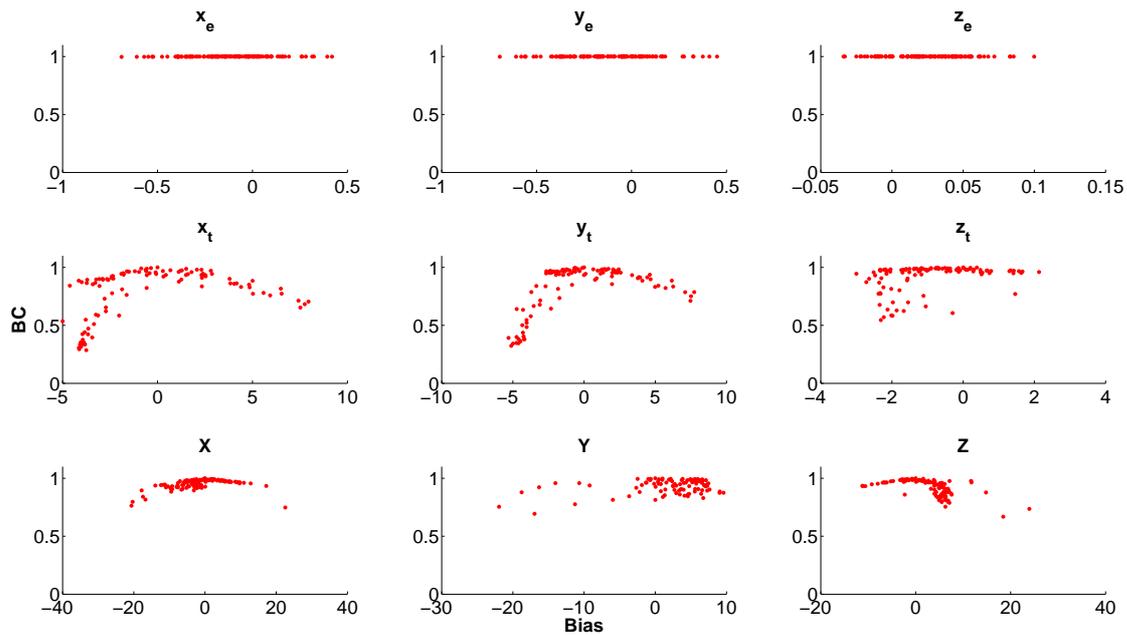
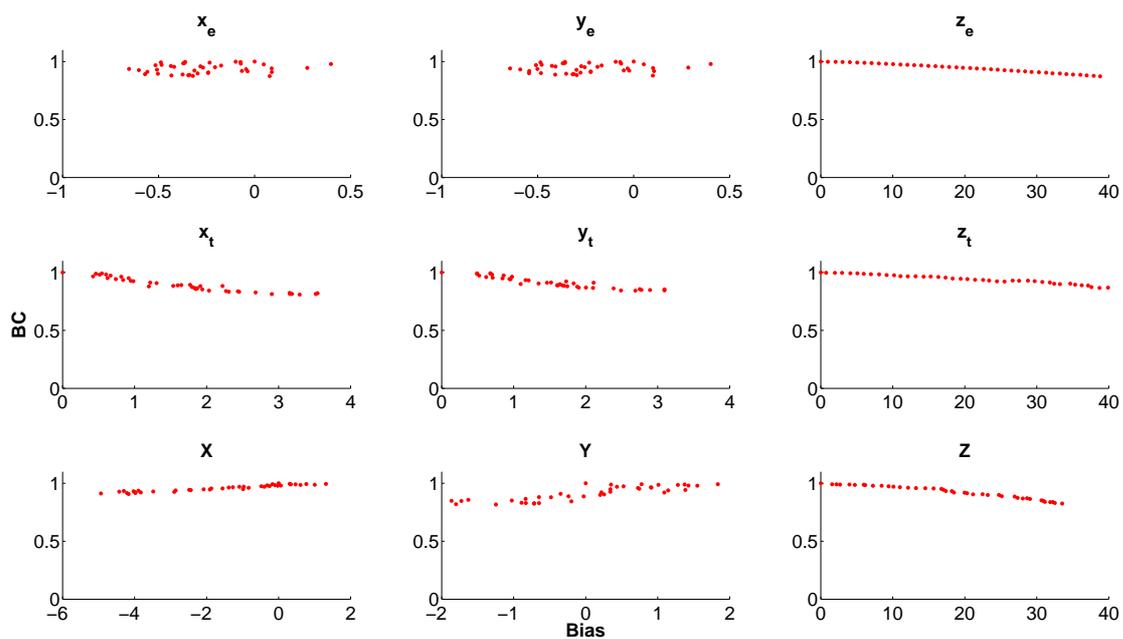
(a) Bhattacharyya coefficient as a function of the bias for $\{c, c_z\}$ -models.(b) Bhattacharyya coefficient as a function of the bias for r -models.

Figure 4.11: Bhattacharyya coefficient measuring the similarity between the model distribution and the distribution associated with the initial conditions for AI, as a function of the bias. Each point in a single panel corresponds to one of (a) the 109 $\{c, c_z\}$ -models (b) the 41 r -models. Panels correspond to the nine variables as indicated.

models, where every point in each of the nine panels corresponds to a single model. In the “extratropical atmosphere”, the coefficient is equal to one for all models, because both distributions are almost entirely equal due to the equal “extratropical” coupling. The remaining variables show a dependency of the coefficient on the model bias. The larger the bias, the less similar the model PDF is compared to the distribution from which the initial conditions for AI are sampled. This shows that the first order bias correction behind Anomaly Initialisation does not result in a good approximation of the model PDF as intended.

Figure 4.11(b) shows the same, but for r -models. The Bhattacharyya coefficient is much less dependent on the bias of the model, clearly pointing towards the fact that the model PDFs are well-approximated by the “corrected” distributions. This result confirms the discussion followed in Fig. 4.9(a)-4.10(b). Thus, Anomaly Initialisation succeeds for r -models, because the first order approximation scheme of the model attractor through addition of the bias succeeds. This success is due to the fact that the model attractors vary from the nature attractor in terms of the first order, and higher order differences are negligible.

4.1.7 Summary of research line 1

With regards to our first research line, we have found that for the initialisation of single compartments, the best performance for long time horizons is obtained when the stabler component of the model, i.e. the “ocean”, is initialised. Having the slowest error growth rate, the “ocean” behaves like the system’s memory. We have also found that FFI depends sensitively on the observational error, whereas AI is not significantly improved when the observational error is reduced. This is not surprising, because the initial conditions error of FFI is reduced for smaller observational errors, but under the addition of the model bias as in AI, the initial error scales less with regards to the observational error (Sect. 3.3.2).

The results investigating the role of the model on the initialisation suggest the following interpretation. AI outperforms FFI for r -models, because a significant model bias is present, and other relevant differences between model and nature PDFs are limited to the second moments. For FFI, the model bias incurs initialisation outside of the model attractor, to which the model’s short-term response is a climate extreme, and resulting oscillations. This is a picture book example of shock-induced forecast errors, and is supported by experiments in which the “perfect model” is perturbed by introducing an artificial bias onto the observations, indicating a direct relationship between drift and perturbation magnitude (not shown).

On the other hand, FFI outperforms AI for $\{c, c_z\}$ -models, because such models define a spectrum of attractors that differ significantly from nature in their third and fourth moments. Although these models have equally significant biases, the biases do not alone account for the differences between model and nature PDFs.

These results are entirely in line with the principle behind AI, and what one would expect of its performance. The results suggest that a comparison of model and nature PDFs can inform the choice between AI and FFI. AI will be the initialisation scheme of choice for models with significant biases that are not accompanied by equally significant higher

order differences in PDFs. AI intends to control and eliminate forecast errors occurring from a “translation” of the model attractor with respect to the nature attractor in phase space.

4.2 Research line 2: Advanced schemes

4.2.1 Least Squares Initialisation (LSI)

In the following section we will compare the performance of FF-LSI against the standard FFI approach for the same example configuration of $\{c = 0.8, c_z = 0.9\}$. The standard deviation of the observational error is equal to $\sigma^o = 1.5\%$, and the error covariance \mathbf{R} is diagonal and contains the correct observational error variance. The background error covariance matrix \mathbf{B}^m is estimated over 50 years of an uninitialised model run.

Figure 4.12 shows the RMSSS of FF-LSI for different observational scenarios (full coloured lines) as a function of the background error scaling coefficient $10^{-5} \geq \alpha \leq 15$. The six panels correspond to the different forecast time horizons as indicated, and the dashed lines correspond to the standard FFI performance. We observe that for small α , the forecast skills of all scenarios converge to zero. A small α suggests a small background error covariance \mathbf{B}^m , implying a high accuracy of the background information. This leads to an over-emphasizing of the background, and a neglect of the observational information, resulting in essentially uninitialised predictions. For large α , we observe an improvement in forecast skill of FF-LSI over standard FFI for partly initialised systems. In this case, the observational information is fully taken advantage of where available, because a large background error variance assimilates the observations as if they were perfect. The observational content limited to the observational subspace is propagated to the entire model domain according to the spatial correlations embedded in the off-diagonal elements of \mathbf{B}^m , thus exceeding the standard FFI skill level. This behaviour is underlined when viewing the full system (black line); the forecast skill of FF-LSI converges towards that of standard FFI. When the full system is being observed and all observations are utilized, no additional information can be obtained through the background error structure. Note also that the skill improvement for the “extratropical atmosphere” is only marginal, due to its weak coupling with the “ENSO” system inducing small covariances.

In order to compare the similarity of the real background error covariance \mathbf{B} and its approximation \mathbf{B}^m , figure 4.13 displays the percentage of the explained variance of each of the eigenvectors of \mathbf{B} (full line) and \mathbf{B}^m (dotted line) in the upper panel, as well as the scalar product of each pair $\mathbf{v}_i(\mathbf{B}) \times \mathbf{v}_i(\mathbf{B}^m)$ in the lower panel, with $\mathbf{v}_i(\mathbf{B})$ being the i th eigenvector of \mathbf{B} , $i = 1, \dots, 9$ and \times indicating the scalar product. The explained variance is given by $Var^{exp}(\mathbf{v}_i) = \lambda_i / (\sum_i \lambda_i) \times 100[\%]$, with λ_i being the i th eigenvalue. We see that the variance distribution is accurately reproduced and that, with exception of the third and fourth that are perpendicular to each other, all remaining eigenvectors are fully aligned.

4.2.2 Exploring the Parameter Uncertainty (EPU)

In the following section we will study the performance of EPU, a drift correction technique introduced in Sect. 3.4.2. EPU is implemented during the forecast run after initialisation with FFI, and will be referred to as FFI-EPU. We compare its performance against standard FFI without a correction procedure.

Figure 4.14 displays the distribution of the RMSE of all 109 $\{c, c_z\}$ -models after Full Field

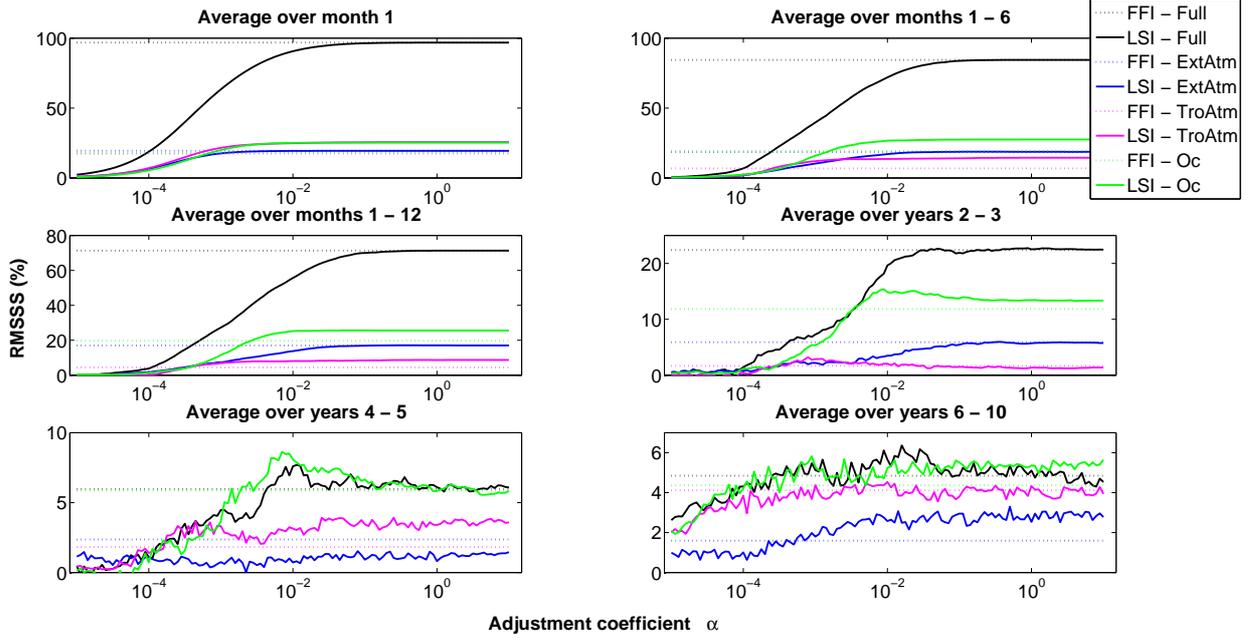


Figure 4.12: RMSSS as a function of the tuning coefficient α for LSI. The panels refer to six different averaging periods as indicated. The colors represent initialization of the full system (black), the “ocean” alone (green), the “tropical atmosphere” alone (pink), and the “extratropical atmosphere” alone (blue). The values of the standard FFI are displayed for reference (dots).

Initialisation with (red lines) and without (black lines) the implementation of EPU. The specifications for EPU include the parameter range $\Delta\Lambda$ given by the 109 model configurations, as well as the correction time interval ΔT_{Bias} equal to one time step. The panels correspond to the same six forecast horizons as earlier. Additionally, the means of both distributions are given in the insets. Results reveal the benefit of implementing EPU up to the third forecast year, with minor improvement for longer horizons. This is evident in the shift of the distributions related to FFI-EPU towards smaller RMSE values, as well as a reduction of the mean error by 17%, 10%, 8%, 4%, 0.6% and 0.2% for the six forecast horizons respectively. Significant skill is limited to the seasonal forecast, which is due to sampling errors, as well as the progressive deviation of the drift behaviour from the linear assumption, on which EPU is based. Note, however, that the use of EPU in conjunction with FFI has implied only a minor increase in computational cost.

In figure 4.15 we investigate the impact of the uncertainty about the width of the sampling interval for parametric error $\Delta\Lambda$, as well as the correction time interval ΔT_{Bias} , on the performance of EPU. Both of these factors control the setup of the drift correction technique. In the left panel, the RMSE of FFI-EPU (full lines) is displayed as a function of the scaling coefficient β , which scales the width of the sampling interval, i.e. $\delta\lambda_i \in U(0, \beta(\lambda - \lambda^{nat}))$, simulating the level of uncertainty about the possible parameter range. The RMSE is

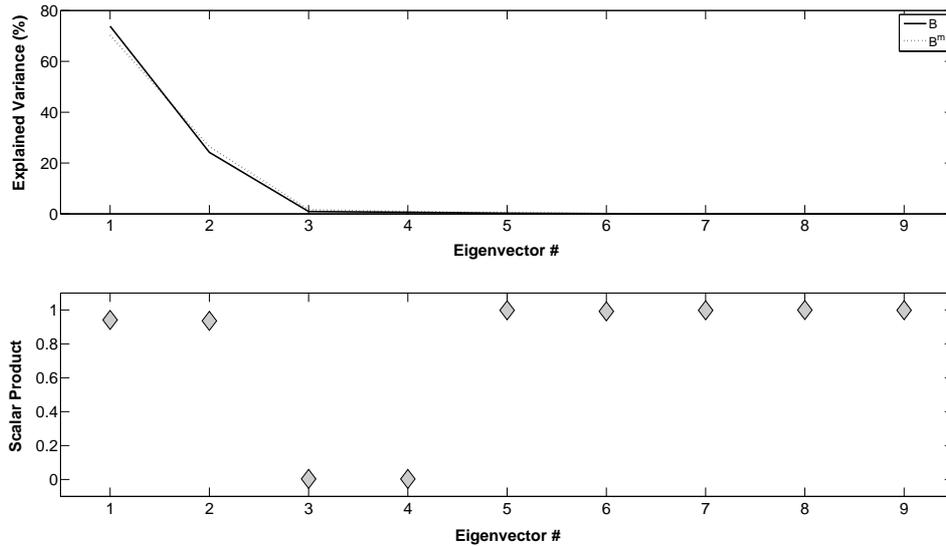


Figure 4.13: Top panel: explained variance over each eigenvector of \mathbf{B}^m used in LSI (dotted line) and the actual background error covariance matrix \mathbf{B} (solid line). Bottom: scalar product between pairs of eigenvectors of \mathbf{B}^m and \mathbf{B} .

calculated over the first year of prediction, and the skill of standard FFI is given in the dashed lines. Both example configurations corresponding to $\{c = 0.8, c_z = 0.9\}$ (black) and $\{c = 0.3, c_z = 1.2\}$ (red) are considered. Results show that best performances are found for $\beta \approx 2$, which is expected because the mean of the distribution coincides with the actual parametric error. More importantly, we observe an improvement of FFI-EPU over standard FFI (dashed line) for the entire range given by $\beta = (0, 3.5)$ and both model configurations.

In the right panel of figure 4.15, we investigate the dependence of the RMSE of FFI-EPU on the length of the short time interval over which the bias is corrected. The same relations apply for the right panel as for the left panel, but we set the scaling parameter $\beta \approx 2$ in order to focus on the impact of ΔT_{Bias} alone. As anticipated, a shorter time interval improves the performance of EPU for both configurations, because the linear error hypothesis is no longer valid for large lead times. We can see, however, that the improvement in skill over standard FFI applies for time intervals up to about 30 time steps for the first configuration (red) and up to as much as 40 time steps, corresponding to two months, for the second configuration (black). These time intervals can be interpreted as a measure of the duration of the linear regime, and are consistent with the largest (in absolute value) Lyapunov exponent of both configurations, $|\gamma_9^{c=0.8, c_z=0.9}| < |\gamma_9^{c=0.3, c_z=1.2}|$ in accordance with the theory of deterministic model dynamics [Nicolis, 2003]. For longer time intervals the correction procedure deteriorates skill.

Overall, the results of this section show the robustness of EPU for simple dynamics, and in situations in which large uncertainties with respect to the parameter range are present.

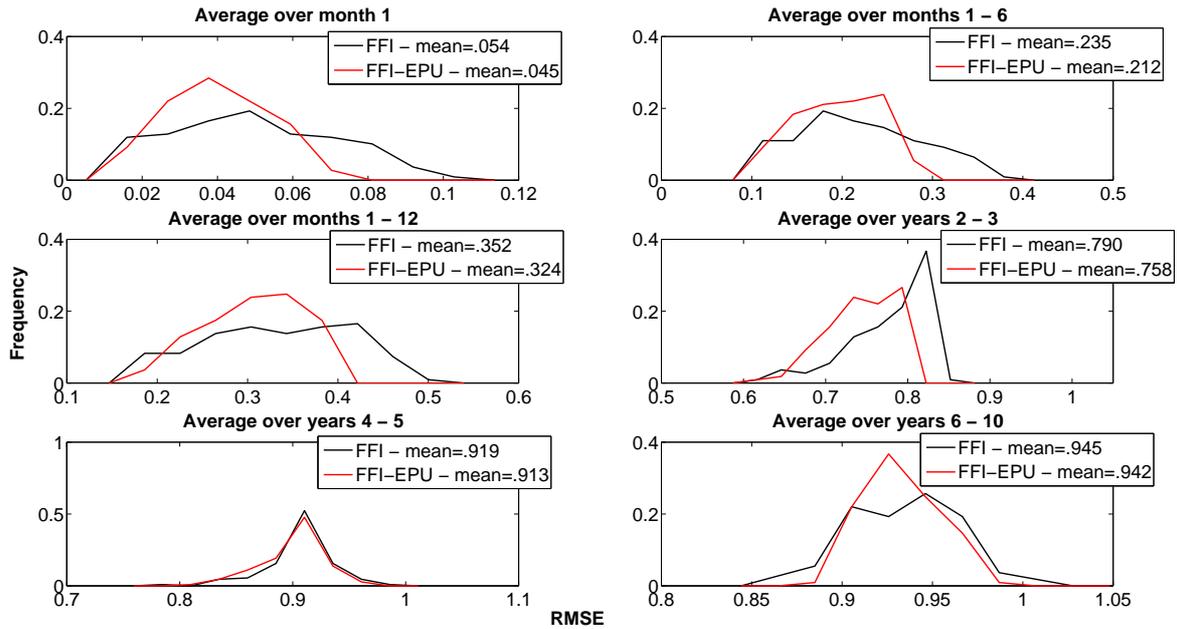


Figure 4.14: Distribution of the RMSE for FFI (black line) and FF-EPU (red line). The six panels refer to six different averaging periods and are indicated in the corresponding labels. Mean values of the distributions are given in the insets.

4.2.3 Summary of research line 2

FF-LSI shows a clear skill improvement over standard FFI for partially initialised systems. The skill improvement is larger than 5% for initialisation of the “tropical atmosphere” on forecast horizons of up to six “months”, and for the “ocean” up to one “year”, with marginal skill on longer horizons. Initialisation of the “extratropical atmosphere” shows no significant improvement, due to its weak coupling. We have also shown that our approximation of the real background error covariance is robust.

The implementation of EPU in conjunction with FFI reveals a reduction of the mean error by 10% up to the first forecast “year”, 5% up to the third, and a minor improvement for longer horizons. Significant skill is thus limited to seasonal time horizons, due to sampling errors, as well as a deviation of the drift behaviour from the linear assumption. We have shown that skill improvement is sensitive to the specification of the uncertainty range, as well as the time correction interval, but seems to be robust with respect to both.

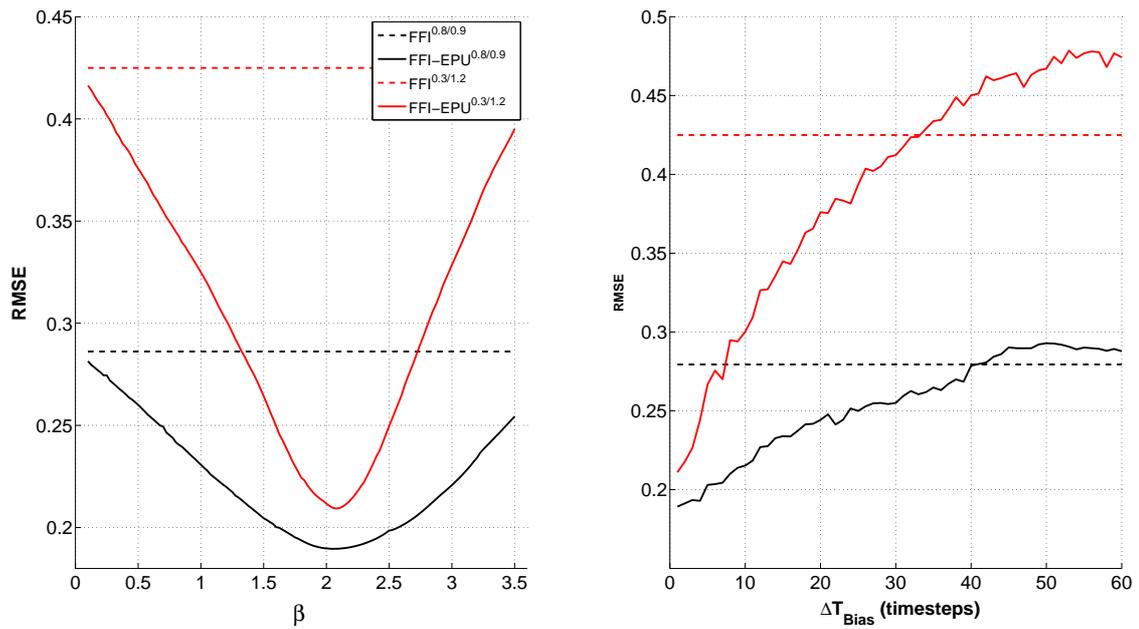


Figure 4.15: Left panel: RMSE of FF-EPU (solid line) over the first forecast year as a function of the width of the sampling distribution scaled with the parameter β for both example $\{c, c_z\}$ -configurations (black and red as indicated in the inset). Right panel: RMSE of FF-EPU over the first forecast year as a function of the bias correction interval ΔT_{Bias} , for the same configurations. FFI skill (dashed lines) is superimposed for reference in both panels.

5 Conclusion

In this study we have investigated the relative performance of Full Field and Anomaly Initialisation of climate predictions on seasonal-to-decadal time horizons under varying conditions of observational and model error. FFI assimilates observations directly into the model, whereas AI assimilates the observed anomalies onto an estimate of the mean model climate [Magnusson *et al.*, 2012].

We have also investigated the skill of two advanced schemes implemented in conjunction with FFI. Least-Squares-Initialisation [Carrassi *et al.*, 2014] borrows central concepts from data assimilation and has the aim of improving the initial conditions based on a least squares approach in which the observational and model statistics are taken into account. LSI approximates the background error covariances using the model anomalies, and is similar to the approach of Smith and Murphy [2007] in which the ocean is initialised. We investigate how LSI can improve forecast skill through the propagation of information across separate model compartments.

Exploring the Parameters Uncertainty [Carrassi *et al.*, 2014] is a short time drift correction technique applied during the forecast run. It exploits the fact that the evolution of the bias due to parametric model error is fully correlated in time, albeit difficult to correctly compute. EPU estimates the model bias by making a first order approximation in time. The Jacobian of the forecast model must be computed at the required time during the forecast run, and the parametric error is sampled from a specified uncertainty range.

Our study has been carried out using a low order climate model [Peña and Kalnay, 2004], necessitating little computational time and effort and allowing for robust inferences. Following the line of an OSSE [Bengtsson *et al.*, 1981] test bed, a “true” trajectory, which we call the nature, is simulated, from which the observations are sampled within a Gaussian white error. The true trajectory is targeted using imperfect models based on parametric error, initialised under the implementation of the above schemes incorporating the sampled observations. A central assumption on which our investigation is based is that the phase space spanned by the model and the nature is the same, neglecting the problem of unresolved scales.

We first discuss the results related to the first main research line spanning the comparison of FFI and AI under different observational error and model error scenarios. First-of-all, we have found that for the initialisation of single compartments, the best performance for long time horizons is obtained when the stabler component of the model, i.e. the “ocean”, is initialised. Having the slowest error growth rate, the “ocean” behaves like the system’s memory. Secondly, we have found that FFI depends sensitively on the observational error, whereas AI is not significantly improved when the observational error is reduced. This result is in line with regards to the error scaling properties of both schemes, which can be estimated using the unified formalism introduced by Carrassi *et al.* [2014]. The initial

conditions error of FFI is reduced for smaller observational errors, but under the addition of the model bias as in AI, the initial error scales less with regards to the observational error.

Thirdly, we have assessed the performance of either schemes for different scenarios of model bias. Both schemes are identical when no model bias is present, i.e. in the face of a perfect model. Anomaly Initialisation intends to avoid model drift occurring in the presence of a model bias, which can be detrimental to forecast skill [Magnusson *et al.*, 2012]. We have identified two scenarios in which either scheme outperforms the other. For models specified by erroneous “tropical” coupling parameters that we have termed $\{c, c_z\}$ -models, the forecast skill of AI decreases for larger model biases while that of FFI is independent of the bias. This scenario suggests that AI is improved only after model enhancements. For models specified by an erroneous forcing parameter r that we have termed r -models, the reverse is true, i.e. AI outperforms FFI. We have identified the reasons behind these various results by comparing the statistics of the model and nature PDFs. AI outperforms FFI for r -models, because a significant model bias is present, and other relevant differences between model and nature PDFs are limited to the second moments. For FFI, the model bias incurs initialisation outside of the model attractor, to which the model’s short-term response is a climate extreme, and resulting oscillations. Our 3D visualisation of this occurrence is a picture book example of shock-induced forecast errors, and is supported by experiments in which the “perfect model” is perturbed by introducing an artificial bias onto the observations, indicating a direct relationship between drift and perturbation magnitude (not shown). Initialisation with AI avoids the occurrence of a shock, because in the case of r -models, the model PDF is well approximated after a subtraction of the model bias. On the other hand, FFI outperforms AI for $\{c, c_z\}$ -models, because such models define a spectrum of attractors that differ significantly from that of the nature in their third and fourth moments. Although these models have equally significant biases, the biases do not alone account for the differences between model and nature PDFs. Thus, implementation of AI can result in initial conditions less consistent with the model PDF.

These results are entirely in line with the principle behind AI, and what one would expect of its performance. AI intends to initialise predictions closer to the model attractor [Magnusson *et al.*, 2012] by effectively adding the model bias onto the observations. This is merely a first order correction applied to the observations with the intention of approximating the model PDF, which is not guaranteed. The results suggest that a comparison of model and nature PDFs can inform the choice between AI and FFI. AI will be the initialisation scheme of choice for models with significant biases that are not accompanied by equally significant higher order differences in PDFs. One can say that AI intends to control and eliminate forecast errors occurring from a “translation” of the model attractor with respect to the nature attractor in phase space.

The results associated with the second main line investigating the performance of advanced schemes are as follows. LS-FFI shows a clear skill improvement over standard FFI for partially initialised systems. The skill improvement is larger than 5% for the “tropical atmosphere” on forecast horizons of up to six “months”, and for the “ocean” up to one “year”, with marginal skill on longer horizons. The “extratropical atmosphere”

shows no significant improvement, due to its weak coupling. We have further shown that our approximation of the real background error covariance is robust. This result contributes towards the discussion on adequate coupled DA schemes, making a case for the approximation of the forecast error covariances based on the model anomalies.

Finally, the implementation of EPU in conjunction with FFI reveals a reduction of the mean error by 10% up to the first forecast “year”, 5% up to the third, and a minor improvement for longer horizons. Significant skill is thus limited to seasonal time horizons, due to sampling errors, as well as a deviation of the drift behaviour from the linear assumption. We have shown that skill improvement is sensitive to the specification of the uncertainty range, as well as the time correction interval, but seems to be robust with respect to both.

Future work will require a validation of the above results from both research lines using models of higher complexity. The development of an alternative scheme to Anomaly Initialisation on the basis of an approximation of the model attractor beyond a first order correction of the bias is desirable. Investigations in this direction have not yet been undertaken by the author.

A Lists

A.1 List of Figures

2.1	Schematic illustrating progression from initial value problems with daily weather forecasts at one end, and century climate change projections as a forced boundary condition problem at the other. Seasonal and decadal prediction falls in between. After Meehl <i>et al.</i> [2009]	15
2.2	Evolution of an ensemble of initial points on the Lorenz (1963) attractor, for three sets of initial conditions in different phase space regions. Predictability is a function of the initial state. After Palmer [1993]	18
2.3	The potential predictability ratio F between the interannual variance of the seasonal mean and the corresponding weather noise induced variance for analysed DJF mean T_{850} as simulated by a ECMWF, b CCC model, c MPI model, d UKMO model. Contours are 0.5, 1, 2, 4, 8, and 16. Light (dark) shading corresponds to F ratios that are significantly greater than one at the 5% (1%) significance level. High terrain areas have been masked. After Zwiers and Kharin [1998]	22
2.4	Total variance for the global decadal mean surface air temperature predictions, split into the three sources of uncertainty. Orange fields represent the internal variability component. Insets: As in the main panel, but only for lead times less than 20 years for (left) the global mean and (right) a North American mean. Lead times shorter than 5 years are plotted using annual mean data to highlight how the internal variability component is vastly reduced when considering decadal mean data. After Hawkins and Sutton [2009]	24
2.5	Example of model bias, shown by the difference in HADCM3 model simulated and observed annual near surface temperatures averaged over the period 1960-2009. After Smith <i>et al.</i> [2013]	25
2.6	Illustration of bias correction for full field (left) and anomaly (right) initialisation. Thin black curves show the observed time series of annual mean global temperature. Coloured curves show the ensemble mean hindcasts, with different colours showing different start dates. (a) and (b) show absolute values (K). (c) and (d) show anomalies after adjustment for model biases. The solid grey and black dashed curves in (a) and (b) show values from the analyses and mean uninitialised model control runs respectively. After Smith <i>et al.</i> [2013]	26

2.7	Illustration of the properties of the probability distribution of the analysis T , given observations T_1 and T_2 , using either the least squares or the Bayesian approach. After Kalnay [2003].	28
3.1	(a)-(c): Attractors of the 3-component coupled system. The x -, y - and z - axes correspond to the same variables. The “ocean” (a) is strongly coupled to the “tropical atmosphere” (b), which in turn is weakly coupled to the “extratropical atmosphere” (c). The time series of the x -variable of all three compartments are plotted in (d).	42
3.2	Schematic of the experimental setup in phase space. The observations (red crosses) are sampled each month from the nature (green line) and used to initialise the forecasts (blue lines). The model control run is shown (orange line) systematically below the nature, indicating a model bias. . .	44
3.3	Characterisation of the configurations using Lyapunov exponents. (a) For $\{c, c_z\}$ -models: Distributions of the nine Lyapunov exponents. (b) For r -models: Lyapunov exponents as a function of the r	48
3.4	Distributions of the Kaplan-Yorke dimensions and the dissipation (top left and right) for $\{c, c_z\}$ -models. Kaplan-Yorke dimension and dissipation (bottom left and right) as a function of r	49
4.1	RMSSS as a function of the forecast lead time for FFI (left panels) and AI (right panels). Top / Bottom panels refer to the configurations $\{c = 0.8, c_z = 0.9\} / \{c = 0.3, c_z = 1.2\}$ respectively. Different colors represent initialisation of the full system (black), “ocean” (green), “tropical atmosphere” (magenta) and “extratropical atmosphere” (blue).	55
4.2	RMSSS as a function of the standard deviation of the observational error, σ^o , expressed as a fraction of the system’s natural variability. The model configuration corresponds to $\{c = 0.8, c_z = 0.9\}$. The six panels refer to six different averaging periods and are indicated in the corresponding labels. FFI (black line), AI (red line).	56
4.3	RMSSS as a function of the RMS Bias for (a) $\{c, c_z\}$ -models and (b) r -models. The six panels refer to six different averaging periods and are indicated in the corresponding labels. FFI (black line), AI (red line). . . .	57
4.4	Normalized mean forecast error (i.e. normalized bias) of the x_t / z_t variables, for three $\{c, c_z\}$ -models / three r -models respectively, as a function of the lead time in “years”. The different colors correspond to the different models indicated in the plot.	58
4.5	RMSE ratio of AI / FFI (colorbar) for all nine variables as a function of the bias on the y-axis and the initial drift on the x-axis. (a) Each coloured point in a single panel corresponds to one of the 109 $\{c, c_z\}$ -models. (b) Each coloured point in a single panel corresponds to one of the 41 r -models.	59

- 4.6 Three-dimensional view of the “extratropical” attractors of the model (red) and nature (blue) for $r = 68$ (left panels) and $r = 34$ (right panels). Top panels show the initial conditions of AI (black triangles) and FFI (green stars). Middle panels show the 1-“month” predictions initialised with FFI. Bottom panels show the 1-“month” predictions initialised with AI. 61
- 4.7 Three-dimensional view of the “tropical” attractors of the model (red) and nature (blue) for $\{c, c_z\} = \{0.8, 0.9\}$. The initial conditions of FFI (green stars) and AI (black triangles) are shown in the top left and right panels respectively. The one “month” predictions are shown in the bottom panels after FFI (left) and AI (right). Note that the perspective has been altered with respect to Fig. 4.6 for better recognition. 64
- 4.8 PDFs of the nature (blue), model (red), the difference between model and nature (green), and the distribution associated with the initial conditions for AI (black). The nine panels correspond to the nine variables as indicated. (a) $\{c, c_z\} = \{0.8, 0.9\}$, (b) $r = 68$ 65
- 4.9 Distributions of (a) the biases and (b) the differences in second order moments for $\{c, c_z\}$ -models (blue) and r -models (red). Panels correspond to the nine variables as indicated. 66
- 4.10 Difference in (a) third order moments and (b) fourth order moments between model and nature plotted as a distribution for $\{c, c_z\}$ -models (blue) and r -models (red). Panels correspond to the nine variables as indicated. 67
- 4.11 Bhattacharyya coefficient measuring the similarity between the model distribution and the distribution associated with the initial conditions for AI, as a function of the bias. Each point in a single panel corresponds to one of (a) the 109 $\{c, c_z\}$ -models (b) the 41 r -models. Panels correspond to the nine variables as indicated. 69
- 4.12 RMSSS as a function of the tuning coefficient α for LSI. The panels refer to six different averaging periods as indicated. The colors represent initialization of the full system (black), the “ocean” alone (green), the “tropical atmosphere” alone (pink), and the “extratropical atmosphere” alone (blue). The values of the standard FFI are displayed for reference (dots). 73
- 4.13 Top panel: explained variance over each eigenvector of \mathbf{B}^m used in LSI (dotted line) and the actual background error covariance matrix \mathbf{B} (solid line). Bottom: scalar product between pairs of eigenvectors of \mathbf{B}^m and \mathbf{B} 74
- 4.14 Distribution of the RMSE for FFI (black line) and FF-EPU (red line). The six panels refer to six different averaging periods and are indicated in the corresponding labels. Mean values of the distributions are given in the insets. 75

- 4.15 Left panel: RMSE of FF-EPU (solid line) over the first forecast year as a function of the width of the sampling distribution scaled with the parameter β for both example $\{c, c_z\}$ -configurations (black and red as indicated in the inset). Right panel: RMSE of FF-EPU over the first forecast year as a function of the bias correction interval ΔT_{Bias} , for the same configurations. FFI skill (dashed lines) is superimposed for reference in both panels. 76

B Bibliography

- Anderson, D. L. T., [2008] Overview of seasonal forecasting *Seasonal climate: forecasting and managing risk* NATO Science Series, pp. 47-68. Dordrecht, The Netherlands: Springer Academic Publishers
- Benettin, G., L. Galgani, A. Giorgilli and J.-M. Strelcyn , [1980] Lyapunov Characteristic Exponents for Smooth Dynamical Systems and for Hamiltonian Systems: A Method for Computing All of Them. Part 2: Numerical Application *Meccanica* **15**, 21
- Bengtsson, L., M. Ghil, and E. Källén (Eds.), [1981] *Dynamic Meteorology: Data Assimilation Methods* Springer Verlag, New York/Heidelberg/Berlin, 330 pp
- Bengtsson, L. and K. I. Hodges, [2006] A note on atmospheric predictability *Tellus*, **58**, 154-157
- Bhattacharyya, A., [1943] On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, **35**, 99-110
- Bocaletti, S., J. Kurths, G. Osipov, D. L. Valladares and C. S. Zhou, [2002] The synchronization of chaotic systems *Physics Reports* **366**, 1-101
- Broecker, W. S., [1991] The Great Ocean Conveyor *Oceanography* **4**, 79–89.
- Buizza, R., M. Miller and T. N. Palmer, [1999] Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System *Q. J. R. Meteorol. Soc.* **125**, 2887-2908
- Buizza, R., [2000] Chaos and weather prediction, Meteorological Training Course Course Lecture Series, European Center for Medium-Range Weather Forecasts (ECMWF), Reading, UK
- Cane, M. A., S. E. Zebiak and S. C. Dolan , [1986] Experimental forecasts of El Niño *Nature* **321**, 827-832
- Carrassi, A., Vannitsem S., and Nicolis C., [2008] Model error and sequential data assimilation: A deterministic formulation *Quart. J. Roy. Meteor. Soc.* **134**, 1297–1313.
- Charney, J. G., R. G. Fleagle, H. Riehl, V. E. Lally and D. Q. Wark, [1966] The feasibility of a global observation and analysis experiment *Bull. Am. Meteorol. Soc.* **47**, 200-220

- Charney, J. G., A. Arakawa, D. J. Baker, B. Bolin, R. E. Dickinson, R. M. Goody, C. E. Leith, H. M. Stommel, and C. I. Wunsch, [1979] *Carbon Dioxide and Climate: A Scientific Assessment*, National Research Council, Ad Hoc Study Group on Carbon Dioxide and Climate, Washington, DC: National Academy Press
- Carrassi, A., R. J. T. Weber, V. Guemas, F. J. Doblas-Reyes and D. Volpi, [2014] Full Field and Anomaly Initialization using a low order climate model: a comparison and proposals for advanced formulations *Nonlin. Processes Geophys.* **in review**
- Daley, R. [1991], *Atmospheric Data Analysis*, (Cambridge University Press), Cambridge, ISBN 9780521382151.
- de Swart, H. E., [1990], Vacillation and predictability properties of low-order spectral models. Ph.D. thesis, Univerisity of Utrecht, Netherlands, 121 pp.
- Dee, D., [2013], Coupled data assimilation development at ECMWF. Oral presentation presented at Sixth Symposium on Data Assimilation, University of Maryland, USA. [http : //das6.umd.edu/program/das6_program.html](http://das6.umd.edu/program/das6_program.html)
- Doblas-Reyes F., García-Serrano J., Lienert F., Pintò Biescas A. and Rodrigues L.R.L., [2013], Seasonal climate predictability and forecasting: status and prospects *WIREs Clim Change* **4**, 245-268, doi:10.1002/wcc.217
- Ghil, M. and A. W. Robertson, [2002], “Waves” vs “particles” in the atmosphere’s phase space: A pathway to long-range forecasting? *PNAS* **99**, 2493-2500
- Guemas, V., Doblas-Reyes, F. J., Andreu-Burillo, I., Asif, M., [2013], Retrospective prediction of the global warming slowdown in the past decade *Nature Climate Change* **3**, 649-653
- Haines, K., L. Hermanson, C. Liu, D. Putt, R. Sutton, A. Iwi and D. Smith, [2009] Decadal climate prediction (project GCEP) *Phil. Trans. R. Soc.*, **367**, 925-937, doi:10.1098/rsta.2008.0178
- Hawkins, E. and R. Sutton, [2009] The potential to narrow uncertainty in regional climate prediction *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107. doi: <http://dx.doi.org/10.1175/2009BAMS2607.1>
- Hazeleger, W., V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F.J. Doblas-Reyes, K. Wyser and M. Caian, [2013] Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett.*, **40**, 1794-1798, doi:10.1002/grl.50355.
- Hegerl, G. C., and Coauthors, [2007] Understanding and attributing climate change. Global climate projections. *Climate Change 2007: The Physical Science Basis*, S. Solomon *et al.*, Eds., Cambridge University Press, 663-745
- Hilborn, R. [1994], *Chaos and Nonlinear Dynamics*, Oxford University Press, Oxford, UK

- Hoke, J. E., Anthes, R. A., [1976] The Initialization of Numerical Models by a Dynamic-Initialization Technique *Mon. Wea. Rev.*, **104**, 1551-1556
- Hoskins, B. J., [1983] Dynamical Processes in the atmosphere and the use of models *Quart. J. Roy. Meteor. Soc.*, **109**, 1-21
- IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Kalman, R., [1960], A new approach to linear filtering and prediction problems *Trans. ASME, J. Basic Eng* **82**, 35–45
- Kalnay, E. [2003], *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK
- Kaplan, J. L. and J. A. Yorke, [1979] Chaotic behaviour of multidimensional difference equations. In *Functional differential equations and approximation of fixed points*, H.-O. Peitgen and H. O. Walther, editors, pp. 204-227. Lecture notes in mathematics, 730. Springer Verlag, Berlin
- Keeling, R.F., S.C. Piper, A.F. Bollenbacher and J.S. Walker, [2009] Atmospheric CO₂ records from sites in the SIO air sampling network. In Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. doi: 10.3334/CDIAC/atg.035
- Lee, T. C. K., F. W. Zwiers, X. Zhang and M. Tsao, [2006], Evidence of decadal climate prediction skill resulting from changes in anthropogenic forcing *J. Clim.* **19**, 5305–5318
- Lorenz, E.N. [1963], Deterministic Nonperiodic Flow *J. Atmos. Sci.* **20**, 130–141
- Lorenz, E.N. [1969], The predictability of flow which possesses many scales of motion *Tellus* **34**, 505–513
- Lorenz, E.N. [1982], Atmospheric predictability experiments with a large numerical model *Tellus* **3**, 289–307
- Madden, R. A., [1976], Estimates of the Natural Variability of Time-Averaged Sea-Level Pressure *Mon. Wea. Rev.* **104**, 942-952
- Madden, R. A. and D. J. Shea [1978], Estimates of the Natural Variability of Time-Averaged Temperatures over the United States *Mon. Wea. Rev.* **106**, 1695-1703
- Madden, R. A., [1989], On Predicting Probability Distributions of Time-Averaged Meteorological Data *J. Clim.* **2**, 922-925

- Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni and T. Stockdale, [2012], Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors *Clim Dyn*, doi 10.1007/s00382-012-1599-2
- Milankovich, M., [1941], Kanon der Erdbestahlung und seine Anwendung auf das Eiszeitenproblem vol. 33, Royal Serbian Sciences, special publication 132, section of Mathematical and Natural Sciences, Belgrad
- Meehl, G.A., L. Goddard, J. Murphy, R.J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M.A. Giorgetta, A.M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer and T. Stockdale, [2012], Decadal Prediction *Bull. Amer. Meteor. Soc.* **90**, 1467–1485, doi: <http://dx.doi.org/10.1175/2009BAMS2778.1>
- Nicolis, C. [2003], Dynamics of model error: Some generic features *J. Atmos. Sci.* **60**, 2208–2218.
- Ott, E., [2002], *Chaos in Dynamical Systems*, Cambridge University Press, Cambridge, UK
- Palmer, T. N., [1993], Extended-Range Atmospheric Prediction and the Lorenz Model *Bull. Amer. Meteor. Soc.* **74**, 49-65 .
- Palmer, T. N., [1998], A Nonlinear Dynamical Perspective on Climate Prediction *J. Clim.*, **12**, 575-590
- Parrish, D.F., and J.C. Derber, [1992], The National Meteorological Center's Spectral Statistical-Interpolation Analysis System *Mon. Wea. Rev.* **120**, 1747-1763 .
- Peña, M., and E. Kalnay, [2004], Separating fast and slow modes in coupled chaotic systems *Nonlin. Proc. Geophys.* **11**, 319-327
- Pierrehumbert, R. T. [2010], *Principles of Planetary Climate*, Cambridge University Press, Cambridge, UK
- Plass, G. N., [1956] Effect of carbon dioxide variations on climate *Am. J. Phys.* **24**, 376-387
- Rowell, D. P., [1998] Assessing Potential Seasonal Predictability with an Ensemble of Multidecadal GCM Simulations *J. Clim.* **11**, 109-120
- Smith, D.M. and J.M. Murphy, [2007] An objective ocean temperature and salinity analysis using covariances from a global model *J. of Geophys. Res.* **112**, C02022, doi:10.1029/2005JC003172
- Smith, D.M., S. Cusack, A.W. Coleman, C.K. Folland, G.R. Harris and J.M. Murphy, [2007], Improved surface temperature prediction for the coming decade from a global climate model *Science* **317**, 796-799.

- Smith, D. M., A. A. Scaife and B. P. Kirtman, [2012], What is the current state of scientific knowledge with regard to seasonal and decadal forecasting? *Environ. Res. Lett.*, **7**, 1-11
- Smith, D. M., R. Eade, and H. Pohlmann [2013] A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Clim. Dyn.*, doi: 10.1007/s00382-013-1683-2.
- Solomon, S., R. R. Garcia, F. S. Rowland and D. J. Wuebbles, [2013] On the depletion of Antarctic Ozone *Nature* **321**, 755-758
- Somerville, R. C. J., [1979] Predictability and prediction of ultra-long planetary waves. Pre-prints of the Amer. Meteor. Soc. Fourth Conference on Numerical Weather Prediction, Silver Spring, MD, 182-185
- von Storch, H. and F. W. Zwiers, [1999], *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, UK
- Strogatz, S. H., [2000], *Nonlinear Dynamics And Chaos*, Westview Press, The Perseus Books Group, Cambridge, MA
- Thompson, L. G., [2000] Ice core evidence for climate change in the Tropics: implications for our future *Quat. Sci. Rev.*, **19**, 19-35
- Trevisan, A. and L. Palatella, [2011] Chaos and weather forecasting: the role of the unstable subspace in predictability and state estimation problems *International Journal of Bifurcation and Chaos*, **21**, 3389-3415
- Toth, Z., M. Peña, [2007] Data assimilation and numerical forecasting with imperfect models: The mapping paradigm *Physica D: Nonlinear Phenomena*, **230**, 146-158
- Tucker, W., [1999] The Lorenz attractor exists *C. R. Acad. Sci. Paris*, **328**, 1197-1202
- Yeager, S., A. Karspeck, G. Danabasoglu, J. Tribbia and H. Teng, [2012] A Decadal Prediction Case Study: Late Twentieth-Century North Atlantic Ocean Heat Content *J. Climate*, **25**, 5173-5189
- Zwiers, F. W., V. V. Kharin, [2007] Intercomparison of interannual variability and potential predictability: an AMIP diagnostic subproject *Clim. Dyn.*, **14**, 517-528

Acknowledgements

I am grateful to Prof. Werner Aeschbach-Hertig for consenting to supervise my work at the external IC3, and helping me obtain the PROMOS grant from the German Academic Exchange Service (DAAD). Without his support, this thesis would not have been possible. I would like to thank Prof. Francisco Doblado-Reyes for his continuous encouragement from the start of the project. His enthusiasm has kept an open door for me with regards to any issue, and helped me to understand the broader scientific context in which this work finds its place.

I am much indebted to Alberto Carrasi. What more can a student wish from his supervisor, than patience, engagement, esteem, and enthusiasm for turning questions into discussions? To be shown what it means to be a full-blooded¹ scientist. Alberto also helped me obtain a grant to take part in the Interdisciplinary Summer School on Data Assimilation in Geosciences at the University of Maryland, and visit the U.S. for the first time.

Last but not least I am grateful to my parents for supporting my (rather long) studies. I also look back on the many friends that helped pull me through - too many to list here. I only mention Michel Wiemer, Dominik Daume, and Phillipp Metzler.

¹italian

Erklärung:

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den (Datum)

.....